PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

University Mohamed Khider of Biskra
Faculty of Exact Sciences and Natural and Life Sciences

Third cycle doctoral thesis in Mathematics

OPTION : **STATISTICS**

Title

Nonparametric Estimation of the Copula Function with Bivariate Twice Censored Data

Presented by : TOUMI *Samia*

Members of the jury:

BRAHIMI Brahim	Pr,	University of Biskra	**President**.
BEN ATIA Fatah	Pr,	University of Biskra	**Supervisor.**
BOUKELOUA Mohamed	MCA,	National Polytechnic Institute of Constantine	**Co-Supervisor.**
YAHIA Djabrane	Pr,	University of Biskra	**Examiner.**
DJEFFAL El Amir	Pr,	University of Batna	**Examiner.**

College year 2024

Aknowledgemnt

I would like to express my sincere gratitude to my supervisor, *Pr. BENATIA Fatah*, *Co-Supervisor. BOUKELOUA Mohamed* and *Dr. IDIOU Nesrine* for their guidance, support, and mentorship throughout my PhD program. I would like to express my sincere thanks to *Pr. BRAHIMI Brahim*, University of Biskra ,*Pr. DJABRANE Yahia*, University of Biskra and *Pr. DJEFFAL El Amir*, University of Batna because they agreed to spend their time for reading and evaluating my thesis.

I would also like to extend my thanks to my family for their love, support, and encouragement during this long and challenging journey.

Finally, I would like express my gratitude to all the professors, colleagues and friends , including all the members of the *LMA* laboratory, and for everyone who has supported, directly or indirectly this work. . .

Résumé

L'objectif de cette thèse est d'étudier l'estimation non paramétrique de la fonction de la copule en présence de données bivariées doublement censurées. En supposant que la copule fonctionne des variables de censure de droite et de gauche sont connues, nous proposons un estimateur de la fonction de distribution conjointe des variables d'intérêt, puis nous dérivons un estimateur de leur fonction de copule. En utilisant une représentation de l'estimateur proposé de la fonction de distribution conjointe comme une somme de variables indépendantes et distribuées de manière identique, nous établissons la faible convergence de la copule empirique et de la simulation. Après cela, nous avons étudié l'estimation du noyau de la fonction copule de deux variables aléatoires censurées deux fois. Nous introduisons donc deux estimateurs de noyau de la fonction de distribution conjointe des deux variables d'intérêt. Ensuite, nous utilisons ces estimateurs pour proposer deux estimateurs lissés de la fonction de la copule. Nous prouvons également la faible convergence des estimateurs proposés vers un processus gaussien étroitement centré

.

Mots-clés : Copules, processus empirique de la copule, données censurées deux fois, limite de produit estimateur, estimateurs lissés, convergence faible.

Abstract

The aim of this thesis is to study the nonparametric estimation of the copula function in the presence of bivariate twice censored data. Assuming that the copula functions of the right and the left censoring variables are known, we propose an estimator of the joint distribution function of the variables of interest, then we derive an estimator of their copula function. Using a representation of the proposed estimator of the joint distribution function as a sum of independent and identically distributed variables, we establish the weak convergence of the empirical copula and simulation. After that, we studied the kernel estimation of the copula function of two twice censored random variables. So, we introduce two kernel estimators of the joint distribution function of the two variables of interest. Then, we use these estimators to propose two smoothed estimators of the copula function. We also prove the weak convergence of the proposed estimators to some tight centered Gaussian processes. Finally, we illustrate the performances of our estimators through a simulation study.

Key words. Copulas, empirical copula process, twice censored data, product-limit estimator, Smoothed estimators, weak convergence.

Contents

Introduction

Copulas represent a very useful tool to describe the dependence structure between two random variables. They allow to study this structure without requiring any knowledge

about the marginal distributions. The estimation of the copula function have aroused the interest of statisticians from many decades. [12] was the first to introduce the empirical copula as a nonparametric estimator of the copula function. The properties of this estimator have been considered by many authors such as [13], [14], [15], [10] and [17]. The copula C of a couple of real random variables (r.r.v.) $X = (X_1, X_2)$, with a joint distribution function F and continuous margins F_{X_1} and F_{X_2}, is defined on $[0,1]^2$ by $C(u,v) = F\left(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)\right)$, where $\varphi^{-1}(u) = \inf\{x \in \mathbb{R} : \varphi(x) \geq u\}$ is the generalized inverse of a non decreasing function φ. Sklar's theorem (see [49]) shows that for all $(x_1, x_2) \in \mathbb{R}^2$, $F(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$. We cite for instance the works of [13], [14], [15] and [10]. Moreover, [17] established the weak convergence of this estimator. However, the empirical copula is based on a sample comprising true realizations of the variable of interest, i.e., complete data; but in the practice, one or more censoring phenomena may prevent the observation of the variable of interest and provide only a partial information about it. For example, in the case of right censoring, when a data is censored, the statistician only knows that the variable of interest is greater than the observed value. Bivariate right censored data have been extensively studied in the literature, given their usefullness in the practice; we quote for instance the works of [40], [34], [2] and [24]. The empirical copula for bivariate right censored data has been introduced by [23] for some particular models. Its weak convergence is also established in the same paper. Other copula models for bivariate right censored data have been studied in [47], [45], [8], [28], [27] and [26].

Althoug the right censoring is the most popular type of censored data, more complicated situations can also be encountered in the practice involving right and left censoring at the same time. [43] considered one of these situations where the variable of interest is right censored by another variable, the minimum of the two variables is itself left censored and the three latent variables are independent. This is the so called twice censored data model. [39] dealt with a practical situation that corresponds to this model. Moreover, [43] proposed and established the asymptotic properties of a product-limit estimator of the survival function under this model. Then, after the pioneer paper of [43], many authors study the model of twice censored data in the univariate case such as [32], [6] and [7] as well as in the conditional case such as [38] and [3]. So, we are interested in the nonparametric estimation of the joint distribution function and the copula function of a couple of r.r.v. $X = (X_1, X_2)$, where X_1 and X_2 are both twice censored. For that, we draw on the work of [23] , we consider a situation that corresponds to one of the three models studied in this paper, by assuming that the copula functions of the right and the left censoring variables are known. This assumption holds for example when the right (resp. left) censoring variables are independent. Under this assumption, we propose a nonparametric estimator F_n of the joint distribution function of X. Then, we derive from this letter the empirical copula C_n that we propose as an estimator of the copula function. In the case of bivariate right censored data, [23] proved the weak convergence of the empirical copula using a representation of the corresponding estimator of the joint distribution function as a sum of independent and identically distributed (i.i.d.) centered random variables. Such a representation was established by [35]. This

thesis is organized as follows:

chapter 1: The first chapter is essentially a reminder some basic notions, we start with foundations definitions like the distribution function, the empirical distribution function, the survival function... etc

chapter 2: The second chapter is mainly devoted to the design of copulas and their properties. We introduce the notion of the bivariate copula, and also we devote a section for different types of copulas, namely the usual copulas...

chapter 3: In this chapter, we difine type of censorship and we intoducde nonparametic estimation for type censoring model.

chapter 4: In this chapter, we have introduced a new copula estimator for censored bivariate data based on the classical estimation method of moments, presented in a non-parametric estimation framework. This chapter is divided into two parts the first focuses on the estimation of this new estimator when the data are twice -censoring, i.e. the two variables are twice-censored at the same time. In the second part, we have presented the weak convergence...and simulation.

chapter 4: The previous chapter also allowed us to conclude smoothed copula estimators for bivariate twice censored data; This chapter is divided into two parts the first focuses on the kernel copula estimators and weak convergence of the proposed estimators and second part, simulation study...

Scientific Contributions

Publications based on this thesis

- S. Toumi, M. Boukeloua, N. Idiou and F. Benatia (2022). Nonparametric Estimation of the Copula Function with Bivariate Twice Censored Data. Boletim da Sociedad Paranaenses da. Matematica (3s.) v. 2024 (42) : 1–22.

Conference papers

- TOUMI S., BOUKELOUA M., IDIOU N., BENATIA F. (26-27 Octobre 2022) “Weak convergence of the empirical Copula with bivariate twise censoring” 6 th International Workshop on applied Mathematics and Modeling WIMAM ’2022’ Guelma university. Guelma, Algerie.

- TOUMI S. (19-21 decembre 2022) “I.I.D.Representation of the product-limit Estimator under twise censoring” International Conference on Operator Theory. ICOT-22, LPMA. Sousse, Tunisie.

- TOUMI S., BOUKELOUA M., IDIOU N., BENATIA F. (11-12 October 2023) “Weak Convergence of kernel copula estimators under twice censoring” International Conference on the evolution of Contemporary Mathematics and their Impact in Science and Technology, Constantine.

- TOUMI S., BENATIA F. (26-27 November 2023) “Simulation study of kernel copula estimation under censoring” International Conference on Contemporary Mathematics and its Applications, ICCMA-, Mila.

- Toumi S. (13 May, 2023). « Nonparametric Estimation of the Copula Function. ».Second national Conference Of Apllied Mathematics and Didactics 2NCAMD2023, ENS Assia Djebar Constantine , Algerie.

- Toumi S., Benatia.F, (14-15 Mai, 2023). “ Empirical Copula for Twice Censored Data ”. first national Applied Mathematics Seminar, 1st-NAM’23, M. Khider University Biskra, Algerie.

Chapter 1

Preliminaries

We describe some of the basic concepts here so that you may utilize them in the next chapter.

Definition 1.1 (***The distribution function***)

The distribution function F_X, describes the probability that a variate X takes on a value less than or equal to a number x in $[0,1]$.
The distribution function is therefore related to a continuous probability density function $f(x)$ by

$$\forall x \in \mathbb{R}, \quad F_X(x) := P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt,$$

so $f(x)$ (when it exists) is simply the derivative of the distribution function

$$f(t) = F_X^{'}(t).$$

Similarly, the distribution function is related to a discrete probability F_X by

$$F_X(x) := P(X \leq x) = \sum_{X \leq x} P(x)$$

Definition 1.2 (***The empirical distribution function***)

Let $X_1, X_2, ..., X_n$ be a sample of size n $(n \geq 1)$, the empirical distribution function F_n is defined by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq x}, \quad \forall x \in \mathbb{R}$$

Theorem 1.1

There exists a constant C such that for all $v > 0$, we have

$$P\left(\sqrt{n}\sup_{x\in\Re}\left|F_n(x) - F(x)\right| > v\right) \leq C\exp\left(-2v^2\right).$$

[37] showed that this Theorem is true for $C = 2$ and that this constant is the best that can be obtained. Indeed, [33] showed that if F is continuous, we have and when v increases the right side is equivalent to $2\exp\left(-2v^2\right)$.

Theorem 1.2

The empirical process $\sqrt{n}(F_n - F)$ converges weakly to a centered Gaussian process with covariance function given by

$$\Gamma(s,t) := F(s)(1 - F(t)) \text{ for } s \leq t.$$

Proof: See [44] page 97. ■

Definition 1.3 (*The survival function*)

Let the lifetime X be a continuous random variable with distribution function $F(x)$ and probability density function $f(x)$ on the interval $[0;\infty[$, its survival function or reliability function is:

$$\begin{aligned} S(x) &= \bar{F}(x) \\ &= P(X > x) = \int_x^\infty f(t)\,dt \\ &= 1 - F(x), \quad x \geq 0 \end{aligned}$$

Every survival function $S(t)$ is monotonically decreasing, i.e. $S(u) \leq S(v)$ for all $u > v$.

The survival function can be related to the probability density function $f(x)$ and the hazard function $\lambda(t)$

- $f(t) = -S'(t)$
- $\lambda(t) = -\frac{d}{dt} \log S(t)$.

Definition 1.4 (*The empirical survival function*)

Let $X_1, X_2, ..., X_n$ be a sample of size n $(n \geq 1)$, the empirical distribution function noted by S_n is given by :

$$\begin{aligned} S_n(x) &= 1 - F_n(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} I_{X_i > x}, \qquad \forall x \in \mathbb{R} \end{aligned}$$

Corollary 1.1

By analogy to the cases of complete and right-censored data, Ren (1997) proposed to estimate the density f of X by

$$f_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right)$$

Definition 1.5

where h_n is called the smoothing parameter and the function K is called the kernel. This estimate is the density, i.e.

1. $f_n(x) \geq 0, \forall t$
2. $\int_{\mathbb{R}} f_n(x)\,dx = 1 \implies \int_{\mathbb{R}} \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right) dx = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} K(w)\,dw = 1.$

Here are some examples of common kernels:

$K(v) = \frac{1}{2} I_{\{|v| \leq 1\}}$: rectangular kernel.

$K(v) = (1 - |v|)\, I_{\{|v| \leq 1\}}$: triangular kernel.

$K(v) = \frac{15}{16}\left(1 - v^2\right)$: quadratic kernel.

$K(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-v^2/2\right)$: Gaussian kernel.

$K(v) = \frac{3}{4}\left(1 - v^2\right) I_{\{|v| \leq 1\}}$:Epanchenikov or parabolic kernel.

Chapter 2

Copula Conseption

In this chapter we define some of the basic conceptions, in order to take advantage of them in the next.

2.1 Bivariate Copula

The construction of the copulas is based on the properties of the distribution functions (fds). We recall below some important properties of bivariate fds. Let (X_1, X_2) be a couple of positive $r.r.v.$ with support $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and joint distribution function F defined like this

$$\forall (x_1, x_2) \in \mathbb{R}^2 : \quad F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2).$$

We call marginal laws the laws of X_1 and X_2 taken separately. We can express the fds of these marginal laws as a function of F. For example, for X_1 we obtain

$$F_1(x) := P(X_1 \leq x) = \lim_{x_2 \to \infty} F(x_1, x_2)$$

and identicall to X_2

$$F_2(x) := P(X_2 \leq x) = \lim_{x_1 \to \infty} F(x_1, x_2)$$

Recall that the X and Y vas are independent if and only if

$$\forall (x_1, x_2) \in \mathbb{R}^2 : \quad F(x_1, x_2) = F_1(x_1) F_2(x_2).$$

Theorem 2.1 (*Sklar's theorem*)

This theory is essential to the theory of copulas. It was established by Sklar in 1959, *whereby it precisely determines the relationship between the two variables, F_1 and*

F_2's marginal univariate distribution and the whole bivariate distribution F, based on the joint distribution F.

Theorem 2.2

If F is the fd of $(X_1; X_2)$, then there is a two-dimensional copule C such that

$$\forall (x_1, x_2) \in \mathbb{R}^2 : \quad F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

Theorem 2.3

F is the bivariate function of the F_1 and F_2 marginals. The copule C associated with F is given by

$$\forall (u, v) \in \mathbb{R}^2 : \quad C(u, v) = F\left(F^{-1}(u), F^{-1}(v)\right)$$

2.1.1 Density of the copula

The copulas accept probabilistic densities. If the density c associated with the copule C is present, it is defined as follows:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v},$$

where $c : I^2 \to \mathbb{R}$ If the fd joint F is absolutely continuous, using Sklar's theory, one

The density of an arbitrary pair $(X; Y)$ may be expressed as the product of its copule's density and its marginal f and g by

$$h(x, y) := c(F(x), G(y)) f(x) g(y).$$

2.1.2 Copula properties

First of all, one observes that the copules exhibit certain characteristics.

a. The margins are uniform, i.e.

$C(u,v) = 0$ if $u \leq 0$ or $v \leq 0$

$C(u,0) = 1$ if $u \geq 1$ and $v \geq 1$

$C(u,v) = u$ if $v \geq 1$ et $C(u,v) = v$ if $u \geq 1$.

b. The continuity : It should be noted that copulas are continuing functions.

$\forall u_1, u_2, v_1, v_2 \in I$, we have

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

c. The symmetry : That C is symmetric is said to

$$\forall u, v \in I \quad :C(u,v) = C(v,u)$$

d. The ordre:

$$\forall u, v \in I \quad :C_1(u,v) \leq C_2(u,v)$$

e. for every $u_1, u_2, v_{1,} v_2$ in I such that $u_1 \leq u_2$ and $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$$

f. Invariance by strictly increasing transformation: One of the fundamental theories in the theory of copules is the invariance via strictly increasing transformations. Let X and Y be a continuous have the marginals F and G and the C_{XY} copule. Given that and are two strictly increasing functions, then

$$C_{\alpha(X)\beta(Y)}(x,y) : C_{X,Y}.$$

2.2 Bivariate copula families

Many families of copulas have been proposed in the literature. The most commonly used ones are introduced in this section. A more comprehensive list can be found in Nelsen.

2.2.1 Usual Copulas

The family of copulas is bounded by what is referred to as the Fréchet-Hoeffding bounds.

We have

$$\forall (u,v) \in I^2 : \quad M(u,v) = \max(u+v-1,0) \le C(u,v) \le \min(u,v) = m(u,v)$$

m is the distribution function of the couple $(U;U)$; it is called the minimum copula or comonotonic copula.

M is the distribution function of the couple $(U;1-U)$ It is called the maximum copula or anticomonotonic copula.

Note that m is a copula; however, M is a copula only in the case where $d=2$. Another special case is that of independent variables: the associated copula is the copula denoted by

$$\forall (u,v) \in I^2 : \quad \pi(u,v) = uv,$$

called the independent copula or product copula.

2.2.2 Elliptique copula

Elliptical copulas are copulas associated with elliptical distributions. Any copula that is written in the following form is called an elliptical copula.

$$C_l(u,v) := \frac{1}{\sqrt{1-l^2}} \int_{-\infty}^{\phi_{g,1}^{-1}(u)} \int_{-\infty}^{\phi_{g,2}^{-1}(v)} g\left(\frac{x^2-2lxy+y^2}{\sqrt{1-l^2}}\right) dxdy = H_l\left(\phi_{g,1}^{-1}(u), \phi_{g,2}^{-1}(v)\right).$$

H is the joint distribution of the random variables X and Y $\phi_{g,1}^{-1}(u), \phi_{g,2}^{-1}(v)$ their respective quantile functions and their correlation coefficients are l. In this family, among others, are the Gaussian copula and the Student's copula.

2.2.3 Gaussian copula

This copula does not allow for measuring dependence between the tails of marginal distributions. This is a limiting property when assessing dependence between rare events. One of the most commonly used types of copulas in modeling is the bivariate normal copula.

It is the most frequently used copula, and it is defined by

$$C_l\left(u,v\right):=\frac{1}{2\pi\sqrt{1-l^2}}\int\limits_{-\infty}^{\phi^{-1}(u)}\int\limits_{-\infty}^{\phi^{-1}(v)}\exp\left(-\frac{x^2-2lxy+y^2}{2\left(1-l^2\right)}\right)dxdy,$$

where ϕ^{-1} is the quantile function of the standard normal distribution $N(0;1)$: We find the following particular cases as limit cases: $C_{-1}=M$, $C_0=\pi$ and $C_1=m$.

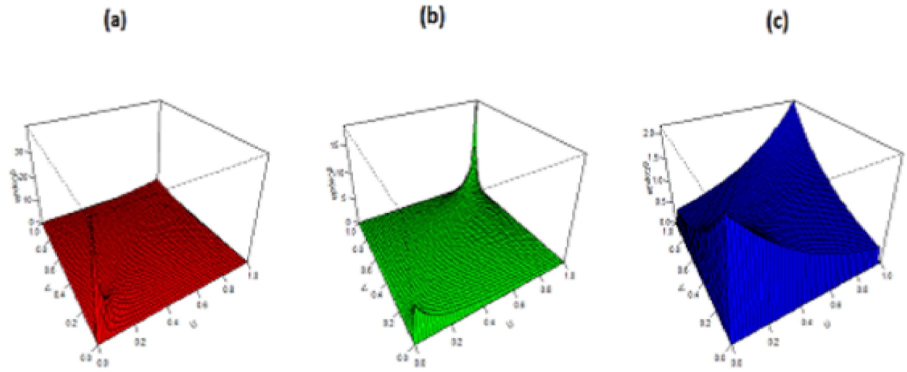

Figure 2.1: Density C_l of three Gaussian copulas according to ł

2.2.4 Student copula

Compared to the Gaussian copula, the Student's copula allows, thanks to its degree of freedom, to better account for thick tails of the distribution. Furthermore, as the degree of freedom tends towards infinity, the Student's copula is equal to the Gaussian copula. The Student's copula is defined as follows:

$$C_{l,v}\left(u,v\right):=\frac{1}{2\pi\sqrt{1-l^2}}\int\limits_{-\infty}^{T_v^{-1}(u)}\int\limits_{-\infty}^{T_v^{-1}(v)}\left(1+\frac{x^2-2lxy+y^2}{v\left(1-l^2\right)}\right)^{-\left(\frac{v}{2}+1\right)}dxdy,$$

Archimedean copula

The class of Archimedean copulas, defined by [19], plays a very important role. On one hand, they allow for the construction of a wide variety of copula families, thus representing a broad range of dependency structures. On the other hand, the copulas generated in this way have closed analytical forms and are easy to simulate. Indeed, unlike Gaussian copulas and Student copulas, Archimedean copulas have the great advantage of describing very diverse dependency structures, including so-called asymmetric dependencies, where the lower and upper tail coefficients differ. For further details on this family of copulas, the reader may refer to the excellent book by Nelsen .

Several reasons justify the use of this type of copulas, among others:

The wide variety of parametric families.

The particular and interesting properties that this class possesses.

The broad variety of different dependency structures.

The ease with which they can be constructed and simulated.

Archimedean copulas were introduced and developed by Professor [19] from Laval University in Quebec. A significant number of families belonging to him and possessing interesting properties will be mentioned in Table 2.4.

Definition 2.1

We call an Archimedean copula with generator the copula given by

$$C\left(u,v\right)=\phi^{-1}\left(\phi\left(u\right)+\phi\left(v\right)\right)$$

where $\phi : I \to [0;\infty[$ *a continuous and strictly decreasing function satisfying* $\phi\left(1\right)=0$
We define the inverse of ϕ *by* $\phi^{[-1]}$ *as*

$$\phi^{[-1]}\left(t\right)=\begin{cases}\phi^{-1}\left(t\right) & \text{if} \quad 0\leq t\leq\phi\left(0\right), \\ 0 & \text{if} \quad \phi\left(0\right)\leq t\leq\infty.\end{cases}$$

if $\phi\left(0\right)=\infty$, *so* $\phi^{[-1]}=\phi^{-1}$
ϕ *is at least twice continuously differentiable such that* $\phi'\left(u\right)<0$ *and* $\phi''\left(u\right)>0$ *for all* $u\in I$.

Remark 2.4

Another characterization of Archimedean copulas can be done using the Kendall's tau function

$$K(s) = P(C(U,V) \leq s) = 1 - \frac{\phi(s)}{\phi'(s)}; s \in I.$$

The table 2.4 presents some classic Archimedean families.

Copula	$\phi_\theta(t)$	$C_\theta(u,v)$
Clayton	$\left(t^{-\theta}-1\right)/\theta$	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}; \theta \in [-1;0[\cup]0,\infty[$
Gumbel	$(-\ln t)^\theta$	$\exp\left\{-\left[(-\ln u)^\theta+(-\ln v)^\theta\right]^{1/\theta}\right\}; \theta \geq 1$
Frank	$-\ln\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$\frac{-1}{\theta}\ln\left(1+\frac{(\exp((-\theta u)-1))(\exp((-\theta v)-1))}{\exp(-\theta)-1}\right); \theta \in \mathbb{R}$
Joe	$-\ln\left(1-(1-t)^\theta\right)$	$1-\left[(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta\right]^{1/\theta}; \theta \geq 1$
AMH	$\ln\left(\frac{1-\theta(1-t)}{t}\right)$	$uv/(1-\theta(1-u)(1-v))$

The obtained results for the estimator $\widehat{C}_n^1$ under weak dependence.

2.2.5 Emprical Copula

La notion of empirical copula was introduced by [13] based on the empirical versions of the distribution functions. One can construct the empirical copula $\hat{C}_n$ based on the empirical marginals F_n, G_n, and H_n as follows:

$$\hat{C}_n(u,v) = H_n\left(F_n^{-1}(u), G_n^{-1}(v)\right),$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}}$and $G_n(y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i \leq y\}}$.

In the case of a bivariate distribution by a noted sample $\{(x_i, y_i)\}_{1 \leq i \leq n}$ for $1 \leq i \leq n$, it is written as follows:

$$H_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x, Y_i \leq y\}}$$

The bivariate empirical copula denoted by $\hat{C}_n$ is defined on the set $\mathcal{L} = \left\{\left(\frac{j}{n}, \frac{k}{n}\right); j,k = 1,2,...,n\right\}$ The emprical copula $\hat{C}_n$ is

$$\hat{C}_n\left(\frac{j}{n},\frac{k}{n}\right)=\frac{1}{n}\sum_{i=1}^{n}I_{\{X_i\leq x(j),Y_i\leq y(k)\}}$$

where $x(j)$ and $y(k)$ represent the rank statistics of the sample $(X_1, X_2, ..., X_n)$ and $(Y_1, Y_2, ..., Y_n)$ respectively.

The empirical copula can also be written as follows:

$$\hat{C}_n(u,v)=\begin{cases}\hat{C}_n\left(\frac{j-1}{n},\frac{k-1}{n}\right) & \text{if } \frac{j-1}{n}\leq u<\frac{j}{n}\text{and } \frac{k-1}{n}\leq v<\frac{k}{n}\\ 1 & \text{if } u=1\text{and } v=1\end{cases}$$

The empirical density function of the copula $\hat{C}_n$ sometimes called the empirical frequency copula and denoted by $\hat{c}_n$ is given by

$$\hat{c}_n\left(\frac{j}{n},\frac{k}{n}\right)=\begin{cases}\frac{1}{n} & \text{if } \left(x_{(j)},y_{(k)}\right)\in\{(x_i,y_i)\}_{1\leq i\leq n},\\ 0 & \text{ifelse}\end{cases}$$

There exists a relationship between $\hat{C}_n$ and $\hat{c}_n$ given by $\hat{C}_n\left(\frac{j}{n},\frac{k}{n}\right)=\sum_{p=1}^{j}\sum_{q=1}^{k}\hat{c}_n\left(\frac{p}{n},\frac{q}{n}\right)$ and $\hat{c}_n\left(\frac{j}{n},\frac{k}{n}\right)=\hat{C}_n\left(\frac{j}{n},\frac{k}{n}\right)-\hat{C}_n\left(\frac{j-1}{n},\frac{k}{n}\right)-\hat{C}_n\left(\frac{j}{n},\frac{k-1}{n}\right)+\hat{C}_n\left(\frac{j-1}{n},\frac{k-1}{n}\right)$.

Chapter 3

Censoring Notion

In actuality, it's not always possible to obtain a sample with all the data.In statistics, censorship is one of the most common events that leads to incomplete data. A data is said to be "censored" if the exact value is unknown, but only an estimate, lower or higher, in other words, approximate details of the type $T \geq C$ or $T \leq C$. Such information is very poor, poorer than saying "T is between a and b ", since only one of the two bounds is known. In the analysis of survival times, censoring occurs when the survival T is only known for some of the individuals "the data for which survival is unknown are said censored". The variable of interest T is not observed and it is limited superially or inferiorly by a variable (of censoring, generally noted C) which has been observed.Given that in biostatistics and epidemiology, the main focus of the studies is the explanation for the occurrence of an event of interest (death, rejection of a transplant, end of study, withdrawal from study, loss of follow-up, etc), all available information must be analyzed. However, due to the fact that the phenomenon of censoring is in itself a special case in completeness of the data, observational studies only very rarely present complete data when within a framework of survival analysis. Thus, it is necessary, for the clinician be quick to use statistical methods that take into account the censored data.

In addition, censoring can be informative or non-informative: in the event of censoring informative, there is a dependence between the survival time and the censoring time. We take the example of a patient lost to follow-up: his voluntary withdrawal may, for example, result from the fact that the patient is near death or decides to stop treatment to die in a certain time dignity, its censoring is then dependent on the time of death.

For an individual i, we consider:

- its survival time T_i
- its censoring time C_i
- the time actually observed X_i

3.0.1 Type of censorship

Right censoring

Right censoring is discussed when observing censoring C (and not the lifetime T) and we know that $T > C$. This model is most common in practice, for example, it is adapted to the case where the event of interest is the survival time of a disease and where the end date of the study is predetermined; patients alive at the end of the study provide right-censored data.

Left censoring

Left censoring corresponds to the case where the individual has already experienced the event before being observed. We only know that the event of interest occurred before a certain known date, $T < C$. An example of such a situation is when an electronic component is mounted in parallel with one or more other components. Failure of this component does not necessarily stop the system: the system may continue to operate until this failure is detected (for example, during inspection or in the event of system shutdown). The observed duration for this component is then left-censored.

Double censoring

In the same sample, one can find data that are right-censored and others that are left-censored. For example, in a study focusing on the age at which children learn to perform certain tasks, at the beginning of the study, some children already knew how to perform the tasks under investigation. We only know then that the age at which they learned is younger than their age at the start of the study. At the end of the study, some children still could not perform these tasks, and we only know then that the age at which they will eventually learn is older than their age at the end of the study. The age at the beginning of the study (left-censoring variable L) is obviously younger than the age at the end of the study (right-censoring variable C). The age of interest is observed if it falls within the study period.

Interval censoring

A date is interval-censored if instead of observing the exact time of the event, the only available information is that it occurred between two known dates $C_1 \leq T \leq C_2$. This occurs, for example, when a patient visits the hospital regularly: if they miss an appointment, the only information available is that their death occurred within the interval between the last visit and the appointment.

Mixed censoring

It is said that there is mixed censoring when two censorship phenomena (one on the left and the other on the right) can prevent the observation of the phenomenon of interest without necessarily being able to determine an interval to which it belongs. Instead of observing a sample of the variable of interest Y, we observe a sample of the pair $(Z; A)$ with

$$Z = max(min(T; C); L)$$

and

$$A = \begin{cases} 0 & \text{if } \ L < T \leq C, \\ 1 & \text{if } \ L < C < T, \\ 2 & \text{if } \ \min(T, C) \leq L. \end{cases}$$

where L and C are censoring variables and A is the censorship indicator.

3.0.2 Non-parametric estimation for right-censoring model

In the case of right-censoring, the empirical survival function of the vari able T is no longer valid because since it involves unobserved quantities.

In particular, estimating the distribution of a duration censored by the empirical distribution function was impossible. In order to estimate the T distribution, it was necessary to construct a survival function estimator in the presence of censored data. The non-parametric estimation problem of a right-censored random variable distribution function, was originally considered by Kaplan and Meier (1958) . They provide a good estima tor of the survival function $S_T(t) = 1 - F_T(t)$, having the following

$$\hat{S}_n(t) = \prod_{j/Z'_j \leq t} \left(1 - \frac{M\left(Z'_j\right)}{N\left(Z'_j\right)}\right)$$

$\left(Z'_j\right)_{1 \leq j \leq M}$ $(M \leq n)$ are the distinct values of $Z_i = \min(T_i, C_i)$ ar-ranged in ascending order.

$M\left(Z'_j\right) = \sum_{i=1}^{n} \delta_i I_{\left\{Z_i = Z'_j\right\}}$ is the exact number of deaths at the moment Z'_j.

$N\left(Z'_j\right) = \sum_{i=1}^{n} I_{\left\{Z_i \geq Z'_j\right\}}$ is the number of individuals at risk just before the moment Z'_j.

This estimator coincides with the empirical distribution function when there are no censored data. Therefore, it is natural for statisticians to have been interested in extending the known results for the empirical distribution function to the case of the **Kaplan-Meier** estimator. The almost sure uniform convergence and the law of the iterated logarithm were respectively shown by **Winter et al**. (1978) and **Földes and Rejtő** (1981a). Then, Stute and Wang (1993) showed the strong law of large numbers in the case of right censoring, which leads, among other things, to the almost sure convergence of F_n^R. As for almost complete convergence, it was shown by **Földes et al.** (1980), with a convergence rate of the order of p log $n/\sqrt{n}$. Then, by imposing the continuity of F and F_R, **Földes and Rejtő** (1981b) improved the convergence rate to the order of p log n/n. Furthermore, **Kitouni et al.** (2015) showed that it is possible to dispense with this continuity assumption by using the following exponential bound.

Theorem 3.1

There exists an absolute constant D such that, for any positive real number u,

$$P\left(\sqrt{n} \sup_{x \in \mathbb{R}} S_R(x) \left|F_n(x) - F(x)\right| > u\right) \leq 2.5 \exp\left(-2u^2 + Du\right)$$

■

Proof:

See Bitouzé et al. (1999).

Regarding weak convergence, Breslow and Crowley (1974) provided the following result.

3.0.3 Non-parametric estimation for mixed censoring model

A new class of estimators is to be presented when the observations T_i are subjected to a censoring mechanism, this model is carried on the non parametric estimate and discussed by Patilea and Rolin (2006) .

The Patilea and Rolin Estimator

Assuming that we have observed a sample $(Z_i;\delta_i)_{1\leq i\leq n}$ of the pair (Z,δ) where $Z = (T\wedge C)\vee L = \max(X,L)$, for $X = (T\wedge C)$ and T, C, L are positive and independent random variables representing respectively the variable of interest, the left-censored variable, and the right-censored variable.

Let H be the distribution function of Z and $H^{(0)}$ its sub-distribution for uncensored observations having the following expressions:

$$H(t) = P(Z\leq t) = F_L(t)F_X(t)F_C(t)(1-S_{T(t)}S_C(t)),$$

and

$$H^{(0)}(t) = P(Z\leq t,\delta=0) = \int_0^t F_L(x)\,S_C(x)\,dF_T(x)$$

As well as their empirical versions are given respectively by:

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i\leq t\}},$$

and

$$H^{(0)}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i\leq t,\delta_i=0\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i\leq t, T_i-C_i\leq 0, L_i-T_i\leq 0\}}$$

We noted $Z_j^{'}\,(1\leq j\leq M)$ the distinct values of Z_i arranged in increasing order and for $k\in\{0,1,2\}$:

$$D_{kj} = \sum_{i=1}^{n} I\left\{Z_i = Z_j^{'},\delta_i = k\right\}$$

The non-parametric estimator, denoted by $\tilde{S}_n$, of S_T, is the bounded product estimator given by **Patilea and Rolin** (2006) by the form:

$$\tilde{S}_n(t) = 1 - \tilde{F}_n(t)$$
$$= \prod_{j/Z'_j \le t} \left(1 - \frac{D_{0j}}{\dot{F}_n\left(Z'_{j-1}\right) - nH_n\left(Z'_{j-1}\right)}\right)$$

where $\dot{F}_n$ is the Kaplan-Meier estimator of the distribution function F_L, defined by inverting time as:

$$\dot{F}_n(t) = \prod_{j/Z'_j \le t} \left(1 - \frac{D_{2j}}{nH_n\left(Z'_j\right)}\right).$$

3.0.4 Smooth estimators of the copula and its density

Let $k : \mathbb{R} \to \mathbb{R}$ be a kernel function (that is a smooth function with integral over $\mathbb{R}$ equal to one) and $K(x) = \int k(u)\,du$ its cumulative integral. Introducing a smoothing parame ter $h > 0$, a classical kernel estimator of the bivariate distribution function of multiplicative form is defined through a convolution of the empirical measure with the measure of density $h^{-2}k(u/h)\,k(v/h)$.

For the right censored data, we replace the empirical measure by the measure defined by the estimator

$$F_n(t_1, t_2) = \frac{1}{n}\sum_{i=1}^{n} W_{in} I_{Y_{1i} < t_1, Y_{2i} < t_2}.$$

This leads us to the following distribution function estimator

$$\hat{F}_n^1(t_1, t_2) = \frac{1}{n}\sum_{i=1}^{n} W_{in} K_h(t_1 - Y_{1i})\, K_h(t_2 - Y_{2i}),$$

where $K_h(x) := K(x/h)$. Let us introduce now a first smooth copula estimator given by

$$\hat{C}_n^1(u, v) = \hat{F}_n^1\left(\left(\hat{F}_{1n}^1\right)^{-1}(u), \left(\hat{F}_{2n}^1\right)^{-1}(v)\right) \tag{3.1}$$

In the complete data, this estimator was introduced and studied by Fermanian et al. (2004). An important drawback of estimator 3.1 is that its performance depends on marginal distribution functions of the variables T_1 an T_2 (this issue was extensively discussed in Omelka et al. (2009)). To get rid of this inconvenient, Omelka et al. (2009) proposed to use a

transformation of the initial variables by some d.f. Φ, designed to avoid corner bias problems.Their method can be extended to our framework, leading to a second smooth estimator C_n^2 of the copula function. Indeed, for some d.f. Φ, consider a couple of variables $.\left(\tilde{T}_1, \tilde{T}_2\right) = \left(\Phi^{-1}\left[F_1\left(T_1\right)\right], \Phi^{-1}\left[F_2\left(T_2\right)\right]\right)$ and pseudo-observations . $\left(\Phi^{-1}\left[F_{1n}\left(Y_{1i}\right)\right], \Phi^{-1}\left[F_{2n}\left(Y_{2i}\right)\right]\right)_{1\leq}$ the variables $\left(\tilde{T}_1, \tilde{T}_2\right)$ are coupled by the same copula as $.(T_1, T_2)$. Next, define an estimator of the joint d.f.of $\left(\tilde{T}_1, \tilde{T}_2\right)$ by

$$\hat{F}_n^2\left(t_1, t_2\right) = \frac{1}{n}\sum_{i=1}^{n} W_{in} K_h\left(t_1 - \Phi^{-1}\left[F_{1n}\left(Y_{1i}\right)\right]\right) K_h\left(t_2 - \Phi^{-1}\left[F_{2n}\left(Y_{2i}\right)\right]\right),$$

where F_{1n} (resp. F_{2n}) are marginal distributions of estimator 3.1 Then, define

$$\hat{C}_n^2\left(u, v\right) = \hat{F}_n^2\left(\Phi^{-1}\left(u\right), \Phi^{-1}\left(v\right)\right)$$

The smoothness of these estimators allows to deduce estimators of the copula density $c\left(u, v\right)$, let

$$\hat{c}^i\left(t_1, t_2\right) = \frac{\partial^2}{\partial t_1 \partial t_2} \hat{C}^i\left(t_1, t_2\right), \text{ for } i = 1, 2.$$

3.0.5 Semi-parametric estimation for Copula models

Maximum Likelihood Estimation (MLE)

Assuming a multivariate parametric copula C_θ, where $\theta = (\theta_1, ..., \theta_d) \in \ominus$ be the vector of copula parameters and 3b2 be the vector of marginal param eters. Given the relatively simple functional form the self-selection like lihood function under an Archimedean copula, MLE can be employed to jointly estimate all parameters of the unknown parameters vector $(\beta_1, ..., \beta_d, \theta)$ at the same time. Assume that we observe d-independent realizations $(X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, d$, specified by p-margins with cumu lative distribution function (CDF) F_i. However, the density of F is given by:

$$f(x_1, ..., x_d) = c_\theta\left[\left(F_{1,\beta}\left(x_1\right), ..., F_{d,\beta}\left(x_d\right)\right); \theta\right] \prod_{i=1}^{d} f_{i,\beta_i}\left(x_i\right) \quad (3.2)$$

That is associated with a sample $(X_{i1}, \ . \ . \ . \ , X_{id})$, $i = 1, \ . \ . \ . \ , d$, where c_θ is a den sity of a parametric copula C_θ and f_i,β_i

is a density of F_i,β_i. A parametric and a semi-parametric approaches both presented seek to maximize a like lihood approximation based on 3.2. Consequently, the parameter vector to be estimated in the parametric approach is $3b1 = (3b2, 3b8)$ and by maximiz ing the log-likelihood function $L(3b2_1, ..., 3b2_d; 3b8)$ defined by:

$$L(3b2_1, ..., 3b2_d; 3b8) = \sum_{i=1}^{n} \log f(x_1, ..., x_d; \theta)$$

$$\begin{aligned} L(3b2_1, ..., 3b2_d; 3b8) &= \log c_\theta \left[\left(F_{1,\beta}(x_1), ..., F_{d,\beta}(x_d) \right); \theta \right] \prod_{j=1}^{d} f_{j,\beta_j}(x_j) \\ &= \sum_{i=1}^{n} \log c_\theta \left[\left(F_{1,\beta}(x_1), ..., F_{d,\beta}(x_d) \right); \theta \right] + \sum_{i=1}^{n} \sum_{j=1}^{d} \log \prod_{j=1}^{d} f_{j,\beta_j}(x_j), \end{aligned}$$

then the estimator of θ, noted $\hat{\theta}_n^{MV}$ is

$$\hat{\theta}_{MLE} = \arg\max L(3b2_1, ..., 3b2_d; 3b8)$$

See Lehmann and Casella [56], for more details. This estimator is con sistent and satisfies the asymptotic normality property:

$$\sqrt{n} \left(\hat{\theta}_{MLE} - \theta \right) \to N(0, 1)$$

such that $I(\theta)$ is the Fisher information matrix. This matrix is estimated by the inverse of the Hessian matrix of the likelihood function.

Simulation of the Copula Function with Bivariate Twice Censored Data

4.1 Empirical copula for twice censored data

Let $X = (X_1, X_2)$ be a couple of positive r.r.v. with support $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and joint distribution function F, and let $R = (R_1, R_2)$ (resp. $L = (L_1, L_2)$) be a couple of positive right (resp. left) censoring variables. We assume that the variables X, L and R are independent. In the twice censoring framework, instead of observing X, we observe the independent copies $(Z_{1i}, Z_{2i}, A_{1i}, A_{2i})_{1\leq i\leq n}$ of the vector (Z_1, Z_2, A_1, A_2), where for $k \in \{1, 2\}$, $Z_k = \max(\min(X_k, R_k), L_k)$ and A_k is the indicator of censoring given by

$$A_k = \begin{cases} 0 & \text{if } L_k < X_k \leq R_k, \\ 1 & \text{if } L_k < R_k < X_k, \\ 2 & \text{if } \min(X_k, R_k) \leq L_k. \end{cases}$$

In all the sequel, for any r.r.v. V, F_V, S_V, I_V and T_V denote, respectively, the distribution function, the survival function, the lower and the upper endpoint of the support of V. Moreover, for any right continuous function $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$, we set $\varphi(t^-) = \lim_{\varepsilon \overset{>}{\to} 0} \varphi(t - \varepsilon)$ the left-hand limit of φ at t when it exists. Furthermore, for any differentiable function $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$, we denote by $\partial_1\psi$ (resp. $\partial_2\psi$) the partial derivative of ψ with respect to its first (resp. second) variable.

We assume that the copula function C of X is twice continuously differentiable on $[0, 1]^2$. Furthermore, following [35], we assume that the copula function C_L of L and the survival copula function [1] $\widetilde{C}_R$ of R are known and twice continuously differentiable on $[0, 1]^2$. We also assume that the functions F_{X_k}, F_{R_k} and F_{L_k} $(k \in \{1, 2\})$ are continuous.

To define the empirical copula $C_n(u, v)$ for $(u, v) \in [0, 1]^2$, we need to introduce the following notations. For $k \in \{1, 2\}$ and $j \in \{0, 1, 2\}$, denote by $H_k^{(j)}(t) = P(Z_k \leq t, A_k =$

[1] The survival copula $\widetilde{C}_R$ of R is defined by $\widetilde{C}_R(u, v) = u + v - 1 + C_R(1 - u, 1 - v)$, where C_R is the copula function of R.

j) the sub-distribution function of Z_k for $A_k = j$, $I_{H_k^{(j)}} = \inf\{t \in \mathbb{R}/H_k^{(j)}(t) > 0\}$ the lower endpoint of the support of $H_k^{(j)}$ and

$$H_{nk}^{(j)}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_{ki}\leq t, A_{ki}=j\}} \quad \text{and} \quad \widehat{F}_{Z_k}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_{ki}\leq t\}}$$

($I_{\{.\}}$ being the indicator function) the empirical versions of $H_k^{(j)}$ and F_{Z_k}, respectively. Furthermore, denote by $\left(Z'_{kj}\right)_{1\leq j\leq m}$ ($m \leq n$) the distinct values of $(Z_{ki})_{1\leq i\leq n}$. The product-limit estimator $\widehat{F}_{L_k}$ of F_{L_k} is defined by

$$\widehat{F}_{L_k}(t) = \prod_{j/Z'_{kj}>t}\left(1 - \frac{\sum_{i=1}^{n} I_{\left\{Z_{ki}=Z'_{kj}, A_{ki}=2\right\}}}{n\widehat{F}_{Z_k}\left(Z'_{kj}\right)}\right),$$

this estimator can be derived from the Kaplan-Meier one by reversing time.

In addition, the product-limit estimator of S_{R_k} is given by [43] as follows.

$$\widehat{S}_{R_k}(t) = \prod_{i/Z_{ki}\leq t}\left(1 - \frac{I_{\left\{A_{k_i}=1\right\}}}{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)}\right).$$

Since X is not observed, the empirical distribution function

$$\widetilde{F}_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{1i}\leq x_1, X_{2i}\leq x_2\}}$$

can not be used to estimate $F(x_1, x_2)$. So, following [35] and remarking that

$$E\left[g(Z_1, Z_2)\, I_{\{A_1=0\}} I_{\{A_2=0\}} I_{\{Z_1\leq x_1, Z_2\leq x_2\}}\right] = E\left[I_{\{X_1\leq x_1, X_2\leq x_2\}}\right] = F(x_1, x_2),$$

where $g(z_1, z_2) = C_L\left(F_{L_1}(z_1), F_{L_2}(z_2)\right)^{-1} \widetilde{C}_R\left(S_{R_1}(z_1), S_{R_2}(z_2)\right)^{-1}$, we propose to replace $I_{\{X_{1i}\leq x_1, X_{2i}\leq x_2\}}$ by the observed quantity

$$\widehat{g}(Z_{1i}, Z_{2i}) I_{\{A_{1i}=0\}} I_{\{A_{2i}=0\}} I_{\{Z_{1i}\leq x_1, Z_{2i}\leq x_2\}},$$

where $\widehat{g}(z_1, z_2) = C_L\left(\widehat{F}_{L_1}(z_1), \widehat{F}_{L_2}(z_2)\right)^{-1} \widetilde{C}_R\left(\widehat{S}_{R_1}(z_1), \widehat{S}_{R_2}(z_2)\right)^{-1}$.

This gives the following estimator of $F(x_1, x_2)$

$$F_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_{\{A_{1i}=0\}} I_{\{A_{2i}=0\}}}{C_L\left(\widehat{F}_{L_1}(Z_{1i}), \widehat{F}_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(\widehat{S}_{R_1}(Z_{1i}), \widehat{S}_{R_2}(Z_{2i})\right)} I_{\{Z_{1i}\leq x_1, Z_{2i}\leq x_2\}}.$$

Using this estimator, we propose to estimate $C(u, v)$ as in [23] (relation (3.3)) by

$$C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_{\{A_{1i}=0\}} I_{\{A_{2i}=0\}}}{C_L\left(\widehat{F}_{L_1}(Z_{1i}), \widehat{F}_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(\widehat{S}_{R_1}(Z_{1i}), \widehat{S}_{R_2}(Z_{2i})\right)} I_{\{F_{n1}(Z_{1i})\leq u,\, F_{n2}(Z_{2i})\leq v\}},$$

where $F_{n1}(x_1) = \lim_{x_2\to\infty} F_n(x_1, x_2)$ and $F_{n2}(x_2) = \lim_{x_1\to\infty} F_n(x_1, x_2)$.

4.2 Main results

In this section, we establish the weak convergence of the processes $\sqrt{n}\left(F_n(x_1,x_2)-F(x_1,x_2)\right)$, $(x_1,x_2)\in\mathcal{X}$ and $\sqrt{n}\left(C_n(u,v)-C(u,v)\right)$, $(u,v)\in[0,1]^2$. Our approach will be based, as in [23], on a representation of F_n-F as a sum of i.i.d. random variables. So, we will first establish this representation. For that, we need to represent $\widehat{g}-g$ as a sum of i.i.d. random variables. In order to prove such a representation, we begin by introducing some assumptions and notations. For $k\in\{1,2\}$, denote by $\mathbb{S}_k=\{z\in\mathbb{R}:I_{H_k^{(1)}}<z<\tau_k\}$, where τ_k is such that $I_{H_k^{(1)}}<\tau_k<T_{Z_k}$ and let $\mathbb{S}=\mathbb{S}_1\times\mathbb{S}_2$. We assume that

H1 $I_{L_k}<I_{R_k}$ and $T_{L_k}<T_{R_k}\leq T_{X_k}$.

H2 There exist $\theta_{k1}>I_{R_k}$ and $\theta_{k2}<T_{R_k}$ such that

$$\forall n\in\mathbb{N}^*,\ \forall 1\leq i\leq n: A_{ki}=1\Rightarrow \theta_{k1}\leq Z_{ki}\leq\theta_{k2}\quad \text{almost surely } (a.s.).$$

H3 $\displaystyle\int_{I_{H_k^{(1)}}}^{+\infty}\frac{dH_k^{(2)}(z)}{\left(F_{Z_k}(z)\right)^2}<+\infty.$

Assumptions **H1** and **H2** are standard in the twice censoring setting (see e.g. [43], [38] and [30]). Assumption **H3** is needed to obtain the weak convergence of $\sqrt{n}\left(\widehat{F}_{L_k}-F_{L_k}\right)$ and $\sqrt{n}\left(\widehat{S}_{R_k}-S_{R_k}\right)$ on $\mathbb{S}_k$ (see [[43], Lemma 7.2 and Theorem 7.3]). This weak convergence ensures that $\sup_{z\in\mathbb{S}_k}\left|\widehat{F}_{L_k}(z)-F_{L_k}(z)\right|=O_P\left(n^{-1/2}\right)$ and $\sup_{z\in\mathbb{S}_k}\left|\widehat{S}_{R_k}(z)-S_{R_k}(z)\right|=O_P\left(n^{-1/2}\right)$ [2]. So, as in [35], we can use a Taylor expansion to get

$$\begin{aligned}\widehat{g}(z_1,z_2)-g(z_1,z_2)=-\sum_{k=1,2}\Bigg(&C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)\frac{\partial_k\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)}{\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)^2}\left(\widehat{S}_{R_k}(z_k)-S_{R_k}(z_k)\right)\\&+\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)\frac{\partial_k C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)}{C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)^2}\left(\widehat{F}_{L_k}(z_k)-F_{L_k}(z_k)\right)\Bigg)\\&+r_n(z_1,z_2),\end{aligned}\tag{4.1}$$

where $\sup_{(z_1,z_2)\in\mathbb{S}}\left|r_n(z_1,z_2)\right|=o_P\left(n^{-1/2}\right)$.

It remains to represent $\widehat{F}_{L_k}-F_{L_k}$ and $\widehat{S}_{R_k}-S_{R_k}$ as a sum of i.i.d. random variables. The representation of $\widehat{F}_{L_k}-F_{L_k}$ can be deduced from [[36], Theorem 1] by reversing time. In

[2]For a sequence of r.r.v. (ζ_n) and a sequence of non-zero real numbers (u_n), $\zeta_n=O_p(u_n)$ means that $\dfrac{\zeta_n}{u_n}$ is bounded in probability and $\zeta_n=o_p(u_n)$ means that $\dfrac{\zeta_n}{u_n}$ converges in probability to zero.

fact, we get for $\delta \in]0,1[$, $I \in \mathbb{R}$ such that $F_{L_k}(I) > \delta$ and $u > I_{Z_k}$

$\widehat{F}_{L_k}(u) - F_{L_k}(u) = F_{L_k}(u)\left(A_{L_k}(n,u) + B_{L_k}(n,u)\right) + R_{L_k}(n,u)$, where

$$A_{L_k}(n,u) = -\frac{H_{nk}^{(2)}(u) - H_k^{(2)}(u)}{F_{Z_k}(u)} - \int_u^{+\infty} \frac{H_{nk}^{(2)}(y) - H_k^{(2)}(y)}{\left(F_{Z_k}(y)\right)^2} dF_{Z_k}(y),$$

$$B_{L_k}(n,u) = \int_u^{+\infty} \frac{\widehat{F}_{Z_k}(y) - F_{Z_k}(y)}{\left(F_{Z_k}(y)\right)^2} dH_k^{(2)}(y)$$

and $R_{L_k}(n,u)$ satisfies

$$\sup_{u \geq I} \left|R_{L_k}(n,u)\right| = O_p\left(\frac{1}{n}\right). \tag{4.2}$$

So,

$$\begin{aligned} \widehat{F}_{L_k}(u) - F_{L_k}(u) = \frac{1}{n}\sum_{i=1}^{n} \Bigg[& -\frac{F_{L_k}(u)}{F_{Z_k}(u)}\left(I_{\{Z_{ki} \leq u, A_{ki}=2\}} - H_k^{(2)}(u)\right) \\ & - F_{L_k}(u) \int_u^{+\infty} \frac{I_{\{Z_{ki} \leq y, A_{ki}=2\}} - H_k^{(2)}(y)}{\left(F_{Z_k}(y)\right)^2} dF_{Z_k}(y) \\ & + F_{L_k}(u) \int_u^{+\infty} \frac{I_{\{Z_{ki} \leq y\}} - F_{Z_k}(y)}{\left(F_{Z_k}(y)\right)^2} dH_k^{(2)}(y) \Bigg] + R_{L_k}(n,u) \end{aligned} \tag{4.3}$$

which is a representation of $\widehat{F}_{L_k} - F_{L_k}$ as a sum of i.i.d. centered random variables. Regarding $\widehat{S}_{R_k} - S_{R_k}$, we give its representation in the following lemma.

Lemma 4.1

*Assume that assumptions **H1-H3** hold and let $\delta \in]0,1[$, I, $T \in \mathbb{R}$ such that $F_{L_k}(I)$ $S_{X_k}(T)\, S_{R_k}(T) > \delta$. We have*

$$\widehat{S}_{R_k}(u) - S_{R_k}(u) = S_{R_k}(u)\left(A_k(n,u) + B_k(n,u)\right) + R_k(n,u),$$

where

$$A_k(n,u) = -\frac{H_{nk}^{(1)}(u) - H_k^{(1)}(u)}{F_{L_k}(u) - F_{Z_k}(u)} + \int_0^{u} \frac{H_{nk}^{(1)}(y) - H_k^{(1)}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} d\left(F_{L_k}(u) - F_{Z_k}(u)\right),$$

$$\begin{aligned} B_k(n,u) = \frac{1}{n}\sum_{i=1}^{n} \int_0^{u} \Bigg\{ & \frac{F_{L_k}(y)}{F_{Z_k}(y)}\left(I_{\left\{Z_{ki} < y, A_{ki}=2\right\}} - H_k^{(2)}(y)\right) + F_{L_k}(y) \int_y^{+\infty} \frac{I_{\{Z_{ki} \leq t, A_{ki}=2\}} - H_k^{(2)}(t)}{\left(F_{Z_k}(t)\right)^2} dF_{Z_k}(t) \\ & - F_{L_k}(y) \int_y^{+\infty} \frac{I_{\left\{Z_{ki} \leq t\right\}} - F_{Z_k}(t)}{\left(F_{Z_k}(t)\right)^2} dH_k^{(2)}(t) \Bigg\} \frac{dH_k^{(1)}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} + \int_0^{u} \frac{\widehat{F}_{Z_k}(y^-) - F_{Z_k}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} dH_k^{(1)}(y), \end{aligned}$$

and $R_k(n,u)$ *satisfies* $\sup_{I\le u\le T}|R_k(n,u)| = O_P\left(\frac{1}{n}\right)$.

From this lemma, we deduce that

$$\begin{aligned}
\widehat{S}_{R_k}(u) - S_{R_k}(u) = \frac{S_{R_k}(u)}{n}\sum_{i=1}^{n}\Bigg[&\frac{1}{F_{L_k}(u)-F_{Z_k}(u)}\left(H_k^{(1)}(u) - I_{\left\{Z_{ki}\le u, A_{ki}=1\right\}}\right)\\
&+\int_0^u \frac{I_{\{Z_{ki}\le y, A_{ki}=1\}} - H_k^{(1)}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2} d\left(F_{L_k}(y)-F_{Z_k}(y)\right)\\
&+\int_0^u \Bigg\{\frac{F_{L_k}(y)}{F_{Z_k}(y)}\left(I_{\{Z_{ki}<y, A_{ki}=2\}} - H_k^{(2)}(y)\right)\\
&+F_{L_k}(y)\int_y^{+\infty}\frac{I_{\{Z_{ki}\le t, A_{ki}=2\}} - H_k^{(2)}(t)}{\left(F_{Z_k}(t)\right)^2} dF_{Z_k}(t)\\
&-F_{L_k}(y)\int_y^{+\infty}\frac{I_{\{Z_{ki}\le t\}} - F_{Z_k}(t)}{\left(F_{Z_k}(t)\right)^2} dH_k^{(2)}(t)\Bigg\}\frac{dH_k^{(1)}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}\\
&+\int_0^u \frac{I_{\{Z_{ki}<y\}} - F_{Z_k}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2} dH_k^{(1)}(y)\Bigg] + R_k(n,u) \qquad (4.4)
\end{aligned}$$

which is a representation of $\widehat{S}_{R_k} - S_{R_k}$ as a sum of i.i.d. centered random variables. Relations (4.1), (4.3) and (4.4) permit to write

$$\widehat{g}(z_1,z_2) - g(z_1,z_2) = \frac{1}{n}\sum_{i=1}^{n}\rho(Z_{1i},Z_{2i},A_{1i},A_{2i};z_1,z_2) + \widetilde{r}_n(z_1,z_2), \qquad (4.5)$$

where

$$\begin{aligned}
\rho(Z_{1i},Z_{2i},A_{1i},A_{2i};z_1,z_2) = -\sum_{k=1,2}\Bigg[&\Bigg(S_{R_k}(u)\,C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)\frac{\partial_k\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)}{\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)^2}\Bigg)\\
\times\Bigg(&\frac{1}{F_{L_k}(u)-F_{Z_k}(u)}\left(H_k^{(1)}(u) - I_{\left\{Z_{ki}\le u, A_{ki}=1\right\}}\right) + \int_0^u \frac{I_{\{Z_{ki}\le y, A_{ki}=1\}} - H_k^{(1)}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2} d\left(F_{L_k}(y)-F_{Z_k}(y)\right)\\
&+\int_0^u \Bigg\{\frac{F_{L_k}(y)}{F_{Z_k}(y)}\left(I_{\{Z_{ki}<y, A_{ki}=2\}} - H_k^{(2)}(y)\right) + F_{L_k}(y)\int_y^{+\infty}\frac{I_{\{Z_{ki}\le t, A_{ki}=2\}} - H_k^{(2)}(t)}{\left(F_{Z_k}(t)\right)^2} dF_{Z_k}(t)\\
&-F_{L_k}(y)\int_y^{+\infty}\frac{I_{\{Z_{ki}\le t\}} - F_{Z_k}(t)}{\left(F_{Z_k}(t)\right)^2} dH_k^{(2)}(t)\Bigg\}\frac{dH_k^{(1)}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2} + \int_0^u \frac{I_{\{Z_{ki}<y\}} - F_{Z_k}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2} dH_k^{(1)}(y)\Bigg)\\
&+\widetilde{C}_R\left(S_{R_1}(z_1),S_{R_2}(z_2)\right)\frac{\partial_k C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)}{C_L\left(F_{L_1}(z_1),F_{L_2}(z_2)\right)^2}\times\Bigg(-\frac{F_{L_k}(u)}{F_{Z_k}(u)}\left(I_{\{Z_{ki}\le u, A_{ki}=2\}} - H_k^{(2)}(u)\right)\\
&-F_{L_k}(u)\int_u^{+\infty}\frac{I_{\{Z_{ki}\le y, A_{ki}=2\}} - H_k^{(2)}(y)}{\left(F_{Z_k}(y)\right)^2} dF_{Z_k}(y) + F_{L_k}(u)\int_u^{+\infty}\frac{I_{\{Z_{ki}\le y\}} - F_{Z_k}(y)}{\left(F_{Z_k}(y)\right)^2} dH_k^{(2)}(y)\Bigg)\Bigg]
\end{aligned}$$

and $\sup_{(z_1,z_2)\in\mathbb{S}} |\widetilde{r}_n(z_1,z_2)| = o_P(n^{-1/2})$.

Note that ρ satisfies assumption 2 of [35]. In fact, it is not difficult to check that ρ is centered and that is uniformly bounded on $\mathbb{S}$ under **H1**. Moreover, one can proceed as in Lemma 7.3. of [35] to show that there exists a Donsker class of functions $\mathcal{G}$ such that the function $\frac{1}{n}\sum_{i=1}^{n}\rho(Z_{1i}, Z_{2i}, A_{1i}, A_{2i}; z_1, z_2)$ belongs to $\mathcal{G}$.

Using relation (4.5), we will represent $F_n - F$ as a sum of i.i.d. centered random variables. For that, we need to introduce more notations and assumptions. For any nonempty set A, we denote by $l^\infty(A)$ the space of all bounded real-valued functions defined on A. Moreover, For $k \in \{1,2\}$, denote by

$$\widehat{\Lambda}_{R_k}(t) = \int_0^t \frac{dH_{nk}^{(1)}(u)}{\widehat{F}_{L_k}(u^-) - \widehat{F}_{Z_k}(u^-)}$$

the estimator of the cumulative hazard function Λ_{R_k} of R_k. Thanks to [[43], Theorem 7.3], the process $\sqrt{n}\left(\widehat{\Lambda}_{R_k}(t) - \Lambda_{R_k}(t)\right)$, $t \in \mathbb{S}_k$, converges weakly, under **H1** and **H3** to a centered Gaussian process G_{R_k}. For $s, t \in \mathbb{S}_k$ such that $s \le t$, we denote by

$$\mathcal{K}_{R_k}(s) = cov\left(G_{R_k}(s), G_{R_k}(t)\right)$$

and by

$$\mathcal{K}_{L_k}(s) = \int_s^{+\infty} \frac{dF_{L_k}(u)}{F_{L_k}^2(u)\, F_{X_k\wedge R_k}(u)}$$

the covariance function of the limiting process of the Nelson-Aalan estimator of the cumulative hazard function of L_k (see [[11], Theorem 4] in reversing time). To prove our claimed result, we need the following assumptions which correspond to assumptions 3-5 of [35] adapted to the twice censored data model.

H4 The first and the second partial derivatives of C_L and $\widetilde{C}_R$ are bounded on $[0,1]^2$. Moreover, $C_L(x_1,x_2) \neq 0$ and $\widetilde{C}_R(x_1,x_2) \neq 0$ for $x_1 \neq 0$ and $x_2 \neq 0$.

H5 For $k \in \{1,2\}$, there exist $0 \leqslant \zeta_k \leqslant 1$ and $0 \leqslant_k \leqslant 1$ such that2

$$C_L(x_1,x_2) \geqslant x_1^{\alpha_1} x_2^{\alpha_2} \quad \text{and} \quad \widetilde{C}_R(x_1,x_2) \geqslant x_1^{\beta_1} x_2^{\beta_2}.$$

H6

$$\int \frac{dF(z_1,z_2)}{C_L\left(F_{L_1}(z_1), F_{L_2}(z_2)\right)\widetilde{C}_R\left(S_{R_1}(z_1), S_{R_2}(z_2)\right)} < \infty$$

and for some $\varepsilon > 0$ arbitrary small

$$\int \left[\frac{F_{L_1}^{1-\alpha_1}(z_1)\,\mathcal{K}_{L_1}^{1/2+\varepsilon}(z_1)}{F_{L_2}^{\alpha_2}(z_2)\,S_{R_1}^{\beta_1}(z_1)\,S_{R_2}^{\beta_2}(z_2)} + \frac{F_{L_2}^{1-\alpha_2}(z_2)\,\mathcal{K}_{L_2}^{1/2+\varepsilon}(z_2)}{F_{L_1}^{\alpha_1}(z_1)\,S_{R_1}^{\beta_1}(z_1)\,S_{R_2}^{\beta_2}(z_2)} + \frac{S_{R_1}^{1-\beta_1}(z_1)\,\mathcal{K}_{R_1}^{1/2+\varepsilon}(z_1)}{F_{L_1}^{\alpha_1}(z_1)\,F_{L_2}^{\alpha_2}(z_2)_2\,S_{R_2}^{\beta_2}(z_2)} \right.$$
$$\left. + \frac{S_{R_2}^{1-\beta_2}(z_2)\,\mathcal{K}_{R_2}^{1/2+\varepsilon}(z_2)}{F_{L_1}^{\alpha_1}(z_1)\,F_{L_2}^{\alpha_2}(z_2)\,S_{R_1}^{\beta_1}(z_1)} \right] dF(z_1, z_2) < \infty.$$

Theorem 4.2

*Under assumptions **H1-H6**, we have*

i) For all $(x_1, x_2) \in \mathcal{X}$

$$F_n(x_1, x_2) - F(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{x_1}\int_0^{x_2}[\rho(Z_{1i}, Z_{2i}, A_{1i}, A_{2i}; t_1, t_2)\, C_L(F_{L_1}(t_1), F_{L_2}(t_2))$$
$$\times \widetilde{C}_R(S_{R_1}(t_1), S_{R_2}(t_2))]dF(t_1, t_2) + R_n(x_1, x_2),$$

where $\sup_{(x_1,x_2)\in\mathcal{X}} |R_n(x_1, x_2)| = o_P\left(n^{-1/2}\right)$.

ii) The process $\sqrt{n}(F_n - F)$ converges weakly in $l^\infty(\mathcal{X})$ to a tight centered Gaussian process G_F.

Note that *ii)* follows directly from *i)* and allows to prove the next theorem which gives the weak convergence of the process $\sqrt{n}(C_n - C)$.

Theorem 4.3

*Under assumptions **H1-H6**, the process $\sqrt{n}(C_n - C)$ converges weakly in $l^\infty\left([0,1]^2\right)$ to the tight Gaussian process*

$$G(u, v) = G^*(u, v) - \partial_1 C(u, v)\, G^*(u, 1) - \partial_2 C(u, v)\, G^*(1, v),$$

where $G^*(u, v) = G_F\left(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)\right)$.

This result is an extension of [[23], Theorem 2] to our case of bivariate twice censored data.

4.3 Simulation study

We carry out a simulation study to illustrate the performance of our estimator. As a starting point, we create a bivariate survival distribution of the Gumbel copula model where the margins are assumed to be Pareto model.That is

$$F_1(t_1) = 1 - t_1^{-\lambda_1} \; and \; F_2(t_2) = 1 - t_2^{-\lambda_2} \quad t_1, t_2 \geq 0.$$

Such that $\lambda_1, \lambda_2 > 0$. We assume that the corresponding percentage of the observed data is given by $p_1 = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$ for the first sample. So that we can select the values 0.3 for λ_1 and $0.95, 0.90, 0.85, 0.80, 0.75$ for p_1, the equation $p_1 = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$ is then resolved to get the pertaining λ_2-values. This step, we permit a certain amount of censoring of T to be $5\%, 10\%, 15\%, 20\%$ and 25%. We generate 1000 independent replicates for each common size n varying from $n = 30, 50, 100, 500, 1000$ for the two samples, to apply the results obtained throughout all replicates as empirical proof for our final show.

We ought to select the survival copula parameter values (α, β), using the link between Kendall's τ and the transformed of the underlying survival copula formulated by $\tau_{\alpha,\beta} = 4E(\tilde{C}_{\alpha,\beta}(u,v)) - 1$, since τ considred as a function of the dependency parameter in Archimedean copula models. Then, we can select the values 0.1 for the first parameter α and low dependence that corresponding to 0.05 Kendall's tau values, next we applying the transformed of the underlying survival copula to obtain β-values. In a similar way, we determine the values of the additional parameters (α, β) for the corresponding Kendall's tau values 0.5 (moderate dependence) and 0.7 (strong dependence)[25], summarized in the following Tables (Tables1;2;3) below.

$\tau = 0.05$, $\alpha = 0.1 \to \beta = 1.6$					
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
% of censoring	*Mise*				
5%	1.20862	1.12774	0.82887	0.59617	0.55740
10%	0.99576	1.3269	0.87537	0.61283	0.55754
15%	1.09038	1.0898	0.93174	0.62488	0.56949
20%	1.56908	1.06111	0.79922	0.6305	0.5765
25%	0.945	1.0518	0.85303	0.63527	0.58982

Table 4.1: The estimator performance based on Gumbel survival copula in the case of weak dependence (τ=0.05). Mise of the estimators are determined for various censoring values.

$\tau = 0.5$, $\alpha = 0.2 \to \beta = 1.82$					
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
% of censoring	*Mise*				
5%	0.90043	1.27155	0.81397	0.60066	0.54418
10%	0.90043	0.98767	0.7141	0.59024	0.56295
15%	1.48457	1.02722	0.83566	0.61167	0.56714
20%	1.25187	0.77571	0.66976	0.62747	0.57606
25%	0.91859	0.90437	0.82308	0.61196	0.54947

Table 4.2: The estimator performance based on Gumbel survival copula in the case of moderate dependence (τ=0.5). Mise of the estimators are determined for various censoring values..

$\tau = 0.7$, $\alpha = 0.4 \rightarrow \beta = 2.99$					
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
% of censoring	*Mise*				
5%	0.86032	0.77042	0.7026	0.55432	0.52912
10%	0.9174	0.79666	0.69767	0.56181	0.53337
15%	0.89035	0.79593	0.63963	0.57308	0.54344
20%	0.92682	0.78261	0.72459	0.57451	0.54861
25%	1.18433	0.73591	0.63566	0.57448	0.56208

Table 4.3: The estimator performance based on Gumbel survival copula in the case of strong dependence (τ=0.7). Mise of the estimators are determined for various censoring values.

Chapter 5

Kernel estimation of the copula function under twice censoring

5.1 Kernel copula estimators

We begin by introducing some general notations and definitions. Let $T = (T_1, T_2)$ be a couple of positive real random variables (r.r.v.) with support $\mathcal{T}_1 \times \mathcal{T}_2$, joint distribution function H and continuous margins H_1, H_2 and let C be the copula function of T defined on $[0,1]^2$ by $C(u,v) = H(H_1^{-1}(u), H_2^{-1}(v))$, where f^{-1} denotes the generalized inverse of a non decreasing function f. Furthermore, let $R = (R_1, R_2)$ be a couple of positive right censoring variables and $L = (L_1, L_2)$ be a couple of positive left censoring ones. We assume that the variables R, L and T are independent. In the twice censoring setting, instead of observing T, we observe a sample $(Z_{1i}, Z_{2i}, \Delta_{1i}, \Delta_{2i})_{1\leq i\leq n}$ of i.i.d. copies of the vector $(Z_1, Z_2, \Delta_1, \Delta_2)$, where for $k \in \{1,2\}$, $Z_k = \max(\min(T_k, R_k), L_k)$ and Δ_k is the indicator of censoring given by

$$\Delta_k = I_{\{L_k<R_k<T_k\}} + 2I_{\{\min(T_k,R_k)\leq L_k\}}.$$

($I_{\{.\}}$ being the indicator function).

For any r.r.v. V, F_V, S_V, I_V and T_V denote, respectively, the distribution function, the survival function, the lower and the upper endpoint of the support of V. Furthermore, denote by $\left(Z'_{kj}\right)_{1\leq j\leq m}$ $(m \leq n)$ the distinct values of $(Z_{ki})_{1\leq i\leq n}$. The product-limit estimator of F_{L_k} is defined by

$$\widehat{F}_{L_k}(t) = \prod_{j/Z'_{kj}>t}\left(1 - \frac{\sum_{i=1}^{n} I_{\left\{Z_{ki}=Z'_{kj},\Delta_{ki}=2\right\}}}{n\widehat{F}_{Z_k}\left(Z'_{kj}\right)}\right),$$

where $\widehat{F}_{Z_k}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_{ki}\leq t\}}$.
In addition, the product-limit estimator of S_{R_k} is given by [43] as follows

$$\widehat{S}_{R_k}(t) = \prod_{i/Z_{ki}\leq t}\left(1 - \frac{I_{\{\Delta_{ki}=1\}}}{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)}\right).$$

Furthermore, [50] proposed the following estimator of H

$$H_n(t_1,t_2) = \frac{1}{n}\sum_{j=1}^{n} W_{jn} I_{\{Z_{1j}\leq t_1, Z_{2j}\leq t_2\}}, \tag{5.1}$$

where $W_{jn} = I_{\{\Delta_{1j}=0\}} I_{\{\Delta_{2j}=0\}} \widehat{g}(Z_{1j}, Z_{2j})$
and $\widehat{g}(z_1,z_2) = C_L\left(\widehat{F}_{L_1}(z_1), \widehat{F}_{L_2}(z_2)\right)^{-1} \widetilde{C}_R\left(\widehat{S}_{R_1}(z_1), \widehat{S}_{R_2}(z_2)\right)^{-1}$ ($C_L$and $\widetilde{C}_R$ being the copula of L and the survival copula of R respectively).

We denote by C the copula function of T and we assume that C and H are twice continuously differentiable. We also assume that the second order derivatives of H are uniformly bounded. Let $k : \mathbb{R} \to \mathbb{R}$ be a symmetric kernel with $\int k(t)\,dt = 1$ and let h_n be a sequence of positive bandwidths. We set $K(x) = \int_{-\infty}^{x} k(t)\,dt$.
Inspired by [23], we define the following smoothed estimators of the distribution function and the copula function of T

$$\widehat{H}_n^1(t_1,t_2) = \frac{1}{n}\sum_{j=1}^{n} W_{jn} K\left(\frac{t_1 - Z_{1j}}{h_n}\right) K\left(\frac{t_2 - Z_{2j}}{h_n}\right) \tag{5.2}$$

and

$$\widehat{C}_n^1(u,v) = \widehat{H}_n^1\left(\left(\widehat{H}_{1n}^1\right)^{-1}(u), \left(\widehat{H}_{2n}^1\right)^{-1}(v)\right). \tag{5.3}$$

where $\widehat{H}_{1n}^1(t_1) = \lim_{t_2\to\infty} \widehat{H}_n^1(t_1,t_2)$ and $\widehat{H}_{2n}^1(t_2) = \lim_{t_1\to\infty} \widehat{H}_n^1(t_1,t_2)$.
As mentioned in [23], this estimator has the disadvantage that its performance depends on the margins H_1 and H_2. So, we use as in [23] and [42], a transformation
$\left(\tilde{T}_1, \tilde{T}_2\right) = \left(\Phi^{-1}(H_1(T_1)), \Phi^{-1}(H_2(T_2))\right)$ for some increasing distribution function Φ, such that Φ' and Φ'^2/Φ are bounded. Note that the variables $\left(\tilde{T}_1, \tilde{T}_2\right)$ and (T_1, T_2) have the same copula. This leads to new estimators of the joint distribution function and the copula function of (T_1, T_2) given by

$$\widehat{H}_n^2(t_1,t_2) = \frac{1}{n}\sum_{j=1}^{n} W_{jn} K\left(\frac{t_1 - \Phi^{-1}(H_{1n}(Z_{1j}))}{h_n}\right) K\left(\frac{t_2 - \Phi^{-1}(H_{2n}(Z_{2j}))}{h_n}\right),$$

where $H_{1n}(t_1) = \lim_{t_2\to\infty} H_n(t_1,t_2)$, $H_{2n}(t_2) = \lim_{t_1\to\infty} H_n(t_1,t_2)$ and

$$\widehat{C}_n^2(u,v) = \widehat{H}_n^2\left(\Phi^{-1}(u), \Phi^{-1}(v)\right).$$

5.2 Weak convergence of the proposed estimators

To establish the weak convergence of the processes $\sqrt{n}\left(\widehat{H}_n^j(t_1,t_2) - H(t_1,t_2)\right)$ and $\sqrt{n}\left(\widehat{C}_n^j(u,v) - C(u,v)\right)$, $(u,v) \in [0,1]^2$ for $j \in \{1,2\}$, we will follow a similar approach as that of [23]. For that we need the following assumptions and notations. For $k \in \{1,2\}$ and $l \in \{0,1,2\}$, denote by $\mathbb{S}_k = \{z \in \mathbb{R} : I_{H_k^{(1)}} < z < \tau_k\}$, where τ_k is such that $I_{H_k^{(1)}} < \tau_k < T_k$ and denote by $H_k^{(l)}(t) = P(Z_k \le t, \Delta_k = l)$ the sub-distribution function of Z_k for $\Delta_k = l$ and $I_{H_k^{(l)}} = \inf\{t \in \mathbb{R}/H_k^{(l)}(t) > 0\}$ the lower endpoint of the support of $H_k^{(l)}$. We assume that

H1: $I_{L_k} < I_{R_k}$ and $T_{L_k} < T_{R_k} \le T_{T_k}$.

H2: There exist $\theta_{k1} > I_{R_k}$ and $\theta_{k2} < T_{R_k}$ such that

$$\forall n \in \mathbb{N}^*, \forall 1 \le i \le n : \Delta_{ki} = 1 \Rightarrow \theta_{k1} \le Z_{ki} \le \theta_{k2} \quad \text{almost surely } (a.s.).$$

H3: $\displaystyle\int_{I_{H_k^{(1)}}}^{+\infty} \frac{dH_k^{(2)}(z)}{\left(F_{Z_k}(z)\right)^2} < \infty.$

H4: The first and the second partial derivatives of C_L and $\widetilde{C}_R$ are bounded on $[0,1]^2$. Moreover, $C_L(t_1,t_2) \neq 0$ and $\widetilde{C}_R(t_1,t_2) \neq 0$ for $t_1 \neq 0$ and $t_2 \neq 0$.

H5: There exist $0 \leqslant \alpha_k \leqslant 1$ and $0 \leqslant \beta_k \leqslant 1$ such that

$$C_L(t_1,t_2) \geqslant t_1^{\alpha_1} t_2^{\alpha_2} \quad \text{and} \quad \widetilde{C}_R(t_1,t_2) \geqslant t_1^{\beta_1} t_2^{\beta_2}.$$

H6:

$$\int \frac{dF(z_1,z_2)}{C_L\left(F_{L_1}(z_1), F_{L_2}(z_2)\right) \widetilde{C}_R\left(S_{R_1}(z_1), S_{R_2}(z_2)\right)} < \infty$$

and for some $\varepsilon > 0$ arbitrary small

$$\int \left[\frac{F_{L_1}^{1-\alpha_1}(z_1)\,\mathcal{K}_{L_1}^{1/2+\varepsilon}(z_1)}{F_{L_2}^{\alpha_2}(z_2)\, S_{R_1}^{\beta_1}(z_1)\, S_{R_2}^{\beta_2}(z_2)} + \frac{F_{L_2}^{1-\alpha_2}(z_2)\,\mathcal{K}_{L_2}^{1/2+\varepsilon}(z_2)}{F_{L_1}^{\alpha_1}(z_1)\, S_{R_1}^{\beta_1}(z_1)\, S_{R_2}^{\beta_2}(z_2)} + \frac{S_{R_1}^{1-\beta_1}(z_1)\,\mathcal{K}_{R_1}^{1/2+\varepsilon}(z_1)}{F_{L_1}^{\alpha_1}(z_1)\, F_{L_2}^{\alpha_2}(z_2)_2\, S_{R_2}^{\beta_2}(z_2)} + \frac{S_{R_2}^{1-\beta_2}(z_2)\,\mathcal{K}_{R_2}^{1/2+\varepsilon}(z_2)}{F_{L_1}^{\alpha_1}(z_1)\, F_{L_2}^{\alpha_2}(z_2)\, S_{R_1}^{\beta_1}(z_1)} \right] dF(z_1,z_2) < \infty,$$

where

$$\mathcal{K}_{R_k}(s) = Var\left(G_{R_k}(s)\right),$$

G_{R_k} being the limiting the process of $\sqrt{n}\left(\widehat{\Lambda}_{R_k}(t) - \Lambda_{R_k}(t)\right), t \in \mathbb{S}_k$; where $\widehat{\Lambda}_{R_k}(t) = \int\limits_0^t \frac{dH_{nk}^{(1)}(u)}{F_{L_k}^*(u^-) - \widehat{F}_{Z_k}(u^-)}$ is the estimator of the cumulative hazard function Λ_{R_k} of R_k and

$$\mathcal{K}_{L_k}(s) = \int\limits_s^{+\infty} \frac{dF_{L_k}(u)}{F_{L_k}^2(u)\, F_{T_k \wedge R_k}(u)}.$$

H7: The kernel function k has a compact support and $\mu_2 = \int u^2 k(u)\, du < \infty$.

H8: $h_n^2 \sqrt{n} \to 0$, as $n \to \infty$.

Assumptions **H1-H6** have been used in [50] and the assumptions **H7** and **H8** on the kernel and the bandwidth have been used in [23].

Theorem 5.1

Under Assumptions **H1-H8**, *we have*

$$\sqrt{n} \sup_{(t_1,t_2)\in \mathcal{T}_1 \times \mathcal{T}_2} \left|\widehat{H}_n^1(t_1,t_2) - H_n(t_1,t_2)\right| \to^P 0, \text{ as } n \to \infty.$$

Proof: Using Lemma 3.1 and relations $(A.13)$ and $(A.15)$ of [50], this Theorem can be proved in the same way as Theorem 3 of [23]. ■

From this Theorem and Theorem $3.2 - ii)$ of [50], we can deduce the weak convergence of the process $\sqrt{n}\left(\widehat{H}_n^1 - H\right)$ to the tight centered Gaussian limit process $\boldsymbol{G}_H(t_1,t_2)$ defined in [50]. This is stated in the following Corollary.

Under Assumptions $\mathbf{H}_1 - \mathbf{H}_8$, the process $\sqrt{n}\left(\widehat{H}_n^1 - H\right)$ converges weakly, in $l^\infty(\mathcal{T}_1 \times \mathcal{T}_2)$[1] to the tight entered Gaussian limit process $\boldsymbol{G}_H$ defined in [50].

Now, we will prove the weak convergence of the processes $\sqrt{n}\left(\widehat{C}_n^1(u,v) - C(u,v)\right)$ and $\sqrt{n}\left(\widehat{C}_n^2(u,v) - C(u,v)\right)$. For that, we need the following additional hypotheses which have been introduced in [23].

[1] For any non empty set A, $l^\infty(A)$ denotes the space of all bounded real-valued functions defined on A.

H9: $\dfrac{\partial^2 C}{\partial u^2} = O\left(\dfrac{1}{u(1-u)}\right), \dfrac{\partial^2 C}{\partial v^2} = O\left(\dfrac{1}{v(1-v)}\right)$ and $\dfrac{\partial^2 C}{\partial u \partial v} = O\left(\dfrac{1}{\sqrt{uv(1-u)(1-v)}}\right)$.

H10: Consider the sample of transformed observations

$\left(N_{1j}, N_{2j}, I_{\{\Delta_{1j}=0\}}, I_{\{\Delta_{2j}=0\}}\right)_{1\leq j\leq n} := \left(\Phi^{-1}\left(H_1\left(Z_{1j}\right)\right), \Phi^{-1}\left(H_2\left(Z_{2j}\right)\right), I_{\{\Delta_{1j}=0\}}, I_{\{\Delta_{2j}=0\}}\right)_{1\leq j\leq n}$ and the corresponding weights $W_{jn}^{\Phi} := I_{\{\Delta_{1j}=0\}} I_{\{\Delta_{2j}=0\}} \widehat{g}^{\Phi}\left(Z_{1j}, Z_{2j}\right)$, where $\widehat{g}^{\Phi}$ is computed by the same method as $\widehat{g}$, but based on the sample of transformed observations. Assume that $W_{jn}^{\Phi} = W_{jn}$.

Theorem 5.2

i) Under $\mathbf{H}_1 - \mathbf{H}_8$, *the process* $\sqrt{n}\left(\widehat{C}_n^1 - C\right)$ *converges weakly in* $l^\infty\left([0,1]^2\right)$ *to the Gaussian process*

$$Z_C(u,v) = G^*(u,v) - \frac{\partial C}{\partial u}(u,v) G^*(u,1) - \frac{\partial C}{\partial v}(u,v) G^*(1,v),$$

where $G^*(u,v) = G_H\left(H_1^{-1}(u), H_2^{-1}(v)\right)$.

ii) Under $\mathbf{H}_1 - \mathbf{H}_{10}$, *the process* $\sqrt{n}\left(\widehat{C}_n^2 - C\right)$ *converges weakly in* $l^\infty\left([0,1]^2\right)$ *to the process* Z_C.

Proof: Using relations $(A.13)$ and $(A.15)$ of [50], this Theorem can be proved in the same way as Theorem 4 of [23]. ■

5.3 Simulation study

We carry out a simulation study to illustrate the performances of our proposed estimators. As a starting point, we create a bivariate survival distribution of the Gumbel copula model where the margins are assumed to have a Pareto distribution. In other words, the copula function of the couple $T = (T_1, T_2)$ is given for all $(u, v) \in [0, 1]^2$ by

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

with the parameters $\alpha > 0$ and $\beta \geq 1$, and the margins of T are given for all $t_1, t_2 \geq 0$ by

$$H_1(t_1) = 1 - t_1^{-\lambda_1} \quad \text{and} \quad H_2(t_2) = 1 - t_2^{-\lambda_2},$$

where λ_1 and λ_2 are positive parameter.

Under this model, the proportion of the observed data in the first sample is given by $p_1 = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$. So, we select the value 0.3 for λ_1 and the values $0.95, 0.90, 0.85, 0.80$ and 0.75 for p_1 and we resolve the equation $p_1 = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$ to get the pertaining λ_2-values. This gives respectively the following censoring rates: $5\%, 10\%, 15\%, 20\%$ and 25%. Moreover, we select the copula parameters α and β using the link between the Kendall's τ and the copula function formulated by

$$\tau_{\alpha,\beta} = 4 \int_{[0,1]^2} C_{\alpha,\beta}(u, v) dC_{\alpha,\beta}(u, v) - 1. \tag{5.4}$$

Then, we select the value 0.1 for the first parameter α and the value 0.05 for the Kendall's tau (low dependence) we apply relation (5.4) to obtain the value of β. In a similar way, we determine the values of the parameters α and β for the corresponding Kendall's tau values of 0.5 (moderate dependence) and 0.7 (high dependence).

In order to show the performances of our estimators $\widehat{C}_n^1$ and $\widehat{C}_n^2$, we compute the mean integrated squared error (MISE) for the two estimators. For that, we generate 1000 samples of size n for each latent variable. We take different values of the sample size n which will be specified later. Then, we compute the integrated squared error of $\widehat{C}_n^1$ and $\widehat{C}_n^2$ for each sample. Finally, we calculate the mean of the obtained values which is an approximation of the MISE of the two estimators. Our obtained results are presented in Tables 5.1–5.6 below. Note that we use the Gaussian kernel and a fixed bandwidth $h_n = 0.2$ to compute the estimators $\widehat{C}_n^1$ and $\widehat{C}_n^2$. Moreover, the transformation Φ that we use to compute the estimator $\widehat{C}_n^2$ is the distribution function of the exponential distribution with parameter 1.

$\tau = 0.05$, $\alpha = 0.1 \rightarrow \beta = 1.6$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.49083	0.4915	0.49767	0.50078
10%	0.48611	0.49793	0.49892	0.50767
15%	0.49806	0.50814	0.50567	0.51082
20%	0.50083	0.51852	0.52474	0.52109
25%	0.51	0.52185	0.52947	0.52949

Table 5.1: The obtained results for the estimator $\widehat{C}_n^1$ under weak dependence.

$\tau = 0.5$, $\alpha = 0.2 \rightarrow \beta = 1.82$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.485	0.4904	0.4984	0.50081
10%	0.488	0.50331	0.49902	0.50801
15%	0.49708	0.51725	0.51224	0.51415
20%	0.53611	0.51154	0.5066	0.52097
25%	0.48333	0.54088	0.5357	0.53317

Table 5.2: The obtained results for the estimator $\widehat{C}_n^1$ under moderate dependence.

From these results, we remark that the estimators $\widehat{C}_n^1$ and $\widehat{C}_n^2$ have good performances. Not surprisingly, the quality of estimation decreases when the rate of censoring increases. However, we remark that the variation of the sample size and the strength of the dependence does not affect the quality of estimation.

$\tau = 0.7$, $\alpha = 0.4 \to \beta = 2.99$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.48444	0.49400	0.49683	0.50132
10%	0.49556	0.49678	0.50419	0.50353
15%	0.49444	0.50002	0.51193	0.51327
20%	0.51797	0.52661	0.51745	0.52639
25%	0.57417	0.51481	0.51541	0.53169

Table 5.3: The obtained results for the estimator $\widehat{C}_n^1$ under strong dependence.

$\tau = 0.05$, $\alpha = 0.1 \to \beta = 1.6$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.48667	0.492	0.49654	0.50099
10%	0.49444	0.50942	0.49671	0.5083
15%	0.49074	0.50907	0.51111	0.5127
20%	0.49665	0.50756	0.52235	0.52594
25%	0.52796	0.50537	0.52947	0.53352

Table 5.4: The obtained results for the estimator $\widehat{C}_n^2$ under weak dependence.

$\tau = 0.5$, $\alpha = 0.2 \to \beta = 1.82$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.48667	0.49495	0.49691	0.50680
10%	0.48778	0.49737	0.51045	0.50661
15%	0.49616	0.50226	0.51452	0.52062
20%	0.51845	0.51419	0.52454	0.52187
25%	0.50605	0.53524	0.52407	0.528

Table 5.5: The obtained results for the estimator $\widehat{C}_n^2$ under moderate dependence.

$\tau = 0.7$, $\alpha = 0.4 \to \beta = 2.99$				
Sample size	$n = 30$	$n = 50$	$n = 100$	$n = 500$
% of censoring	*MISE*			
5%	0.4875	0.4928	0.49639	0.50186
10%	0.4875	0.49695	0.50026	0.50606
15%	0.50105	0.50731	0.51531	0.50731
20%	0.51690	0.50809	0.51553	0.50836
25%	0.50361	0.49866	0.51561	0.50929

Table 5.6: The obtained results for the estimator $\widehat{C}_n^2$ under strong dependence.

Conclusion

In this thesis, we introduce the empirical copula function in the case of bivariate twice censored data and we establish its weak convergence with simulation. Our approach is based on a representation of the corresponding joint distribution function estimator as a sum of i.i.d. centered random variables. The results we obtain extend those given in [28] and [19] in the setting of bivariate right censored data. We prove our results only in the case where the copula functions of the left and the right censoring variables are known. It would be interesting to consider a general bivariate twice censoring model and to look also at other types of censored data, such as doubly or interval censored data. Our obtained results allows to propose and study smoothed copula estimators for bivariate twice censored data.

Appendix

5.4 Proofs

Proof Lemma 4.1 : We follow the same steps of the proof of [[36], Theorem 1]. Let K, C, λ and δ be some positive universal constants. For $k \in \{1,2\}$, we set $T_k(u) = \log(S_{R_k}(u))$, $\widehat{T}_k(u) = \log\left(\widehat{S}_{R_k}(u)\right) = \sum_{i=1}^{n} I_{\{Z_{ki} \leq u, A_{ki}=1\}} \log\left(1 - \frac{1}{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)}\right)$ and $\widetilde{T}_k(u) = -\sum_{i=1}^{n} \frac{I_{\{Z_{ki} \leq u, A_{ki}=1\}}}{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)}$.
Proceeding as in [36], we can show that for $I \leq u \leq T$

$$\widehat{S}_{R_k}(u) - S_{R_k}(u) = S_{R_k}(u)\left(A_k(n,u) + \widetilde{B}_k(n,u)\right) + \widetilde{R}_k(n,u), \tag{5.5}$$

where

$$A_k(n,u) = -\frac{H_{nk}^{(1)}(u) - H_k^{(1)}(u)}{F_{L_k}(u) - F_{Z_k}(u)} + \int_0^u \frac{H_{nk}^{(1)}(y) - H_k^{(1)}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} d\left(F_{L_k}(u) - F_{Z_k}(u)\right),$$

$$\widetilde{B}_k(n,u) = -\int_0^u \frac{\widehat{F}_{L_k}(y^-) - \widehat{F}_{Z_k}(y^-) - F_{L_k}(y) + F_{Z_k}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} dH_k^{(1)}(y)$$

and

$$\widetilde{R}_k(n,u) = S_{R_k}(u)\left(R_{k2}(u) + R_{k3}(u) + R_{k4}(u)\right) + R_{k1}(u),$$

with

$$R_{k1}(u) = \widehat{S}_{R_k}(u) - S_{R_k}(u) - S_{R_k}(u)\left(\widehat{T}_k(u) - T_k(u)\right),$$

$$R_{k2}(u) = \widehat{T}_k(u) - \widetilde{T}_k(u),$$

$$R_{k3}(u) = \widetilde{T}_k(u) + \sum_{i=1}^{n} \frac{I_{\{Z_{ki} \leq u, A_{ki}=1\}}}{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)} \times \left[1 - \frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^{-}\right) - \widehat{F}_{Z_k}\left(Z_{ki}^{-}\right) - F_{L_k}(Z_{ki}) + F_{Z_k}(Z_{ki})\right)}{n\left(F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})\right)}\right]$$

and

$$R_{k4}(u)=\frac{1}{n}\int_{0}^{+\infty}\int_{0}^{+\infty}\frac{I_{\{y\leq u\}}I_{\{x<y\}}}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}d\left[\sqrt{n}\left(H_{nk}^{(1)}(y)-H_{k}^{(1)}(y)\right)\right]$$
$$d\left[\sqrt{n}\left(\widehat{F}_{L_k}(x^-)-\widehat{F}_{Z_k}(x^-)-F_{L_k}(x)+F_{Z_k}(x)\right)\right].$$

As in [36], we will prove the following lemmas.

Lemma 5.3

$$\sup_{I\leq u\leq T}|R_{k2}(u)|=O_{a.s.}\left(\frac{1}{n}\right).$$

Lemma 5.4

$$P\left(\sup_{I\leq u\leq T}|nR_{k3}(u)|>x\right)\leq K\exp\left\{-\lambda\delta^2x\right\}\ \textit{if}\ 0\leq x<\frac{2n}{\delta}.$$

Lemma 5.5

$$P\left(\sup_{I\leq u\leq T}|nR_{k4}(u)|>x\right)\leq K\exp\left\{-\lambda\delta^2x\right\}\ \textit{for}\ x>0.$$

From these lemmas, we deduce that

$$\sup_{I\leq u\leq T}\left|\widetilde{R}_k(n,u)\right|=O_P\left(\frac{1}{n}\right). \tag{5.6}$$

Moreover, relation (4.3) permits to write

$$\widetilde{B}_k(n,u)=-\int_{0}^{u}\frac{\left(\widehat{F}_{L_k}(y^-)-F_{L_k}(y)\right)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}dH_k^{(1)}(y)+\int_{0}^{u}\frac{\widehat{F}_{Z_k}(y^-)-F_{Z_k}(y)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}dH_k^{(1)}(y)$$
$$=B_k(n,u)-\int_{0}^{u}\frac{R_{L_k}(n,y^-)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}dH_k^{(1)}(y).$$

Combining this with (5.5), we obtain

$$\widehat{S}_{R_k}(u)-S_{R_k}(u)=S_{R_k}(u)\left(A_k(n,u)+B_k(n,u)\right)+R_k(n,u),$$

where

$$R_k(n,u)=-S_{R_k}(u)\int_{0}^{u}\frac{R_{L_k}(n,y^-)}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}dH_k^{(1)}(y)+\widetilde{R}_k(n,u)$$
$$=:\widetilde{\widetilde{R}}_k(n,u)+\widetilde{R}_k(n,u). \tag{5.7}$$

Since $F_{L_k}(y) - F_{Z_k}(y) = F_{L_k}(y) S_{X_k}(y) S_{R_k}(y)$, we get

$$\sup_{I \leq u \leq T} \left| \widetilde{\widetilde{R}}_k(n,u) \right| \leq \sup_{I \leq u \leq T} |R_{L_k}(n,u)| \int_0^u \frac{dH_k^{(1)}(y)}{(F_{L_k}(y) S_{X_k}(y) S_{R_k}(y))^2}$$
$$\leq \frac{1}{(F_{L_k}(I_{R_k}) S_{X_k}(T) S_{R_k}(T))^2} \sup_{I_{R_k} \leq u \leq T} |R_{L_k}(n,u)|.$$

This together with (4.2), (5.6) and (5.7) permits to write

$$\sup_{I \leq u \leq T} |R_k(n,u)| = O_P\left(\frac{1}{n}\right),$$

which gives the claimed result. ■

It remains to prove lemmas $2 - 4$.

Proof of Lemma 5.3 : We have

$$|R_{k_2}(u)| = \left| \widehat{T}_k(u) - \widetilde{T}_k(u) \right|$$
$$\leq \sum_{i=1}^n \left| I_{\{Z_{k_i} \leq u, A_{ki}=1\}} \log\left(1 - \frac{1}{n\left(\widehat{F}_{L_k}\left(Z_{k_i}^-\right) - \widehat{F}_{Z_k}\left(Z_{k_i}^-\right)\right)}\right) + \frac{I_{\{Z_{k_i} \leq u, A_{ki}=1\}}}{n\left(\widehat{F}_{L_k}\left(Z_{k_i}^-\right) - \widehat{F}_{Z_k}\left(Z_{k_i}^-\right)\right)} \right|$$
$$\leq \sum_{i=1}^n \left| \log\left(1 - \frac{1}{n\left(\widehat{F}_{L_k}\left(Z_{k_i}^-\right) - \widehat{F}_{Z_k}\left(Z_{k_i}^-\right)\right)}\right) + \frac{1}{n\left(\widehat{F}_{L_k}\left(Z_{k_i}^-\right) - \widehat{F}_{Z_k}\left(Z_{k_i}^-\right)\right)} \right| I_{\{Z_{k_i} \leq T\}}.$$

Since $|\log(1-z) + z| \leq z^2$ for $0 \leq z \leq 1/2$ and
$\inf_{I_{R_k} \leq t \leq T} n\left\{ \widehat{F}_{L_k}(t^-) - \widehat{F}_{Z_k}(t^-) \right\} \geq \frac{n}{2}\left(F_{L_k}(I_{R_k}) S_{X_k}(T) S_{R_k}(T)\right) \geq 2$ for n large enough,
we deduce that for $I \leq u \leq T$

$$|R_{k_2}(u)| \leq \frac{1}{n^2} \sum_{i=1}^n \frac{I_{\{Z_{ki} \leq T\}}}{\left(\widehat{F}_{L_k}\left(Z_{k_i}^-\right) - \widehat{F}_{Z_k}\left(Z_{k_i}^-\right)\right)^2}$$
$$\leq \frac{1}{n} \frac{1}{\inf_{I_{R_k} \leq t \leq T} \left(\widehat{F}_{L_k}(t^-) - \widehat{F}_{Z_k}(t^-)\right)^2}$$
$$\leq \frac{1}{n} \frac{4}{(F_{L_k}(I_{R_k}) S_{X_k}(T) S_{R_k}(T))^2} \quad a.s. \text{ for n large enough.}$$

Thus $\sup_{I \leq u \leq T} |R_{k_2}(u)| = O_{a.s.}\left(\frac{1}{n}\right).$ ■

Proof of Lemma 5.4 : To prove this lemma, we need to apply some exponential inequalities for $\widehat{F}_{Z_k}$, $H_{nk}^{(1)}$ and $\widehat{F}_{L_k}$. Regarding $\widehat{F}_{Z_k}$, [16] proved that there exists a positive constant D such that for all $x > 0$

$$P\left(\sqrt{n}\sup_{t\in\mathbb{R}}\left|\widehat{F}_{Z_k}(t) - F_{Z_k}(t)\right| > x\right) \leq D\exp\left(-2x^2\right). \tag{5.8}$$

Moreover, writing

$$H_{nk}^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{R_{ki}\leq t,\, L_{ki}-R_{ki}<0,\, R_{ki}-X_{ki}<0\}}$$

allows to apply [[31], Theorem 1-m] to get for all $x > 0$ and $\varepsilon > 0$

$$P\left(\sqrt{n}\sup_{t\in\mathbb{R}}\left|H_{nk}^{(1)}(t) - H_k^{(1)}(t)\right| > x\right) \leq D\exp\left(-(2-\varepsilon)x^2\right), \tag{5.9}$$

where D is a positive constant.

Furthermore, adapting [[5], Theorem 1], we get for all $x > 0$ and $\theta > \min(I_{X_k}, I_{R_k})$

$$P\left(\sqrt{n}\sup_{t\geq\theta}\left|\widehat{F}_{L_k}(t) - F_{L_k}(t)\right| > \frac{x}{F_{X_k\wedge R_x}(\theta)}\right) \leq 2.5\exp\left(-2x^2 + Dx\right), \tag{5.10}$$

where D is a positive constant.

Set

$$\Gamma_{kn} = \left\{\inf_{\theta_{k1}\leq t\leq\theta_{k2}}\left\{\widehat{F}_{L_k}\left(t^-\right) - \widehat{F}_{Z_k}\left(t^-\right)\right\} \geq \frac{2}{n}\right\}$$

and

$$\Delta_{kn}(u) = \left\{\left|\widehat{F}_{L_k}\left(Z_{ki}^-\right) - \widehat{F}_{Z_k}\left(Z_{ki}^-\right) - F_{L_k}(Z_{ki}) + F_{Z_k}(Z_{ki})\right| < \frac{1}{2}\left(F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})\right) \text{ or } \right.$$
$$\left. Z_{ki} \leq u \text{ for all } 1\leq i\leq n \text{ such that } A_{ki} = 1\right\}.$$

Remarking that

$$\widetilde{T}_k(u) = -\sum_{i=1}^{n}\frac{I_{\{Z_{ki}\leq u, A_{ki}=1\}}}{n\left(F_{L_k}\left(Z_{ki}^-\right) - F_{Z_k}\left(Z_{ki}^-\right)\right)}\frac{1}{1+\frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)-F_{L_k}(Z_{ki})+F_{Z_k}(Z_{ki})\right)}{n\left(F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})\right)}}$$

and that $\left|\frac{1}{1+z} - 1 + z\right| < 2z^2$ for $|z| < 1/2$, we deduce that on Γ_{kn}, we have

$$|R_{k3}(u)| \leq 2\sum_{i=1}^{n}\frac{I_{\{Z_{ki}\leq u, A_{ki}=1\}}}{n\left(F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})\right)}\left[\frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^-\right) - \widehat{F}_{Z_k}\left(Z_{ki}^-\right) - F_{L_k}(Z_{ki}) + F_{Z_k}(Z_{ki})\right)}{n\left(F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})\right)}\right]^2.$$

Therefore

$$\begin{aligned}P\left(\sup_{I\leq u\leq T}|nR_{k3}(u)| > x\right) &\leq P\left(\Gamma_{kn}^c\right) + P\left(\left(\Delta_{kn}(I)\right)^c\right)\\ &+ P\left[\sup_{I\leq u\leq T}\left\{2\sum_{i=1}^{n}\frac{I_{\{Z_{ki}\leq u, A_{ki}=1\}}}{F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})}\right.\right.\\ &\left.\left.\times\left(\frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^-\right) - \widehat{F}_{Z_k}\left(Z_{ki}^-\right) - F_{L_k}(Z_{ki}) + F_{Z_k}(Z_{ki})\right)}{n\left(F_{L_k}(Z_{ki}) - F_{Z_k}(Z_{ki})\right)}\right)^2\right\} > x\right].\end{aligned} \tag{5.11}$$

Moreover, we have for n large enough
$\inf_{\theta_{k1}\le t\le\theta_{k2}}\left\{\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right\}<\frac{2}{n}\Rightarrow\exists t_0\in[\theta_{k1},\theta_{k2}]$ such that $\widehat{F}_{L_k}\left(t_0^-\right)-\widehat{F}_{Z_k}\left(t_0^-\right)<\frac{2}{n}$

$$
\begin{aligned}
\Rightarrow \sup_{\theta_{k1}\le t\le\theta_{k2}}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)-F_{L_k}(t)+F_{Z_k}(t)\right| &\ge F_{L_k}(t_0)-F_{Z_k}(t_0)-\left(\widehat{F}_{L_k}\left(t_0^-\right)-\widehat{F}_{Z_k}\left(t_0^-\right)\right)\\
&\ge F_{L_k}(\theta_{k1})\,S_{X_k}(\theta_{k2})\,S_{R_k}(\theta_{k2})-\frac{2}{n}\\
&> \frac{F_{L_k}(\theta_{k1})\,S_{X_k}(\theta_{k2})\,S_{R_k}(\theta_{k2})}{2}=:\frac{a}{2}.
\end{aligned}
$$

So

$$
\begin{aligned}
P\left(\inf_{\theta_{k1}\le t\le\theta_{k2}}\left(\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right)<\frac{2}{n}\right) &\le P\left(\sqrt{n}\sup_{\theta_{k1}\le t\le\theta_{k2}}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right.\right.\\
&\qquad\left.\left.-F_{L_k}(t)+F_{Z_k}(t)\right|>\frac{a\sqrt{n}}{2}\right)\\
&\le K\exp\{-Cn\}\quad\text{(thanks to relations (5.8) and (5.10))}\\
&\le K\exp\left\{-\lambda\delta^2x\right\}\ \text{ for } n \text{ large enough.}
\end{aligned}
\tag{5.12}
$$

Furthermore, we have

$$
\begin{aligned}
P\left((\Delta_{kn}(I))^c\right) &= P\left(\bigcup_{i=1}^{n}\left(\left\{\frac{\left|\widehat{F}_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)-F_{L_k}(Z_{ki})+F_{Z_k}(Z_{ki})\right|}{F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})}>2\right\}\right.\right.\\
&\qquad\left.\left.\cap\{Z_{ki}>I,A_{ki}=1\}\right)\right)\\
&\le nP\left(\frac{\left|\widehat{F}_{L_k}\left(Z_k^-\right)-\widehat{F}_{Z_k}\left(Z_k^-\right)-F_{L_k}(Z_k)+F_{Z_k}(Z_k)\right|}{F_{L_k}(Z_k)-F_{Z_k}(Z_k)}>2,Z_k>I,A_k=1\right)
\end{aligned}
$$

and for $t>I$, we have

$$
\begin{aligned}
&P\left(\left.\frac{\left|\widehat{F}_{L_k}\left(Z_k^-\right)-\widehat{F}_{Z_k}\left(Z_k^-\right)-F_{L_k}(Z_k)+F_{Z_k}(Z_k)\right|}{F_{L_k}(Z_k)-F_{Z_k}(Z_k)}>2,Z_k>I,A_k=1\right|Z_k=t\right)\\
&=P\left(\left.\frac{\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)-F_{L_k}(t)+F_{Z_k}(t)\right|}{F_{L_k}(t)-F_{Z_k}(t)}>2,Z_k>I,A_k=1\right|Z_k=t\right)\\
&\le P\left(\left|\widehat{F}_{L_k}(t^-)-F_{L_k}(t)\right|>F_{L_k}(t)-F_{Z_k}(t),Z_k>I,A_k=1\mid Z_k=t\right)\\
&+P\left(\left|\widehat{F}_{Z_k}(t^-)-F_{Z_k}(t)\right|>F_{L_k}(t)-F_{Z_k}(t),Z_k>I,A_k=1\mid Z_k=t\right).
\end{aligned}
$$

On the one hand, the Bernstein inequality (see [[18], Corollary $A.9$]) allows to write

$$\begin{aligned}
&P\left(\left|\widehat{F}_{Z_k}\left(t^-\right)-F_{Z_k}\left(t\right)\right|>F_{L_k}\left(t\right)-F_{Z_k}\left(t\right),Z_k>I,A_k=1\mid Z_k=t\right)\\
&=P\left(\left|\sum_{j=1}^{n}\left(I_{\left\{Z_{kj}<t\right\}}-F_{Z_k}\left(t\right)\right)\right|>n\left(F_{L_k}\left(t\right)-F_{Z_k}\left(t\right)\right),Z_k>I,A_k=1\mid Z_k=t\right)\\
&\leq 2\exp\left\{\frac{-2\left(F_{Z_k}\left(t\right)\right)^2n}{F_{Z_k}\left(t\right)S_{Z_k}\left(t\right)\left(1+\frac{4F_{Z_k}(t)}{F_{Z_k}(t)S_{Z_k}(t)}\right)}\right\}\\
&\leq 2\exp\left\{-CnF_{Z_k}\left(t\right)\right\}
\end{aligned}$$

and the probability equals to zero if $t\leq I$.

On the other hand, proceeding as in [[46], proof of Theorem 2], we get

$$\sup_{u\geq I}\left|\widehat{F}_{L_k}\left(u^-\right)-F_{L_k}\left(u\right)\right|\leq C\left[\sup_{u\geq I}\left|\widehat{F}_{Z_k}\left(u^-\right)-F_{Z_k}\left(u\right)\right|+\sup_{u\geq I}\left|H_{kn}^{(2)}\left(u^-\right)-H_k^{(2)}\left(u\right)\right|\right].$$

So, for $t>I$

$$\begin{aligned}
&P\left(\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)\right|>F_{L_k}\left(t\right)-F_{Z_k}\left(t\right),Z_k>I,A_k=1\right|Z_k=t\Big)\\
&\leq P\left(\sup_{u\geq I}\left|\frac{\widehat{F}_{Z_k}\left(u^-\right)-F_{Z_k}\left(u\right)}{F_{Z_k}\left(u\right)}\right|>\frac{b}{2C},Z_k>I,A_k=1\right|Z_k=t\Big)\\
&+P\left(\sup_{u\geq I}\left|\frac{H_{kn}^{(2)}\left(u^-\right)-H_k^{(2)}\left(u\right)}{H_k^{(2)}\left(u\right)}\right|>\frac{b}{2C},Z_k>I,A_k=1\right|Z_k=t\Big)\ (\text{where } b=F_{L_k}(I)S_{X_k}(\theta_{k2})S_{R_k}(\theta_{k2}))\\
&\leq K_1\exp\left\{-C_1n\right\}+K_2\exp\left\{-C_2n\right\}
\end{aligned}$$

(thanks to lemma 3 of [52]; K_1, K_2, C_1 and C_2 being some positive constants)

$$\leq K\exp\left\{-CnF_{Z_k}\left(t\right)\right\}$$

and the probability equals to zero if $t\geq I$. Thus

$$\begin{aligned}
P\left(\left(\Delta_{kn}\left(I\right)\right)^c\right)&\leq Kn\int_I^{+\infty}\exp\left\{-CnF_{Z_k}\left(t\right)\right\}dF_{Z_k}\left(t\right)\\
&\leq K\exp\left\{-\frac{C\delta n}{2}\right\}. \qquad (5.13)
\end{aligned}$$

It remains to deal with the following probability

$$P\left[\sup_{I\le u\le T}\left(2\sum_{i=1}^{n}\frac{I_{\{Z_{ki}\le u,A_{ki}=1\}}}{F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})}\left[\frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)-F_{L_k}(Z_{ki})+F_{Z_k}(Z_{ki})\right)}{n\left(F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})\right)}\right]^2>x\right)\right]$$

$$\le P\left(2\sum_{i=1}^{n}\frac{I_{\{Z_{ki}\le T,A_{ki}=1\}}}{n\left(F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})\right)}\left[\frac{n\left(\widehat{F}_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)-F_{L_k}(Z_{ki})+F_{Z_k}(Z_{ki})\right)}{\sqrt{n}\left(F_{L_k}(Z_{ki})-F_{Z_k}(Z_{ki})\right)}\right]^2>x\right)$$

$$\le P\left(\frac{2}{a^3}\sup_{\theta_{k1}\le u\le\theta_{k2}} n\left(\left(\widehat{F}_{L_k}(u^-)-F_{L_k}(u)\right)-\left(\widehat{F}_{Z_k}(u^-)-F_{Z_k}(u)\right)\right)^2>x\right)$$

$$\le P\left(\frac{2}{C\delta}\sup_{\theta_{k1}\le u\le\theta_{k2}}\left(\sqrt{n}\frac{\widehat{F}_{L_k}(u^-)-F_{L_k}(u)}{F_{L_k}(u)}\right)^2>\frac{x}{2}\right)$$

$$+P\left(\frac{2}{C\delta}\sup_{\theta_{k1}\le u\le\theta_{k2}}\left(\sqrt{n}\frac{\widehat{F}_{Z_k}(u^-)-F_{Z_k}(u)}{F_{Z_k}(u)}\right)^2>\frac{x}{2}\right)$$

$$\le P\left(\frac{2C_1}{C\delta}\sup_{\theta_{k1}\le u\le\theta_{k2}}\left(\sqrt{n}\frac{\widehat{F}_{Z_k}(u^-)-F_{Z_k}(u)}{F_{Z_k}(u)}\right)^2>\frac{x}{4}\right)$$

$$+P\left(\frac{2C_2}{C\delta}\sup_{\theta_{k1}\le u\le\theta_{k2}}\left(\sqrt{n}\frac{H_{kn}^{(2)}(u^-)-H_k^{(2)}(u)}{H_k^{(2)}(u)}\right)^2>\frac{x}{4}\right)$$

$$+P\left(\frac{2}{C\delta}\sup_{\theta_{k1}\le u\le\theta_{k2}}\left(\sqrt{n}\frac{\widehat{F}_{Z_k}(u^-)-F_{Z_k}(u)}{F_{Z_k}(u)}\right)^2>\frac{x}{2}\right)\ (C_1\text{ and }C_2\text{ are positive constants})$$

$$\le K\exp\left\{-\lambda\delta^2x\right\}\quad\text{(thanks to [[52], Lemma 3])}.$$

Combining this with (5.11),(5.12) and (5.13) gives the claimed result. ■

Proof of Lemma 5.5 : We have $nR_{k4}(u)=J_{k1}(u)+J_{k2}(u)$, where

$$J_{k1}(u)=\int_0^{+\infty}\int_0^{+\infty}\frac{I_{\{y\le u\}}I_{\{x<y\}}}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}d\left[\sqrt{n}\left(H_{nk}^{(1)}(y)-H_k^{(1)}(y)\right)\right]d\left[\sqrt{n}\left(\widehat{F}_{Z_k}(x^-)-F_{Z_k}(x)\right)\right]$$

and

$$J_{k2}(u)=\int_0^{+\infty}\int_0^{+\infty}\frac{I_{\{y\le u\}}I_{\{x<y\}}}{\left(F_{L_k}(y)-F_{Z_k}(y)\right)^2}d\left[\sqrt{n}\left(H_{nk}^{(1)}(y)-H_k^{(1)}(y)\right)\right]d\left[\sqrt{n}\left(\widehat{F}_{L_k}(x^-)-F_{L_k}(x)\right)\right].$$

So

$$P\left(\sup_{I\le u\le T}\left|nR_{k4}(u)\right|>x\right)\le P\left(\sup_{I\le u\le T}\left|J_{k1}(u)\right|>\frac{x}{2}\right)+P\left(\sup_{I\le u\le T}\left|J_{k2}(u)\right|>\frac{x}{2}\right).\tag{5.14}$$

On the one hand, we can prove as in [[36], Lemma 3] that

$$P\left(\sup_{I\le u\le T}n\left|J_{k1}(u)\right|>\frac{x}{2}\right)\le K\exp\left\{-\lambda\delta^2x\right\}.\tag{5.15}$$

On the other hand, we have for $I \leq u \leq T$

$$J_{k2}(u) = \int_0^{+\infty} \frac{\widehat{F}_{L_k}(y^-) - F_{L_k}(y)}{\left(F_{L_k}(y) - F_{Z_k}(y)\right)^2} I_{\{y \leq u\}} d\left(H_{nk}^{(1)}(y) - H_k^{(1)}(y)\right).$$

Therefore

$$\begin{aligned} |J_{k2}(u)| &\leq \frac{1}{a^2} \sup_{\theta_{k1} \leq u \leq \theta_{k2}} \left|\widehat{F}_{L_k}(u^-) - F_{L_k}(u)\right| \left|H_{nk}^{(1)}(u) - H_k^{(1)}(u)\right| \\ &\leq \frac{1}{a^2} \sup_{\theta_{k1} \leq u \leq \theta_{k2}} \left|\widehat{F}_{L_k}(u^-) - F_{L_k}(u)\right| \sup_{I \leq u \leq T} \left|H_{nk}^{(1)}(u) - H_k^{(1)}(u)\right| \end{aligned}$$

which implies that

$$\begin{aligned} P\left(\sup_{I \leq u \leq T} n |J_{k2}(u)| > \frac{x}{2}\right) &\leq P\left(\sqrt{n} \sup_{\theta_{k1} \leq u \leq \theta_{k2}} \left|\widehat{F}_{L_k}(u^-) - F_{L_k}(u)\right| > \sqrt{\frac{ax}{2}}\right) \\ &\quad + P\left(\sqrt{n} \sup_{I \leq u \leq T} \left|H_{nk}^{(1)}(u) - H_k^{(1)}(u)\right| > \sqrt{\frac{ax}{2}}\right) \\ &\leq K \exp\{-Cx\} \ \text{(thanks to relations (5.9) and (5.10))} \\ &\leq K \exp\left\{-\lambda \delta^2 x\right\}. \end{aligned}$$

This together with (5.14) and (5.15) gives the claimed result. ■

Proof of Theorem4.2-*i*) : Using the following lemma, theorem 4.2-*i*) can be proved in the same way as in [[35], Theorem 3], for the class of functions

$$\mathcal{F} = \left\{(t_1, t_2) \mapsto I_{[0,x_1]\times[0,x_2]}(t_1, t_2),\ x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\right\}.$$

Lemma 5.6

*Under assumptions **H1-H6**, we have for all* $\varepsilon > 0$

$$\max_{1 \leq i \leq n} \left| \frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}}}{C_L\left(\widehat{F}_{L_1}(Z_{1i}), \widehat{F}_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(\widehat{S}_{R_1}(Z_{1i}), \widehat{S}_{R_2}(Z_{2i})\right)} - \frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}}}{C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}\right.}\right.$$

$$\begin{aligned} &\leq \mathcal{M}_n \left(\frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}} F_{L_1}^{1-\alpha_1}(Z_{1i}) \mathcal{K}_{L_1}^{1/2+\varepsilon}(Z_{1i})}{F_{L_2}^{\alpha_2}(Z_{2i}) S_{R_1}^{\beta_1}(Z_{1i}) S_{R_2}^{\beta_2}(Z_{2i}) C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}(Z_{2i})\right)} \right. \\ &+ \frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}} F_{L_2}^{1-\alpha_2}(Z_{2i}) \mathcal{K}_{L_2}^{1/2+\varepsilon}(Z_{2i})}{F_{L_1}^{\alpha_1}(Z_{1i}) S_{R_1}^{\beta_1}(Z_{1i}) S_{R_2}^{\beta_2}(Z_{2i}) C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}(Z_{2i})\right)} \\ &+ \frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}} S_{R_1}^{1-\beta_1}(Z_{1i}) \mathcal{K}_{R_1}^{1/2+\varepsilon}(Z_{1i})}{F_{L_1}^{\alpha_1}(Z_{1i}) F_{L_2}^{\alpha_2}(Z_{2i}) S_{R_2}^{\beta_2}(Z_{2i}) C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}(Z_{2i})\right)} \\ &+ \left. \frac{I_{\{A_{1i}=1\}} I_{\{A_{2i}=1\}} S_{R_2}^{1-\beta_2}(Z_{2i}) \mathcal{K}_{R_2}^{1/2+\varepsilon}(Z_{2i})}{F_{L_1}^{\alpha_1}(Z_{1i}) F_{L_2}^{\alpha_2}(Z_{2i}) S_{R_1}^{\beta_1}(Z_{1i}) C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}(Z_{2i})\right)} \right), \end{aligned}$$

with $\mathcal{M}_n = O_P\left(n^{-1/2}\right)$.

This lemma is the equivalent of [[35], Lemma 7.2] in the case of twice censoring.

Proof of Lemma 5.6 : Let $Z_{(kn)} = \max_{k \leq i \leq n} Z_{ki}$. Proceeding as in [[35], Lemma 7.2], we obtain

$$\left| \frac{I_{\{A_{1i}=1\}}\, I_{\{A_{2i}=1\}}}{C_L\left(\widehat{F}_{L_1}(Z_{1i}), \widehat{F}_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(\widehat{S}_{R_1}(Z_{1i}), \widehat{S}_{R_2}(Z_{2i})\right)} - \frac{I_{\{A_{1i}=1\}}\, I_{\{A_{2i}=1\}}}{C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}\right.}\right.$$

$$\leq \frac{M\; I_{\{A_{1i}=1\}}\, I_{\{A_{2i}=1\}}}{C_L\left(F_{L_1}(Z_{1i}), F_{L_2}(Z_{2i})\right) \widetilde{C}_R\left(S_{R_1}(Z_{1i}), S_{R_2}(Z_{2i})\right)} \times \left[\frac{\left|\widehat{F}_{L_1}(Z_{1i}) - F_{L_1}(Z_{1i})\right|}{\widehat{F}_{L_1}^{\alpha_1}(Z_{1i})\, \widehat{F}_{L_2}^{\alpha_2}(Z_{2i})\, \widehat{S}_{R_1}^{\beta_1}(Z_{1i})\, \widehat{S}_{R_2}^{\beta_2}(Z_{2i})} \right.$$

$$+ \frac{\left|\widehat{F}_{L_2}(Z_{2i}) - F_{L_2}(Z_{2i})\right|}{\widehat{F}_{L_1}^{\alpha_1}(Z_{1i})\, \widehat{F}_{L_2}^{\alpha_2}(Z_{2i})\, \widehat{S}_{R_1}^{\beta_1}(Z_{1i})\, \widehat{S}_{R_2}^{\beta_2}(Z_{2i})} + \frac{\left|\widehat{S}_{R_1}(Z_{1i}) - S_{R_1}(Z_{1i})\right|}{\widehat{F}_{L_1}^{\alpha_1}(Z_{1i})\, \widehat{F}_{L_2}^{\alpha_2}(Z_{2i})\, \widehat{S}_{R_1}^{\beta_1}(Z_{1i})\, \widehat{S}_{R_2}^{\beta_2}(Z_{2i})}$$

$$\left. + \frac{\left|\widehat{S}_{R_2}(Z_{2i}) - S_{R_2}(Z_{2i})\right|}{\widehat{F}_{L_1}^{\alpha_1}(Z_{1i})\, \widehat{F}_{L_2}^{\alpha_2}(Z_{2i})\, \widehat{S}_{R_1}^{\beta_1}(Z_{1i})\, \widehat{S}_{R_2}^{\beta_2}(Z_{2i})} \right],$$

where M is a positive constant. So, to prove the lemma, we have to show as in [35] that for $k \in \{1, 2\}$

$$\sup_{t \geq \theta_{k_1}} \frac{F_{L_k}(t)}{\widehat{F}_{L_k}(t)} = O_P(1), \tag{5.16}$$

$$\sup_{t \geq \theta_{k_1}} \frac{\left|\widehat{F}_{L_k}(t) - F_{L_k}(t)\right|}{\mathcal{K}_{L_k}^{1/2+\varepsilon}(t) F_{L_k}(t)} = O_P\left(\frac{1}{\sqrt{n}}\right), \tag{5.17}$$

$$\sup_{t \leq \theta_{k_2}} \frac{S_{R_k}(t)}{\widehat{S}_{R_k}(t)} = O_P(1) \tag{5.18}$$

and

$$\sup_{t \leq Z_{(k_n)}} \frac{\left|\widehat{S}_{R_k}(t) - S_{R_k}(t)\right|}{\mathcal{K}_{R_k}^{1/2+\varepsilon}(t) S_{R_k}(t)} = O_P\left(\frac{1}{\sqrt{n}}\right). \tag{5.19}$$

Relation (5.16) follows from the fact that for $t \geq \theta_{k_1}$

$$\frac{F_{L_k}(t)}{\widehat{F}_{L_k}(t)} \leq \frac{1}{\widehat{F}_{L_k}(\theta_{k_1})} \leq \frac{2}{F_{L_k}(\theta_{k_1})} \; a.s. \text{ for n large enough.}$$

Relation (5.17) can be proved in the same way as in [[22], Theorem 2.1].

Relation (5.18) follows from the fact that for $t \leq \theta_{k_2}$

$$\frac{S_{R_k}(t)}{\widehat{S}_{R_k}(t)} \leq \frac{1}{\widehat{S}_{R_k}(\theta_{k_2})} \leq \frac{2}{S_{R_k}(\theta_{k_2})} \; a.s. \text{ for n large enough.}$$

It remains to deal with relation (5.19). Set

$$\xi_{nk}(t) = \sqrt{n}\left(\frac{\widehat{S}_{R_k}(t) - S_{R_k}(t)}{S_{R_k}(t)}\right)$$

and

$$h(t)=\frac{1}{\mathcal{K}_{R_k}^{1/2+\varepsilon}(t)}.$$

It view of [[43], Theorem 7.3], the process $\xi_{nk}(t)$ converges weakly to a centered Gaussian process in $l^{\infty}([0,\tau])$, for any τ such that $\theta_{k_2}<\tau<T_{Z_k}$. So, relation (5.19) can be proved as in [[22], Theorem 2.1]. In fact, it suffices to prove that for all $\varepsilon>0$

$$\lim_{\tau\uparrow T_{Z_k}}\limsup_{n\to\infty}P\left(\sup_{\tau\le t\le Z_{(kn)}}\left|\int_{\tau}^{t}h(s)\,d\xi_{nk}(s)\right|>\varepsilon\right)=0. \tag{5.20}$$

For that, we set

$$F_{L_k}^{*}(t)=\frac{1}{n}\sum_{i=1}^{n}I_{\{L_{ki}\le t\}}$$

and

$$S_{R_k}^{*}(t)=\prod_{i/Z_{ki}\le t}\left(1-\frac{I_{\{A_{k_i}=1\}}}{n\left(F_{L_k}^{*}\left(Z_{ki}^{-}\right)-\widehat{F}_{Z_k}\left(Z_{ki}^{-}\right)\right)}\right).$$

We have $\xi_{nk}(t)=\xi_{nk}^{*}(t)+R_{nk}^{*}(t)$, where

$$\xi_{nk}^{*}(t)=\sqrt{n}\left(\frac{S_{R_k}^{*}(t)-S_{R_k}(t)}{S_{R_k}(t)}\right)$$

and

$$R_{nk}^{*}(t)=\sqrt{n}\left(\frac{\widehat{S}_{R_k}(t)-S_{R_k}^{*}(t)}{S_{R_k}(t)}\right).$$

Therefore

$$\begin{aligned}P\left(\sup_{\tau\le t\le Z_{(kn)}}\left|\int_{\tau}^{t}h(s)\,d\xi_{nk}(s)\right|>\varepsilon\right)&\le P\left(\sup_{\tau\le t\le Z_{(kn)}}\left|\int_{\tau}^{t}h(s)\,d\xi_{nk}^{*}(s)\right|>\varepsilon/2\right)\\&\quad+P\left(\sup_{\tau\le t\le Z_{(kn)}}\left|\int_{\tau}^{t}h(s)\,dR_{nk}^{*}(s)\right|>\varepsilon/2\right)\\&=:P_1+P_2\end{aligned} \tag{5.21}$$

We start by dealing with P_1. For that, we need the following lemma. Note $\mathcal{F}_{kt}$ the filtration defined by

$$\begin{aligned}\mathcal{F}_{kt}=\mathcal{N}\vee\sigma\Big(\Big\{&I_{\{X_{ki}\le s\}},I_{\{L_{ki}\le s\}},I_{\{R_{ki}\le s\}},I_{\{X_{ki}\le s,A_{ki}=1\}},I_{\{L_{ki}\le s,A_{ki}=1\}},I_{\{R_{ki}\le s,A_{ki}=1\}},\\&0<s\le t,1\le i\le n\Big\}\Big),\end{aligned}$$

where $\mathcal{N}$ is the family of negligible sets.

Lemma 5.7

We have

i) $\xi_{nk}^{*}(t)$ *is an* $\mathcal{F}_{kt}-$ *martingale.*

ii) $\forall\beta\in\,]0,1[\,,\ P\left(\sup_{t\leq Z_{(kn)}}\frac{S_{R_k}^{*}(t)}{S_{R_k}(t)}\leq\frac{1}{\beta}\right)\geq 1-\beta.$

iii) $\sup_{I_{R_k}\leq t\leq Z_{(kn)}}\left|\frac{F_{L_k}(t)-F_{Z_k}(t)}{F_{L_k}^{*}(t^{-})-\widehat{F}_{Z_k}(t^{-})}\right|=O_P(1).$

Proof of Lemma 5.7 : *i)* Set

$$\Lambda_{R_k}^{*}(t)=\int_0^t\frac{dH_{nk}^{(1)}(u)}{F_{L_k}^{*}(u^{-})-\widehat{F}_{Z_k}(u^{-})}.$$

In view of [[21], Proposition A.4.1], we have

$$\frac{S_{R_k}^{*}(t)}{S_{R_k}(t)}=1-\int_0^t\frac{S_{R_k}^{*}(u^{-})}{S_{R_k}(u)}d\left(\Lambda_{R_k}^{*}(u)-\Lambda_{R_k}(u)\right). \tag{5.22}$$

Moreover, consider the $\mathcal{F}_{kt}-$martingale

$$M_k(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[I_{\{Z_{ki}\leq t,A_{ki}=1\}}-\int_0^t I_{\{L_{k_i}\leq u,Z_{ki}\geq u\}}d\Lambda_{R_k}(u)\right].$$

Since $d\left(\Lambda_{R_k}^{*}(u)-\Lambda_{R_k}(u)\right)=\frac{\sqrt{n}dM_k(u)}{n\left(F_{L_k}^{*}(u^{-})-\widehat{F}_{Z_k}(u^{-})\right)}$, relation (5.22) implies that

$$\frac{S_{R_k}^{*}(t)}{S_{R_k}(t)}=1-\sqrt{n}\int_0^t\frac{S_{R_k}^{*}(u^{-})}{nS_{R_k}(u)\left(F_{L_k}^{*}(u^{-})-\widehat{F}_{Z_k}(u^{-})\right)}dM_k(u) \tag{5.23}$$

which implies that

$$\xi_{nk}^{*}(t)=-\int_0^t\frac{S_{R_k}^{*}(u^{-})}{S_{R_k}(u)\left(F_{L_k}^{*}(u^{-})-\widehat{F}_{Z_k}(u^{-})\right)}dM_k(u).$$

Since $\frac{S_{R_k}^{*}(u^{-})}{S_{R_k}(u)\left(F_{L_k}^{*}(u^{-})-\widehat{F}_{Z_k}(u^{-})\right)}$ is predictable with respect to $\mathcal{F}_{kt}$, Theorem 1 page 890 of [48] shows that $\xi_{nk}^{*}(t)$ is an $\mathcal{F}_{kt}-$ martingale.

ii) Using [[48], Theorem 1 page 890], relation (5.23) shows that $\frac{S_{R_k}^{*}(t)}{S_{R_k}(t)}$ is an $\mathcal{F}_{kt}-$martingale. So, the claimed result can be proved in the same way as in [[21], Theorem 3.2.1].

iii) Set $Y_{nk}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{ki}\geq t, R_{ki}\geq t\}}$ and $e_{nk}(t) = F^*_{L_k}(t^-) - \widehat{F}_{Z_k}(t^-) - F^*_{L_k}(t^-)\, Y_{nk}(t)$.

We have for all $t \in \left[I_{R_k}, Z_{(kn)}\right]$

$$F^*_{L_k}(t^-) - \widehat{F}_{Z_k}(t^-) \geq F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t).$$

So

$$\frac{1}{F^*_{L_k}(t^-) - \widehat{F}_{Z_k}(t^-)} \leq \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t)}. \tag{5.24}$$

Furthermore, we have

$$\begin{aligned}
&\sup_{I_{R_k}\leq t\leq Z_{(kn)}} \left| \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t)} - \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t)} \right| \\
&= \sup_{I_{R_k}\leq t\leq Z_{(kn)}} \left| \frac{e_{nk}(t)}{\left(F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t)\right) F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t)} \right|.
\end{aligned}$$

Since for $t > T_{L_k}$, we have $F^*_{L_k}(t^-) = 1$, so

$$\begin{aligned}
e_{nk}(t) &= 1 - \widehat{F}_{Z_k}(t^-) - Y_{nk}(t) \\
&= \frac{1}{n}\sum_{i=1}^{n} I_{\left\{X_{k_i}\wedge R_{k_i}\geq t\right\}} - \frac{1}{n}\sum_{i=1}^{n} I_{\left\{X_{k_i}\geq t, R_{k_i}\geq t\right\}} \\
&= 0 \quad a.s.
\end{aligned}$$

$\left(I_{\left\{Z_{k_i}\geq t\right\}} = I_{\left\{X_{k_i}\wedge R_{k_i}\geq t\right\}}\ a.s. \text{ because } L_{ki} < t\ \ a.s. \text{ since } t > T_{L_k}\right).$
Therefore

$$\begin{aligned}
&\sup_{I_{R_k}\leq t\leq Z_{(kn)}} \left| \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t)} - \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t)} \right| \\
&= \sup_{I_{R_k}\leq t\leq T_{L_k}} \left| \frac{e_{nk}(t)}{F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) \left(F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) + e_{nk}(t)\right)} \right|.
\end{aligned}$$

Moreover, for $I_{R_k} \leq t \leq T_{L_k}$, we have for n large enough

$$\begin{aligned}
F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}(t) &\geq F^*_{L_k}\left(I^-_{R_k}\right) Y_{nk}\left(T_{L_k}\right) \\
&\geq \frac{F_{L_k}\left(I_{R_k}\right) S_{X_k}\left(T_{L_k}\right) S_{R_k}\left(T_{L_k}\right)}{2} \\
&=: \alpha
\end{aligned}$$

and $F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)+e_{nk}(t)\geq \alpha+e_{nk}(t)>0$, for n large enough (since $\sup\limits_{I_{R_k}\leq t\leq T_{L_k}}|e_{nk}(t)|=o_{a.s.}(1)$). Thus

$$\begin{aligned}\sup_{I_{R_k}\leq t\leq Z_{(kn)}}\left|\frac{1}{F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)+e_{nk}(t)}-\frac{1}{F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)}\right| &\leq \sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\frac{e_{nk}(t)}{\alpha(\alpha+e_{nk}(t))}\right|\\ &\leq \frac{2}{\alpha^2}\sup_{I_{R_k}\leq t\leq T_{L_k}}|e_{nk}(t)| \text{ for } n \text{ large enough}\\ &= o_{a.s.}(1).\end{aligned}$$

So relation (5.24) implies that

$$\frac{1}{F^*_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)}\leq \frac{1}{F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)}+o_{a.s.}(1)$$

where the $o_{a.s.}(1)$ is uniform on $t\in\left[I_{R_k},Z_{(kn)}\right]$. So

$$\begin{aligned}\frac{F_{L_k}(t)-F_{Z_k}(t)}{F^*_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)} &\leq \frac{F_{L_k}(t)S_{X_k}(t)S_{R_k}(t)}{F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)}+o_{a.s.}(1)\\ &\leq \frac{S_{X_k}(t)S_{R_k}(t)}{F^*_{L_k}\left(I^-_{R_k}\right)Y_{nk}(t)}+o_{a.s.}(1).\end{aligned}\tag{5.25}$$

Since

$$\frac{1}{F^*_{L_k}\left(I^-_{R_k}\right)}\xrightarrow{a.s.}\frac{1}{F_{L_k}(I_{R_k})},\quad \text{as } n\to\infty$$

we have

$$\frac{1}{F^*_{L_k}\left(I^-_{R_k}\right)}=O_P(1)\tag{5.26}$$

and

$$\sup_{I_{R_k}\leq t\leq Z_{(kn)}}\left|\frac{S_{X_k}(t)S_{R_k}(t)}{Y_{nk}(t)}\right|=O_P(1)$$

(see [[52], Remark 1 (ii)]). Combining this with (5.25) and (5.26) gives the claimed result. This ends the proof of Lemma 5.7. ■

Using this lemma, we can show that

$$\lim_{\tau\uparrow T_{Z_k}}\limsup_{n\to\infty}P\left(\sup_{\tau\leq t\leq Z_{(kn)}}\left|\int_\tau^t h(s)\,d\xi^*_{nk}(s)\right|>\varepsilon/2\right)=0\tag{5.27}$$

in the same way as in [[22], Theorem 2.1], see also [[1], Proposition 3].

Now, we deal with the probability P_2. We have for $\tau \leq t \leq Z_{(kn)}$

$$\begin{aligned}\left|\frac{\widehat{S}_{R_k}(t)-S^*_{R_k}(t)}{S_{R_k}(t)}\right| &\leq \frac{1}{S_{R_k}(t)}\sum_{i/Z_{ki}\leq t}\left|\frac{I_{\{A_{ki}=1\}}}{n}\left[\frac{1}{\widehat{F}_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)}-\frac{1}{F^*_{L_k}\left(Z_{ki}^-\right)-\widehat{F}_{Z_k}\left(Z_{ki}^-\right)}\right]\right| \\ &\leq \sup_{\tau\leq t\leq Z_{(kn)}}\left|\frac{\widehat{F}_{L_k}(t^-)-F^*_{L_k}(t^-)}{S_{R_k}(t)\left(\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right)\left(F^*_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right)}\right|\times\frac{1}{n}\sum_{i=1}^{n}I_{\{A_{ki}=1\}} \\ &\leq \sup_{\tau\leq t\leq T_{L_k}}\left|\frac{\widehat{F}_{L_k}(t^-)-F^*_{L_k}(t^-)}{S_{R_k}(t)\left(\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right)\left(F^*_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right)}\right|\end{aligned}$$

(since for $t > T_{L_k}$, $\widehat{F}_{L_k}(t^-) = F^*_{L_k}(t^-) = 1$).

Therefore

$$\begin{aligned}\sqrt{n}\sup_{\tau\leq t\leq Z_{(kn)}}\left|\frac{\widehat{S}_{R_k}(t)-S^*_{R_k}(t)}{S_{R_k}(t)}\right| &\leq \frac{1}{S_{R_k}(T_{L_k})} \\ &\times\frac{\sqrt{n}\sup_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}(t^-)-F^*_{L_k}(t^-)\right|}{\inf_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right|\inf_{\tau\leq t\leq T_{L_k}}\left|F^*_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right|}.\end{aligned} \tag{5.28}$$

Moreover, since $\sqrt{n}\left(\widehat{F}_{L_k}-F_{L_k}\right)$ and $\sqrt{n}\left(F^*_{L_k}-F_{L_k}\right)$ converge weakly in $l^\infty\left([\tau, T_{L_k}]\right)$, we get

$$\sqrt{n}\sup_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}(t^-)-F^*_{L_k}(t^-)\right| = O_P(1) \tag{5.29}$$

and we have for all $\tau \leq t \leq T_{L_k}$

$$\begin{aligned}&\sup_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)-\left(F_{L_k}(t)-F_{Z_k}(t)\right)\right| \\ &\geq F_{L_k}(t)-F_{Z_k}(t)-\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right| \\ &\geq F_{L_k}(I_{R_k})\,S_{X_k}(T_{L_k})\,S_{R_k}(T_{L_k})-\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right|.\end{aligned}$$

Thus

$$\begin{aligned}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)\right| &\geq F_{L_k}(I_{R_k})\,S_{X_k}(T_{L_k})\,S_{R_k}(T_{L_k}) \\ &\quad-\sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}(t^-)-\widehat{F}_{Z_k}(t^-)-\left(F_{L_k}(t)-F_{Z_k}(t)\right)\right|\end{aligned}$$

and

$$\inf_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)\right|\geq F_{L_k}\left(I_{R_k}\right)S_{X_k}\left(T_{L_k}\right)S_{R_k}\left(T_{L_k}\right)$$
$$-\sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)-\left(F_{L_k}\left(t\right)-F_{Z_k}\left(t\right)\right)\right|$$
$$=:\beta-\sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)-\left(F_{L_k}\left(t\right)-F_{Z_k}\left(t\right)\right)\right|.$$

Therefore

$$\frac{1}{\inf_{\tau\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)\right|}\leq\frac{1}{\beta-\sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)-\left(F_{L_k}\left(t\right)-F_{Z_k}\left(t\right)\right)\right|}=O_P\left(1\right)$$

because

$$\sup_{I_{R_k}\leq t\leq T_{L_k}}\left|\widehat{F}_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)-\left(F_{L_k}\left(t\right)-F_{Z_k}\left(t\right)\right)\right|\xrightarrow{a.s.}0,\quad\text{as } n\rightarrow\infty$$

and with the same manner we can show that

$$\frac{1}{\inf_{\tau\leq t\leq T_{L_k}}\left|F^*_{L_k}\left(t^-\right)-\widehat{F}_{Z_k}\left(t^-\right)\right|}=O_P\left(1\right).$$

Combining this with (5.28), (5.29) and (**??**), we obtain

$$\sup_{\tau\leq t\leq Z_{(kn)}}\left|R^*_{nk}\left(t\right)\right|=O_P\left(1\right).$$

Moreover, using an integration by parts we can write

$$\left|\int_\tau^t h\left(s\right)dR^*_{nk}\left(s\right)\right|=\left|h\left(t\right)R^*_{nk}\left(t\right)-h\left(\tau\right)R^*_{nk}\left(\tau\right)-\int_\tau^t R^*_{nk}\left(s\right)dh\left(s\right)\right|$$
$$\leq 2\sup_{\tau\leq s\leq Z_{(kn)}}\left|h\left(s\right)\right|\sup_{\tau\leq s\leq Z_{(kn)}}\left|R^*_{nk}\left(s\right)\right|+\sup_{\tau\leq s\leq Z_{(kn)}}\left|R^*_{nk}\left(s\right)\right|\left|h\left(t\right)-h\left(\tau\right)\right|$$
$$\leq 4\sup_{\tau\leq s\leq Z_{(kn)}}\left|h\left(s\right)\right|\sup_{\tau\leq s\leq Z_{(kn)}}\left|R^*_{nk}\left(s\right)\right|.$$

So

$$P\left(\sup_{\tau\leq t\leq Z_{(kn)}}\left|\int_\tau^t h\left(s\right)dR^*_{nk}\left(s\right)\right|>\varepsilon/2\right)\leq P\left(\sup_{\tau\leq t\leq T_{Z_k}}\left|h\left(t\right)\right|\sup_{\tau\leq t\leq Z_{(kn)}}\left|R^*_{nk}\left(t\right)\right|>\varepsilon/8\right).\tag{5.30}$$

It remains to show that

$$\lim_{t\uparrow T_{Z_k}}\limsup_{n\rightarrow\infty}P\left(\sup_{\tau\leq t\leq T_{Z_k}}\left|h\left(t\right)\right|\sup_{\tau\leq s\leq Z_{(kn)}}\left|R^*_{nk}\left(t\right)\right|>\varepsilon/8\right)=0.$$

This is equivalent to

$$\forall \delta > 0, \exists \eta_\delta > 0 : |t - T_{Z_k}| < \eta_b \Rightarrow \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le T_{Z_k}} |h(t)| \sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > \varepsilon/8 \right) \le \delta.$$

Let $\delta > 0$, since $\sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| = O_P(1)$, there exist $b_\delta > 0$ and $m_\delta \in \mathbb{N}^*/\forall m \ge m_\delta$

$$P\left(\sup_{\tau\le t\le Z_{(kn)}} |R^*_{mk}(t)| > b_\delta \right) < \delta$$

$$\Rightarrow \sup_{m\ge m_\delta} P\left(\sup_{\tau\le t\le Z_{(kn)}} |R^*_{mk}(t)| > b_\delta \right) \le \delta$$

$$\Rightarrow \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > b_\delta \right) \le \delta$$

and since $\lim_{\tau\uparrow T_{Z_k}} \sup_{\tau\le t\le T_{Z_k}} |h(t)| = 0$, there exists $\eta_\delta > 0$ such that

$$|\tau - T_{Z_k}| < \eta_\delta \Rightarrow \sup_{\tau\le t\le T_{Z_k}} |h(t)| \le \frac{\varepsilon}{8b_\delta}.$$

So, for τ such that $|\tau - T_{Z_k}| < \eta_\delta$, we have

$$P\left(\sup_{\tau\le t\le T_{Z_k}} |h(t)| \sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > \varepsilon/8 \right) \le P\left(\sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > b_\delta \right)$$

$$\Rightarrow \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le T_{Z_k}} |h(t)| \sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > \varepsilon/8 \right) \le \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > b_\delta \right) \le \delta.$$

Thus

$$\lim_{\tau\uparrow T_{Z_k}} \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le T_{Z_k}} |h(t)| \sup_{\tau\le t\le Z_{(kn)}} |R^*_{nk}(t)| > \varepsilon/8 \right) = 0$$

and relation (5.30) gives

$$\lim_{\tau\uparrow T_{Z_k}} \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le Z_{(kn)}} \left| \int_\tau^t h(s)\, dR^*_{nk}(s) \right| > \varepsilon/2 \right) = 0.$$

Combining this with (5.21) and (5.27) shows that relation (5.20) is satisfied which ends the proof.

■

■

Proof of Theorem 4.3 : Using theorem 4.2, this theorem can be proved in the same way as [[23], Theorem 2] i.e The first step of the proof consists of reducing the problem to the case where the marginals T_1 and T_2 are uniformly distributed due to Assumptions $H1 - H6$ and Lemma 4.2. ■

Bibliography

[1] Akritas, M. G., *The central limit theorem under censoring*, Bernoulli 6, 1109-1120, (2000).

[2] Andersen, P. K., Ekstrøm, C. T., Klein, J. P., Shu, Y., and Zhang, M.-J., *A class of goodness of fit tests for a copula based on bivariate right-censored data*, Biometrical Journal 47, 815-824, (2005).

[3] Aouicha, L. and Messaci, F., *Kernel estimation of the conditional density under a censorship model*, Statistics and Probability Letters 145, 173-180, (2019).

[4] Imane, B. E. N. E. L. M. I. R. Modélisation de la Dépendance par les Copules. Diss. Université de Biskra, 2018.

[5] BitouzÃ©, D., Laurent, B. and Massart, P., *A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator*, Annales de l'Institut Henri Poincare (B) Probability and Statistics 35, 735-763, (1999).

[6] Boukeloua, M., *Rates of Mean Square Convergence of Density and Failure Rate Estimators Under Twice Censoring*, Statistics and Probability Letters 106, 121-128, (2015).

[7] Boukeloua, M., *Study of semiparametric copula models via divergences with bivariate censored data*, Communications in Statistics Theory and Methods 50, 5429-5452, (2021).

[8] Boukeloua, M. and Messaci, F., *Asymptotic normality of kernel estimators based upon incomplete data*, Journal of Nonparametric Statistics 28, 469-486, (2016).

[9] S. Bouzebda, and A. Keziou, it New estimates and tests of independence in semi-parametric copula models, Kybernetika, no. 1, 178–201. (2010)

[10] Bouzebda, S. and Zari, T., *Strong approximation of empirical copula processes by Gaussian processes*, Statistics 47, 1047-1063, (2013).

[11] Breslow, N. and Crowley, J., *A large sample study of the life table and product limit estimates under random censorship*, The Annals of Statistics 2, 437-453, (1974).

[12] DEHEUVELS, P. *P*roprietes d'existence et proprietes topologiques des fonctions de dependance avec applications à la convergence des types pour des lois multivariées (1979).

[13] Deheuvels, P., *Nonparametric tests of independence*, In Nonparametric Asymptotic Statistics (Proceedings of the Conference held in Rouen in 1979), J.-P. Raoult, ed. NewYork: Springer 95-107, (1980).

[14] Deheuvels, P., *A Kolmogorov-Smirnov type test for independence and multivariate samples*, Rev. Roumaine Math. Pures Appl. 26, 213-226, (1981).

[15] Deheuvels, P., *A multivariate Bahadurâ"Kiefer representation for the empirical copula process*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI). 364, 120-147 (2009).

[16] Dvoretzky, A., Kiefer, J. and Wolfowitz, J., *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, The Annals of Mathematical Statistics 642-669, (1956).

[17] Fermanian, J.-D., Radulovic, D. and Wegkamp, M., *Weak convergence of empirical copula processes*, Bernoulli 10, 847-860, (2004).

[18] Ferraty, F. and Vieu, P., *Nonparametric functional data analysis: Theory and practice*, Springer, Berlin, (2006).

[19] Genest, C., & Rivest, L. P. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American statistical Association, 88(423), 1034-1043 (1993).

[20] Genest, C., Ghoudi, K., & Rivest, L. P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3), 543-552 (1995).

[21] Gill, R. D., *Censoring and stocasthic integrals*, Mathematisch Centrum tracts, Amsterdam, (1980).

[22] Gill, R. D., *Large sample behaviour of the product-limit estimator on the whole line*, The Annals of Statistics 11, 49-58, (1983).

[23] Gribkova, S. and Lopez, O., *Non-parametric Copula estimation under bivariate censoring*, Scandinavian Journal of Statistics 42, 925-946, (2015).

[24] Gribkova, S., Lopez, O. and Saint-Pierre, P., *A simplified model for studying bivariate mortality under right-censoring*, Journal of Multivariate Analysis 115, 181-192, (2013).

[25] IDIOU, Nesrine. MULTI-PARAMETRIC COPULA ESTIMATION BASED ON MOMENTS METHOD UNDER CENSORING. Diss. Université de mohamed kheider biskra, 2022.

[26] Idiou, N. and Benatia, F., *Survival copula parameters estimation for Archimedean family under singly censoring*, Advances in Mathematics: Scientific Journal 10, 1-4, (2021).

[27] Idiou, N., Benatia, F., and Brahimi, B., *A semi-parametric estimation of copula models based on moments method under right censoring*, Journal of TWMS J. App and Eng. Math. (To appear).

[28] Idiou, N., Benatia, F. and Mesbah, M., *Copulas and frailty models in multivariate survival data*, Journal of Biostatistics and Health Sciences. ISTE OpenScience, BHS 2, 13-39, (2021).

[29] Kaplan, E. L., & Meier, P. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481 (1958).

[30] Kebabi, K. and Messaci, F. *Rate of the almost complete convergence of a kernel regression estimate with twice censored data*, Statistics and Probability Letters 82, 1908-1913, (2012).

[31] Kiefer, J., *On large deviations of the empiric df of vector chance variables and a law of the iterated logarithm*, Pacific Journal of Mathematics 11, 649-660, (1961).

[32] Kitouni, A., Boukeloua, M. and Messaci, F., *Rate of strong consistency for nonparametric estimators based on twice censored data*, Statistics and Probability Letters 96, 255-261, (2015).

[33] Stephens, M. A. Introduction to Kolmogorov (1933) on the empirical determination of a distribution. In Breakthroughs in Statistics: Methodology and Distribution (pp. 93-105). New York, NY: Springer New York (1992).

[34] Liang, K.-Y., Self, S. G. and Chang, Y.-C., *Modelling marginal hazards in multivariate failure time data*, Journal of the Royal Statistical Society. Series B (Methodological) 55, 441-453, (1993).

[35] Lopez, O. and Saint-Pierre, P., *Bivariate censored regression relying on a new estimator of the joint distribution function*, Journal of Statistical Planning and Inference 142, 2440-2453, (2012).

[36] Major, P. and Rejto, L. *Strong embedding of the estimator of the distribution functionunder random censorship*, The Annals of Statistics 16, 1113-1132, (1988).

[37] Massart, P. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. The annals of Probability, 1269-1283 (1990).

[38] Messaci, F., *Local averaging estimates of the regression function with twice censored data*, Statistics and Probability Letters 80, 1508-1511, (2010).

[39] Morales, D., Pardo, L. and Quesada, V., *Bayesian survival estimation for incomplete data when the life distribution is proportionally related to the censoring time distribution*, Communications in Statistics, Theory and Methods 20, 831-850, (1991).

[40] Nielsen, G. G., Gill, R. D. , Andersen, P. K. and S∅rensen, T. I. A., *A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models*, Scandinavian Journal Of Statistics 19, 25-43, (1992).

[41] Oakes, D. Multivariate survival distributions. Journaltitle of Nonparametric Statistics, 3(3-4), 343-354 (1994).

[42] Omelka, M., Gijbels, I., & Veraverbeke, N. Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing (2009).

[43] Patilea, V. and Rolin, J.-M., *Product limit estimators of the survival function with twice censored data*, The Annals of Statistics 34, 925-938, (2006).

[44] Pollard, D. A. A review of ecological studies on seagrass—fish communities, with particular reference to recent studies in Australia. Aquatic Botany, 18(1-2), 3-42 (1984).

[45] Romeo, J. S., Tanaka, N. I. and Pedroso-de-Lima, A. C., *Bivariate survival modeling: a Bayesian approach based on copulas*, Lifetime Data Analysis 12, 205-222, (2006).

[46] Rouabah, N.H., Nemouchi, N. and Messaci, F., *A rate of consistency for nonparametric estimators of the distribution function based on censored dependent data*, Statistical Methods and Applications 28, 259-280, (2019).

[47] Shih, J. H. and Louis, T. A., *Inferences on the association parameter in copula models for bivariate survival data*, Biometrics 51, 1384-1399, (1995).

[48] Shorack, G. R. and Wellner, J. A., *Empirical processes applications to statistics*, John Wily and sons, New York, (1986).

[49] Sklar, A., *Fonctions de repartition Ã n-dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris A. 8, 229-231, (1959).

[50] Samia, T., Mohamed, B., Nesrine, I., & Fatah, B. Nonparametric Estimation of the Copula Function with Bivariate Twice Censored Data(2022).

[51] Tsukahara, H. . Semiparametric estimation in copula models. Canadian Journal of Statistics, 33(3), 357-375. (2005).

[52] Wellner, J. A., *Limit Theorems for the Ratio of the Empirical Distribution Function to the True Distribution Function*, Z. Wahrscheinlichkeitstheorie verw. Gebiete 45, 73-88, (1978).