# Nonparametric copula estimation under bivariate censoring

Svetlana Gribkova, Olivier Lopez

# Nonparametric copula estimation under bivariate censoring

Svetlana GRIBKOVA[1], Olivier LOPEZ[2]

April 24, 2013

**Abstract**

In this paper, we consider nonparametric copula inference under bivariate censoring. Based on an estimator of the joint cumulative distribution function, we define a discrete and two smooth estimators of the copula. The construction that we propose is valid for a large number of estimators of the distribution function, and therefore for a large number of bivariate censoring frameworks. Under some conditions on the tails of the distributions, the weak convergence of the corresponding copula processes is obtained in $l^\infty([0,1]^2)$. We derive the uniform convergence rates of the copula density estimators deduced from our smooth copula estimators. Investigation on the practical behavior of our estimators is done through a simulation study and two real data applications, corresponding to different censoring settings. We use our nonparametric estimators to define a goodness-of-fit procedure for parametric copula models. A new bootstrap scheme is proposed to compute the critical values.

**Key words:** Copula function, nonparametric estimation, bivariate censoring, copula density, survival analysis, Kaplan-Meier estimator, goodness-of-fit, bootstrap.

**Short title:** Copula estimation under censoring.

[1] Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France, E-mail: svetlana.gribkova@etu.upmc.fr

[2] Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France, E-mail: olivier.lopez0@upmc.fr

# 1 Introduction

When it comes to analyzing dependence between random variables, copula models have appeared as a common and flexible tool. According to Sklar's Theorem (Sklar (1959)), the multivariate distribution function $F(t_1, \ldots, t_d) = \mathbb{P}(T_1 \leq t_1, \ldots, T_d \leq t_d)$ of a random vector $(T_1, \ldots, T_d)$ can be coupled to its univariate marginal distributions $F_j(t_j) = \mathbb{P}(T_j \leq t_j)$, for $j = 1, ..., d$, by the relation

$$F(t_1, \ldots, t_d) = \mathfrak{C}(F_1(t_1), \ldots, F_d(t_d)), \tag{1.1}$$

where $\mathfrak{C}$ is a copula function, that is, by definition, a multivariate distribution on $[0, 1]^d$ with uniform marginal distributions. The copula function appears as a quantity that contains all the information about the dependence structure of the random vector, and is unique in the case where the marginal distributions are continuous. In numerous situations in lifetime data analysis, estimation of $\mathfrak{C}$ must be performed from indirect observations of the variables $(T_1, ..., T_d)$, due to the presence of censoring (see e.g. Fleming and Harrington (1991) for a description of censoring mechanism). The aim of the present paper is to define a new class of nonparametric estimators of the copula function $\mathfrak{C}$ that is adapted to various schemes of multivariate random censoring. We derive asymptotic properties of these estimators, and investigate their practical behavior through a simulation study and some applications to real data.

Copula models represent a flexible alternative to fully parametric models of multivariate distribution function, allowing to study the dependence structure separately from the marginal distributions. This property of copula becomes of prime importance for applications in economics and insurance. A detailed introduction to copula theory and dependence modeling can be found in Joe (1997) or Nelsen (2006). For a recent survey on copula models in econometrics we refer to Patton (2012). Various applications in finance and/or insurance are considered by Frees and Valdez (1998), Embrechts et al. (2003), Cherubini et al. (2004) and Bouyé et al. (2007).

Copula estimation for uncensored data has been extensively studied in the literature. Several statistical procedures are available (see e.g. Genest and Rivest (1993), Tsukahara (2005) for parametric or semiparametric modeling, and Choros et al. (2010) for a global review of the existing methods). A nonparametric estimation approach was introduced by Deheuvels (1979), who defined the empirical copula function. We refer to Fermanian et al. (2004) and Segers (2012) for some recent studies of asymptotic properties of this estimator. Unlike parametric or semiparametric models, the empirical copula function

is not affected by misspecification (see Fermanian and Scaillet (2005) for more details). Therefore, nonparametric approaches are required to construct goodness-of-fit procedures, such as those studied in Fermanian (2005). The empirical copula is an irregular discrete estimator and does not permit a copula density estimation. Its smooth versions were considered by several authors, see e.g. Fermanian et al. (2004), and Omelka et al. (2009) who proposed techniques designed to reduce boundary bias.

Such nonparametric estimators all rely on the empirical distribution function of $(T_1, ..., T_d)$. Under random censoring, this empirical estimator is unavailable, since the variables $T_j$ subject to censoring are not directly observed. Parametric or semiparametric approaches can be adapted, by performing maximum likelihood estimation with a form of the likelihood criterion which takes the incompleteness of the observations into account, see Shih and Louis (1995). On the other hand, the extension of nonparametric procedures to the censored framework requires to replace the unavailable empirical distribution function by a suitable nonparametric estimator of $F$. Many estimators of $F$ have been proposed in the literature, see e.g. Dabrowska (1988), Akritas and Van Keilegom (2003), van der Laan (1996), see also a review of most existing procedures in Lopez and Saint-Pierre (2012). Depending on the identifiability assumptions on the censoring mechanism, different procedures can be introduced to take specific forms of censoring into account. Wang and Wells (1997) and Lopez and Saint-Pierre (2012) considered estimators which are consistent under some restrictions on the dependence structure of the censored variables, Gribkova et al. (2013) proposed an estimator which is adapted to a simplified censoring framework, corresponding to some specificity of insurance datasets. Goodness-of-fit procedures for censored copula models have been studied by Wang and Wells (2000), Luciano et al. (2008), Gribkova et al. (2013) (who used an extension of the procedure of Genest and Rivest (1993)), but are only valid in the particular case of Archimedean copula models.

In the present paper, we propose a general procedure of nonparametric copula estimation under multivariate censoring, based on the availability of a nonparametric estimator of the distribution function. This estimator is required to be of some generic form, which is compatible with many multivariate censoring schemes (that is under various types of identifiability assumptions). We construct three classes of copula estimators. The first one is non-smooth and can be considered as an extension of Deheuvels (1979). The two others are smooth estimators, based on kernel estimation of the distribution of either the variables $(T_j)_{1\leq j\leq d}$, or a transformed version of them (such as the one proposed by Omelka

et al. (2009) to reduce boundary effects in copula smoothing). The weak convergence of the corresponding copula processes is obtained. Moreover, we derive uniform asymptotic convergence rates of the copula density estimators obtained through differentiation of our smooth copula estimators. The practical behavior of these new estimators is investigated through a simulation study, and two applications to real datasets. As a by-product of our estimators, we propose a goodness-of-fit procedure (with computation of the critical values through a new bootstrap scheme) which is consistent for a large number of copula models, even in a non-Archimedean framework.

The rest of the paper is organized as follows. In section 2, we introduce the general multiple censoring model that we consider. We describe some examples of specific censoring schemes that will be covered by our general framework. Empirical copula estimators are defined and studied in section 3. Section 4 is devoted to the construction and theoretical study of smooth copula estimators. Simulation studies and real data applications are considered in section 5. Technical arguments are presented in the Appendix section.

# 2 Model description and examples

In this section, we describe the general framework of the present paper. Section 2.1 presents the general setup along with some notations and first model assumptions. Several classical examples are recalled in section 2.2.

## 2.1 General setup

Throughout this paper, we will focus on the two-dimensional case for the sake of simplicity. The extension of our results to higher dimensions is straightforward.

Consider a random vector $(T_1, T_2)$ with the cumulative distribution function $F(t_1, t_2) = \mathbb{P}(T_1 \leq t_1, T_2 \leq t_2)$ and marginal distribution functions $F_1, F_2$. We will denote by $\mathfrak{C}$ the associated copula function, that is

$$F(t_1, t_2) = \mathfrak{C}(F_1(t_1), F_2(t_2)).$$

To ensure the unicity of the copula function $\mathfrak{C}$, we will assume that the variables $T_1$ and $T_2$ are continuous.

In some cases, in particular in lifetime data analysis, one of these two variables (or even both of them) may be subject to censoring, and thus may not be directly observed. Instead of observing the variable $T_j$ $(j = 1, 2)$, one observes a minimum between it and

another (censoring) random variable, which will be denoted by $C_j$. The available data is then composed of i.i.d. replications $(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i})_{1\leq i\leq n}$ of random vector $(Y_1, Y_2, \delta_1, \delta_2)$, where $Y_j = \min(T_j, C_j)$ and $\delta_j = \mathbb{I}_{T_j\leq C_j}$ for $j = 1, 2$. Sometimes, in addition to this information, one may observe some auxiliary variables. Such situation is illustrated by our Example 3 below. Throughout this paper, we will assume that the support of the distribution of $T_j$ is included in the support of the corresponding censoring variable $C_j$ for $j = 1, 2$. This assumption is classical in lifetime data analysis. If it is not verified, a part of the distribution is not observed and thus the distribution function cannot be estimated consistently on the whole support of $(T_1, T_2)$.

As we have already mentioned in the introduction, different bivariate distribution function estimators can be considered, depending on the censoring scheme. However, most of them can be written in some generic form. We will assume throughout the paper that $F$ can be estimated consistently by an estimator $\mathbb{F}_n$ of the following form,

$$\mathbb{F}_n(t_1, t_2) = \frac{1}{n}\sum_{i=1}^{n} W_{in}\mathbb{I}_{Y_{1i}\leq t_1, Y_{2i}\leq t_2}, \tag{2.1}$$

where $W_{in}$ are random weights, designed to compensate asymptotically the bias caused by the particular structure of the data. The appropriate weight $W_{in}$ to be used depends strongly on the identifiability conditions that are required to infer on $F$. Basically, these assumptions describe a dependence structure between $(C_1, C_2)$ and $(T_1, T_2)$, and may differ from one application to another. In all the examples that we consider, we will assume that $W_{in} = \delta_{1i}\delta_{2i}\hat{g}(Y_{1i}, Y_{2i})$, where $\hat{g}$ is a function estimated from the data, converging towards a limit function $g$, where $g$ satisfies the following condition,

$$\forall \phi \in L^1,\ E\left[\delta_1\delta_2 g(Y_1, Y_2)\phi(Y_1, Y_2)\right] = E[\phi(T_1, T_2)]. \tag{2.2}$$

Classical situations where the indicated assumptions hold, are described in the following subsection.

## 2.2 Examples

In the examples that we propose, we only consider estimators of the distribution function that correspond to a positive measure. The results that we derive in the following can be adapted to distributions with eventual negative masses at some observations, such as the estimator considered by Dabrowska (1988) (see also Pruitt (1991)). However, the resulting copula estimators are not true copula functions due to this negative mass, therefore we do not focus on such cases.

**Example 1: censoring acts only on one of the two variables**

In this situation, $C_2 = \infty$ a.s., and consequently $Y_2 = T_2$ and $\delta_2 = 1$ a.s. In such a setting, the estimator $\mathbb{F}_n^{(1)}$ defined by Stute (1993) is of the form (2.1), that is

$$\mathbb{F}_n^{(1)}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} W_{in}^{(1)} \, \mathbb{I}_{Y_{1i} \leq t_1, Y_{2i} \leq t_2}, \tag{2.3}$$

where the random weights $W_{in}^{(1)}$ are the jumps of the Kaplan-Meier estimator of the distribution function of $T_1$. This estimator is consistent provided that $C_1$ is independent from $T_1$, and $\mathbb{P}(T_1 \leq C_1 | T_2, T_1) = \mathbb{P}(T_1 \leq C_1 | T_2)$. A practical way of rewriting $W_{in}^{(1)}$ consists of linking this jump to the Kaplan-Meier estimator of the censoring variable $C_1$. Indeed, defining a Kaplan-Meier estimator of the censoring variable,

$$\hat{G}(t) = 1 - \prod_{i:Y_{1i} \leq t} \left( 1 - \frac{1}{\sum_{j=1}^{n} \mathbb{I}_{Y_{1j} \geq Y_{1i}}} \right)^{1-\delta_{1i}},$$

according to Satten and Datta (2001),

$$W_{in}^{(1)} = \frac{\delta_{1i}}{1 - \hat{G}(Y_{1i}-)}.$$

Introducing $G(t) = \mathbb{P}(C_1 \leq t)$, this weight can be seen as an approximation of

$$W_{i}^{(1)} = \frac{\delta_{1i}}{1 - G(Y_{1i}-)}.$$

**Example 2: censoring variables linked through a copula function**

This situation is described in Lopez and Saint-Pierre (2012). In this framework, $(C_1, C_2)$ is supposed to be independent from $(T_1, T_2)$. Another assumption is made on the joint distribution of the random vector $(C_1, C_2)$. It is assumed that the joint survival function $S_G$ can be expressed as

$$S_G(c_1, c_2) = \mathbb{P}(C_1 \geq c_1, C_2 \geq c_2) = \mathcal{C}_G(S_1(c_1), S_2(c_2)),$$

where $\mathcal{C}_G$ is a known survival copula, and $S_1$ and $S_2$ are the marginal survival functions of $C_1$ and $C_2$. Note that the last assumption can be relaxed by estimating $\mathcal{C}_G$ from a parametric model. The only impact of this additional modeling is a modification of the asymptotic distribution of the distribution function estimator, without any modification of the convergence rate provided that the parametric model is sufficiently regular. Denoting by $\hat{S}_1$ and $\hat{S}_2$ the Kaplan-Meier estimators of $S_1$ and $S_2$, the estimator of Lopez and Saint-Pierre (2012) corresponds to

$$W_{in}^{(2)} = \frac{\delta_{1i}\delta_{2i}}{\mathcal{C}_G(\hat{S}_1(Y_{1i}), \hat{S}_2(Y_{2i}))},$$

with its limit equal to $W_i^{(2)} = \delta_{1i}\delta_{2i}[\mathcal{C}_G(S_1(Y_{1i}), S_2(Y_{2i}))]^{-1}$.

**Example 3: censoring variables differ only through an additional observed variable**

Here we consider a model, studied in Gribkova et al. (2013). It corresponds to a classical situation, appearing when it comes to study the insurance contracts, related to two individuals (generally two members of a same couple). If, for example, $T_1$ (resp. $T_2$) is the total lifetime of the husband (resp. his wife), the censoring variables $C_1$ and $C_2$ are their ages at a moment of the exit from the study, for a reason different from death. The observed variables are then $Y_1 = T_1 \wedge C_1$ and $Y_2 = T_2 \wedge C_2$. Besides these variables, the age difference $\varepsilon$ between two members of the couple is generally observed. In most cases, the two members of the couple, if both alive, exit the study at the same time. This leads to a link between two censoring variables through the relationship $C_2 - C_1 = \varepsilon$. The observations are then formed of i.i.d. replications $(Y_{1i}, Y_{2i}, \varepsilon_i, \delta_{1i}, \delta_{2i})_{1\leq i\leq n}$. In such setting, a consistent estimator of the distribution function $F$ is of the form (2.1), with the weights given by

$$W_{in}^{(3)} = \frac{\delta_{1i}\delta_{2i}}{1 - \tilde{G}(\max(Y_{1i}, Y_{2i} - \varepsilon)-)}, \tag{2.4}$$

where $\tilde{G}(y)$ is a Kaplan-Meier estimator of $G(y) = \mathbb{P}(C_1 \leq y)$ (see Gribkova et al. (2013)). The censoring variable is observed since one of the two lifetimes is censored, so the estimator $\tilde{G}(y)$ is based on the observations $(\inf(A_i, C_i), \mathbb{1}_{C_i\leq A_i})_{1\leq i\leq n}$, where $A_i = \inf(T_{1i}, T_{2i} - \varepsilon_i)$.

**Additional examples:**

An estimator of the form (2.1) has also been considered by Lopez (2012). It can be applied to a large set of situations, since it requires only that $(T_1, T_2)$ is independent from $(C_1, C_2)$, without making any assumption on the dependence structure of two censoring variables. Moreover, this estimator can be used in presence of bivariate left-truncation. Since, for now, there does not exist any weak convergence result for this estimator, we are unable to prove a weak convergence of the corresponding empirical copula process. However, the uniform consistency property of this estimator permits to establish a uniform convergence of the copula estimator, see Theorem 3.1 below.

# 3 Discrete nonparametric copula estimator

In this section, we first define a nonparametric copula estimator, which extends the empirical copula of Deheuvels (1979) in our framework. The asymptotic properties of this

estimator are considered in section 3.2 (uniform $n^{1/2}-$consistency) and in section 3.3 (weak convergence of the corresponding empirical process).

## 3.1 Definition of the estimators

By the definition of the copula function $\mathfrak{C}$, we have

$$\mathfrak{C}(u,v) = F(F_1^{-1}(u), F_2^{-1}(v)), \quad 0 \leq u, v \leq 1, \tag{3.1}$$

where $L^{-1}$ denotes the generalized inverse of a monotone function $L$. Therefore, the copula function can be estimated nonparametrically by considering an empirical version of (3.1), that is

$$\mathfrak{C}_n(u,v) = \mathbb{F}_n(\mathbb{F}_{1n}^{-1}(u), \mathbb{F}_{2n}^{-1}(v)), \quad 0 \leq u, v \leq 1, \tag{3.2}$$

where $\mathbb{F}_n$ is defined in (2.1) and $\mathbb{F}_{1n}(t_1) = \mathbb{F}_n(t_1, \infty)$, and $\mathbb{F}_{2n}(t_2) = \mathbb{F}_n(\infty, t_2)$. In the uncensored case, this definition reduces to that of the empirical copula estimator introduced in Deheuvels (1979). If $\mathbb{F}_n$ is a true distribution function (that is, a monotonic function with $\mathbb{F}_n(+\infty, +\infty) = 1$), then $\mathfrak{C}_n$ is a true copula function. For the examples considered in the previous sections, in some situations, the total mass of $\mathbb{F}_n$ may be strictly less than one (this is a classical issue for Kaplan-Meier estimator in the univariate case), leading to estimators $\mathfrak{C}_n$ with total mass less than one. In this case, the residual mass may be affected to the point $(1,1)$ in order to retrieve a true copula function. In the case where $\mathbb{F}_n$ allocates negative mass to some observations (for example the estimator of Dabrowska (1988)), this definition can not be used since $\mathbb{F}_{jn}$ (for $j = 1, 2$) may not be monotonic and $\mathbb{F}_{jn}^{-1}$ may not be defined. Nevertheless, (3.2) is asymptotically equivalent, up to the terms of order $O_P(1/n)$ uniformly on $u$ and $v$, to

$$\tilde{\mathfrak{C}}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} W_{in} \mathbb{1}_{\mathbb{F}_{1n}(Y_{1i}) \leq u, \mathbb{F}_{2n}(Y_{2i}) \leq v}, \tag{3.3}$$

which is still valid for non-monotonic functions $\mathbb{F}_{jn}$ (although, in this case, $\mathfrak{C}_n$ is not a copula).

In the following sections 3.2 and 3.3, we study the asymptotic behavior of the estimator $\mathfrak{C}_n$. Compared to the uncensored case, the main difficulty here is to handle the weights $W_{in}$. Indeed, in absence of censoring, one observes a sample composed of i.i.d. observations, each of them contributing to the estimator with the same weight equal to $n^{-1}$, while in our setting each weight is random and depends on the whole sample.

## 3.2 Uniform $n^{1/2}-$consistency

As in the uncensored case, asymptotic properties of the empirical copula estimator are derived from the corresponding properties of the underlying distribution function estimator $\mathbb{F}_n$. Thus, the $n^{1/2}-$consistency of $\mathfrak{C}_n$ only requires the $n^{1/2}-$convergence of $\mathbb{F}_n$, as stated in the following Theorem.

**Theorem 3.1** *Assume, without loss of generality, that $(T_1, T_2)$ are almost surely positive, and assume that $\mathbb{F}_{jn}$ is monotonic for $j = 1, 2$. Let $\mathcal{T}_1 = [-\infty, A_1]$, and $\mathcal{T}_2 = [-\infty, A_2]$, such that*

$$\sup_{t_1 \in \mathcal{T}_1, t_2 \in \mathcal{T}_2} |\mathbb{F}_n(t_1, t_2) - F(t_1, t_2)| = O_P(n^{-1/2}),$$

*and, for $j = 1, 2$,*

$$\sup_{t \in \mathcal{T}_j} |\mathbb{F}_{jn}(t) - F_j(t)| = O_P(n^{-1/2}).$$

*Moreover, assume that $\sup_{i=1,\ldots,n: Y_{1i} \leq A_1, Y_{2i} \leq A_2} W_{in} = O_P(1)$. Denoting $F_1(\mathcal{T}_1) = [0, u_1]$, and $F_2(\mathcal{T}_2) = [0, u_2]$, then, for any $\eta > 0$,*

$$\sup_{u \leq u_1 - \eta, v \leq u_2 - \eta} |\mathfrak{C}_n(u, v) - C(u, v)| = O_P(n^{-1/2}).$$

**Proof.** Let $\varepsilon < \eta$. Note that,

$$\sup_{u \leq u_1 - \eta, v \leq u_2 - \eta} |\mathfrak{C}_n(u, v) - \tilde{\mathfrak{C}}_n(u, v)| \leq \frac{2}{n} \sup_{i=1,\ldots,n: Y_{1i} \leq A_1, Y_{2i} \leq A_2} W_{in},$$

where $\tilde{\mathfrak{C}}_n$ is defined in (3.3).Therefore, it suffices to prove the uniform consistency of $\tilde{\mathfrak{C}}_n$. Next, observe that, for $j = 1, 2$, on the set $\mathcal{A}_\varepsilon = \{\sup_{t \in \mathcal{T}_j, j=1,2} |\mathbb{F}_{jn}(t) - F_j(t)| \leq \varepsilon\}$,

$$|\mathbb{1}_{\mathbb{F}_{jn}(Y_{ji}) \leq u} - \mathbb{1}_{F_j(Y_{ji}) \leq u}| \leq \mathbb{1}_{|F_j(Y_{ji}) - u| \leq \varepsilon},$$

for $Y_{1i} \leq A_1$ and $Y_{2i} \leq A_2$. Defining, for $u \leq u_1 - \eta$ and $v \leq u_2 - \eta$,

$$\mathbb{F}_n^*(u, v) = \frac{1}{n} \sum_{i=1}^{n} W_{in} \mathbb{1}_{F_1(Y_{1i}) \leq u} \mathbb{1}_{F_2(Y_{2i}) \leq v},$$

we can deduce that, on $\mathcal{A}_\varepsilon$,

$$|\tilde{\mathfrak{C}}_n(u, v) - \mathbb{F}_n^*(u, v)| \leq \frac{2}{n} \sum_{i=1}^{n} W_{in} \left\{ \mathbb{1}_{|F_1(Y_{1i}) - u| \leq \varepsilon} + \mathbb{1}_{|F_2(Y_{2i}) - v| \leq \varepsilon} \right\} \mathbf{1}_{Y_{1i} \leq A_1, Y_{2i} \leq A_2}.$$

The presence of $\mathbf{1}_{Y_{1i} \leq A_1, Y_{2i} \leq A_2}$ in the last equation is due to the fact that only the terms with $Y_{ji} \leq A_j$ (for $j = 1, 2$) give a positive contribution to the sum that defines $\mathbb{F}_n^*$ and

$\mathfrak{C}_n$. This is clear for $\mathbb{F}_n^*$. For $\tilde{\mathfrak{C}}_n$, observe that, due to monotonicity, $Y_{ji} > A_j$ implies $\mathbb{F}_{jn}(Y_{ji}) \geq \mathbb{F}_{jn}(A_j) \geq F_j(A_j) - \varepsilon$. Moreover,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} W_{in}\left\{\mathbb{1}_{|F_1(Y_{1i})-u|\leq\varepsilon} + \mathbb{1}_{|F_2(Y_{2i})-v|\leq\varepsilon}\right\} &= \mathbb{F}_{1n}(F_1^{-1}(u+\varepsilon)) - \mathbb{F}_{1n}(F_1^{-1}(u-\varepsilon)) \\ &\quad + \mathbb{F}_{2n}(F_2^{-1}(v+\varepsilon)) - \mathbb{F}_{2n}(F_1^{-1}(v-\varepsilon)).\end{aligned}$$

Using the uniform convergence of $\mathbb{F}_{jn}$ for $j = 1, 2$, we can deduce that, on $\mathcal{A}_\varepsilon$,

$$|\tilde{\mathfrak{C}}_n(u,v) - \mathbb{F}_n^*(u,v)| \leq 8\varepsilon,$$

Next, we take $\varepsilon = 8^{-1}Mn^{-1/2}$. We get, for $n$ large enough, $\mathbb{P}(n^{1/2}\sup_{u\leq u_1-\eta, u\leq u_2-\eta}|\tilde{\mathfrak{C}}_n(u,v) - \mathbb{F}_n^*(u,v)| > M) \leq \mathbb{P}(\mathcal{A}^c_{8^{-1}Mn^{-1/2}})$, where $\mathcal{A}^c$ denotes the complementary of the set $\mathcal{A}$. We then can deduce that $\lim_{M\to\infty}\bar{\lim}_{n\to\infty}\mathbb{P}(n^{1/2}\sup_{u\leq u_1-\eta,u\leq u_2-\eta}|\tilde{\mathfrak{C}}_n(u,v) - \mathbb{F}_n^*(u,v)| > M) = 0$. Moreover, $\mathbb{F}_n^*$ converges uniformly towards $\mathfrak{C}$ at rate $n^{1/2}$ for $u \leq u_1 - \eta$ and $v \leq u_2 - \eta$ from the rate of uniform convergence of $\mathbb{F}_n$. ■

## 3.3 Weak convergence of the censored empirical copula process

To obtain asymptotic weak convergence of $\mathfrak{C}_n$, some additional properties on the estimator $\mathbb{F}_n$ are required. First of them is the weak convergence of the empirical process $n^{1/2}(\mathbb{F}_n(t_1,t_2) - F(t_1,t_2))$, which is stated in Assumption 1. Let $l^\infty(\mathfrak{T})$ denote a space of all bounded functions $f : \mathfrak{T} \to \mathbb{R}$.

**Assumption 1** *Assume that $\mathbb{F}_n(t_1,t_2)$ is an $n^{1/2}-$consistent estimator of $F(t_1,t_2)$ satisfying*

$$\mathbb{H}_n(t_1,t_2) := \sqrt{n}(\mathbb{F}_n(t_1,t_2) - F(t_1,t_2)) \rightsquigarrow \mathbb{G}_F(t_1,t_2) \quad in \quad l^\infty(\mathbb{R}^2), \tag{3.4}$$

*where $\mathbb{G}_F(t_1,t_2)$ is a tight gaussian process and $\rightsquigarrow$ denotes the weak convergence.*

The next requirement is related to the fact that the copula estimator (3.2) must be invariant under the probability integral transform, see Fermanian et al. (2004) in the uncensored case. To express this invariance in our framework, define the pseudo-variables $(T_1^*, T_2^*, C_1^*, C_2^*)$ as

$$(T_1^*, T_2^*, C_1^*, C_2^*) = (F_1(T_1), F_2(T_2), F_1(C_1), F_2(C_2)),$$

and, for $j = 1, 2$,

$$Y_j^* = \min(T_j^*, C_j^*),\ \delta_j^* = \mathbb{I}_{T_j^*\leq C_j^*}.$$

We will denote by $(Y_{1i}^*, Y_{2i}^*, \delta_{1i}^*, \delta_{2i}^*)_{1\leq i\leq n}$ a corresponding i.i.d. sample. As $F_1$ and $F_2$ are monotonic, $\delta_j^* = \delta_j$. Joint distribution function of $(T_1^*, T_2^*)$ is to be estimated by $\mathbb{F}_n^*$, which is the estimator calculated by the same way as (2.1), but using the pseudo-sample $(Y_{1i}^*, Y_{2i}^*, \delta_{1i}^*, \delta_{2i}^*)_{1\leq i\leq n}$.

**Assumption 2** *The weights attributed by the estimator* (2.1) *are invariant under the probability integral transform, i.e.*

$$W_{in}^* = W_{in}, \quad i = 1, \ldots, n.$$

In all the examples that we consider, for each $i$, the weight $W_{in}$ depends only on the indicators $(\delta_{1i}, \delta_{2i})$, and on the ranks $R_{ji}$, for $j = 1, 2$, of the observations $Y_{ji}$ in samples $(Y_{j1}, ..., Y_{jn})$. Denote $R_{ji}^*$ the rank of $Y_{ji}^*$ in the transformed sample $(Y_{j1}^*, ..., Y_{jn}^*)$, then the monotonicity of $F_1$ and $F_2$ ensures that $R_{ji}^* = R_{ji}$. Therefore, Assumption 2 is naturally satisfied in the examples that we consider.

From Assumptions 1 and 2, the process $n^{1/2}(\mathbb{F}_n^*(u, v) - \mathfrak{C}(u, v))$ converges in $l^\infty([0, 1]^2)$ to the gaussian process $\mathbb{Z}_C^*(u, v) = \mathbb{G}_F(F_1^{-1}(u), F_2^{-1}(v))$. Defining the empirical process corresponding to the introduced copula estimator by

$$\mathbb{Z}_n(u, v) = \sqrt{n}(\mathfrak{C}_n(u, v) - \mathfrak{C}(u, v)), \qquad 0 \leq u, v \leq 1, \tag{3.5}$$

we now state the main result of this section.

**Theorem 3.2** *Suppose that $F$ has continuous marginal distribution functions and partial derivatives of its copula function exist and are continuous. Then the censored empirical copula process $\{\mathbb{Z}_n(u, v),\ 0 \leq u, v \leq 1\}$ converges weakly in $l^\infty([0, 1]^2)$ to the tight Gaussian process,*

$$\mathbb{Z}_C(u, v) = \mathbb{Z}_C^*(u, v) - \partial_1 C(u, v)\mathbb{Z}_C^*(u, 1) - \partial_2 C(u, v)\mathbb{Z}_C^*(1, v).$$

Theorem 3.2 is an extension of Theorem 3 in Fermanian et al. (2004), which establishes, in absence of censoring, the weak convergence of the empirical copula process in $l^\infty([0, 1]^2)$. The arguments that we develop are similar to those used in the uncensored case and are based mainly on functional Delta-Method.

**Proof.** 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. This is done through the following Lemma (proved in the Appendix section), which is a consequence of Assumption 2.

**Lemma 3.3** *Assume that $F_1$ and $F_2$ are continuous distribution functions and let $\mathfrak{C}_n(u,v)$ and $\mathfrak{C}_n^*(u,v)$ be the estimators of $\mathfrak{C}$ based respectively on the observations $(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i})_{1\leq i\leq n}$ and $(Y_{1i}^*, Y_{2i}^*, \delta_{1i}^*, \delta_{2i}^*)_{1\leq i\leq n}$. The equation $\mathfrak{C}_n(u,v) = \mathfrak{C}_n^*(u,v)$ holds.*

Lemma 3.3 shows that $\mathbb{Z}_n(u,v) = \mathbb{Z}_n^*(u,v) = \sqrt{n}(\mathfrak{C}_n^*(u,v) - \mathfrak{C}(u,v))$ where $\mathbb{C}_n^*(u,v) = \mathbb{F}_n^*(\mathbb{F}_{1n}^{*-1}(u), \mathbb{F}_{2n}^{*-1}(v))$, where $\mathbb{F}_{1n}^*, \mathbb{F}_{2n}^*$ are the marginals of $\mathbb{F}_n^*$. Then Lemma 2 of Fermanian et al. (2004) can be applied with $H^*(u,v) = \mathfrak{C}(u,v)$. The limiting process is then obtained by applying the Delta-Method and Theorem 3.9.28 in van der Vaart and Wellner (1996). ■

**Examples 1 to 3 (continued).** As we have already mentioned, Assumption 2 is quite natural and satisfied for all examples that we consider. Let us discuss now Assumption 1. In each example that we give, the difference between the distribution function and its estimator of the form (2.1) can be represented as

$$\mathbb{F}_n^{(j)}(t_1,t_2) - F^{(j)}(t_1,t_2) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(t_1,t_2) + R_n^j(t_1,t_2),$$

with $\sup_{t_1,t_2} |R_n^j(t_1,t_2)| = o_P(n^{-1/2})$, where $n^{-1/2}\sum_{i=1}^n \eta_i(t_1,t_2)$ is a sum of i.i.d. terms, which converges to a Gaussian process from the empirical process theory, and $j = 1, ..., 3$. For Example 1, such a representation has been derived by Stute (1996) for a fixed point $(t_1, t_2)$. The uniform convergence of the remainder term can be seen as a particular case of the results obtained by Lopez and Saint-Pierre (2012), which also cover Example 2. Gribkova et al. (2013) provide the representation for Example 3. In each case, the uniform convergence of the remainder term is obtained by adding some assumptions on the tail of the distributions of two lifetimes, compared to the tail of the distributions of the censoring variables.

# 4 Smoothed copula estimators

The procedure described in section 2 introduces a discrete nonparametric estimator of the copula function. However, if the underlying copula is continuous or even smooth, it may be reasonable to estimate it by a smooth function. In section 4.1, we introduce two nonparametric smooth copula estimators, which are valid in presence of censored observations. By taking their derivatives, we deduce two copula density estimators. Section 4.2 establishes the weak convergence of the empirical processes associated with the smooth copula estimators. Section 4.3 deals with the uniform convergence of the resulting copula density estimators.

## 4.1 Smooth estimators of the copula and its density

Let $k : \mathbb{R}^2 \mapsto \mathbb{R}$ denote a bivariate kernel function (that is a smooth function with integral over $\mathbb{R}^2$ equal to one), and $K$ its cumulative integral, that is

$$K(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} k(u, v) du dv.$$

Introducing a smoothing parameter $h > 0$, a kernel estimator of the distribution function can be deduced through a convolution of the measure defined by the estimator (2.1) with the measure with density $k_h(u, v) = h^{-2}k(u/h, v/h)$.

A natural extension of this classical Parzen-Rosenblatt estimator in the case of censored data leads to an estimator of $F$ of the form

$$\hat{\mathbb{F}}_n^1(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} W_{in} K \left( \frac{t_1 - Y_{1i}}{h}, \frac{t_2 - Y_{2i}}{h} \right). \tag{4.1}$$

Although we use, for notation convenience, the same $h$ for both components, different bandwidths may be used. Like all smoothing techniques, this estimator is sensitive to the choice of the bandwidth parameter(s). This question will be discussed in section 5. Let us introduce now a first smooth copula estimator given by

$$\hat{\mathfrak{C}}_n^1(u, v) = \hat{\mathbb{F}}_n^1((\hat{\mathbb{F}}_{1n}^1)^{-1}(u), (\hat{\mathbb{F}}_{2n}^1)^{-1}(v)). \tag{4.2}$$

In the uncensored case, this estimator reduces to the kernel estimator studied by Fermanian et al. (2004), who established a functional central limit theorem for the associated empirical process.

In order to ensure the weak convergence of the estimator $\hat{\mathfrak{C}}_n^1$, one must control its bias through an assumption on the boundedness of the second order partial derivatives of the underlying joint distribution function $F$. Let us notice that these regularity conditions do not impose the uniform boundedness of the second order derivatives of the corresponding copula function itself, which would have excluded from consideration several important families of copulas. An important drawback of estimator (4.2) 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) introduced some transformation of the initial variables. It leads to construction of a kernel estimator of the distribution function F of $(T_1, T_2)$ based on pseudo-observations, whose marginal distributions are asymptotically equal to some distribution function $\Phi$, designed to avoid corner bias problems.

This method can be extended to our framework, leading to a second smooth estimator $\mathfrak{C}_n^2$ of the copula function. Indeed, for some distribution function $\Phi$, let us consider a couple of variables $(\tilde{T}_1, \tilde{T}_2) = (\Phi^{-1}[F_1(T_1)], \Phi^{-1}[F_2(T_2)])$ and pseudo-observations $(\Phi^{-1}[\mathbb{F}_{1n}(Y_{1i})], \Phi^{-1}[\mathbb{F}_{2n}(Y_{2i})])_{1\leq i\leq n}$. Since the copula function is invariant under monotone transformations, the variables $(\tilde{T}_1, \tilde{T}_2)$ are coupled by the same copula as $(T_1, T_2)$. Next, we define an estimator of the joint distribution function of $(\tilde{T}_1, \tilde{T}_2)$ by

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

where $\mathbb{F}_{1n}$ (resp. $\mathbb{F}_{2n}$) denotes the Kaplan-Meier estimator of the distribution of $T_1$ (resp. $T_2$). Then, the corresponding copula estimator is defined as,

$$\hat{\mathfrak{C}}_n^2(u, v) = \hat{\mathbb{F}}_n^2(\Phi^{-1}(u), \Phi^{-1}(v)). \tag{4.3}$$

Since these copula estimators are smooth, two estimators of the copula density $c(u, v)$ can be deduced by considering

$$\hat{c}^i(t_1, t_2) = \frac{\partial^2}{\partial_{t_1}\partial_{t_2}} \hat{\mathfrak{C}}_n^i(t_1, t_2), \tag{4.4}$$

for $i = 1, 2$.

## 4.2 Functional CLT for the smooth copula estimators

The proof of the weak convergence of the copula processes associated with the defined smooth estimators relies on asymptotic properties of the weights $W_{in}$ and of the estimators of type (2.1). We first state some key assumptions that will allow, in our proofs, the replacement of the weights $W_{in}$ by their limit quantities $W_i$ (up to some additional terms).

**Assumption 3** *Let us recall that $W_{in} = \delta_{1i}\delta_{2i}\hat{g}(Y_{1i}, Y_{2i})$, where $\hat{g}(y_1, y_2)$ is a random function converging to its deterministic counterpart $g(y_1, y_2)$ satisfying (2.2). For $j = 1, 2$, denote $\tau_j = \inf\{t : F_j(t) = 1\}$. Assume that, on every set $\mathcal{Y} = [-\infty, t_1] \times [-\infty, t_2]$ with $t_1 < \tau_1$ and $t_2 < \tau_2$, $g$ is bounded and is twice continuously differentiable with respect to its arguments with uniformly bounded partial derivatives up to order two. Moreover, assume that:*

1. $\sup_{t\leq t_1, u\leq t_2} |\hat{g}(t_1, t_2) - g(t_1, t_2)| = O_P(n^{-1/2})$, *and that the restrictions of $\hat{g}$ and $g$ to the set $\{(t, u) : t \leq t_1, u \leq t_2\}$ both belong to Donsker classes of functions;*

2. *for all* $\psi \in \mathcal{F}$, *where* $\mathcal{F}$ *denotes a Donsker class of bounded functions such that* $\psi(y_1, y_2) = 0$ *for* $y_1 > t_1$ *or* $y_2 > t_2$, *we have the following representation,*

$$\sum_{i=1}^{n} [W_{in} - W_i]\psi(Y_{1i}, Y_{2i}) = \frac{1}{n}\sum_{i=1}^{n} \eta^{\psi}(Y_{1,i}, Y_{2,i}, \delta_{1i}, \delta_{2i}) + R_n(\psi), \tag{4.5}$$

*with* $\sup_{\psi\in\mathcal{F}} |R_n(\psi)| = o_P(n^{-1/2})$, *and* $E[\eta^{\psi}(Y_{1,i}, Y_{2,i}, \delta_{1i}, \delta_{2i})] = 0$ *for all* $\psi$;

3. $\{(t_1, t_2, d_1, d_2) \to \eta^{\psi_{h,y_1,y_2}}(t_1, t_2, d_1, d_2) : h \in [0, 1/4], (y_1, y_2) \in \mathcal{Y}\}$ *is a Donsker class of functions, with* $\sup_{(y_1,y_2)\in\mathcal{Y}} E[(\eta^{\psi_{h,y_1,y_2}}(Y_{1,i}, Y_{2,i}, \delta_{1i}, \delta_{2i}))^2] \to_{h\to 0} 0$, *defining*

$$\begin{aligned} \phi_{h,y_1,y_2}(t_1, t_2) &= K_h(y_1 - t_1) K_h(y_2 - t_2), \\ \psi_{h,y_1,y_2}(t_1, t_2) &= \phi_{h,y_1,y_2}(t_1, t_2) - \mathbb{1}_{t_1\le y_1, t_2\le y_2}. \end{aligned}$$

I.i.d. representations of the type (4.5) are classical tools when it comes to studying the asymptotic properties of the estimators (2.1). Moreover, the previous assumptions are valid for all the practical examples that we consider. Indeed, the desired representations were obtained in Theorem 3.3 of Lopez and Saint-Pierre (2012) for Examples 1 and 2 and in Theorem 3.1 of Gribkova et al. (2013) for Example 3. The uniform convergence of $\hat{g}$ is, in all the examples, a consequence of the uniform consistency of the Kaplan-Meier type estimators on compact sets. In each case, the restrictions of $\hat{g}$ and $g$ to compact sets can easily seen to be in Donsker class of functions from Theorem 2.7.5 in van der Vaart and Wellner (1996). Point 3 is more technical, but is reasonable, having in mind the particular shape of function $\eta^{\psi}$ coming from the Examples. We refer to section 6.5 to see how this Assumption can be checked in the Examples.

Assumption 3 would be sufficient if we restrain ourselves to proving the convergence of $\hat{\mathfrak{C}}_n^i$ on $[0, a] \times [0, b]$ for $a$ and $b$ strictly less than 1. Indeed, in this case, the convergence is not affected by the observations with $Y_{1i}$ close to $\tau_1$ or $Y_{2i}$ close to $\tau_2$, which give no contribution to the value of $\hat{\mathfrak{C}}_n^i(u, v)$ for $u < a$ and $v < b$. These large observations (close to the tail of at least one of the marginal distributions) give rise to particular difficulties, which are similar to those encountered in the univariate case. For Kaplan-Meier estimator, i.i.d. representations of the same type as in Assumption 3 can be obtained under standard conditions if one avoids investigating the right tail of the distribution (see e.g. Gijbels and Veraverbeke (1991)). To obtain representations valid on the whole real line, some assumptions on the distribution of the lifetime and of the censoring are required, as in Stute (1995). This is the purpose of our Assumption 4 below.

**Assumption 4** *Assume that* $E[\delta_1\delta_2 g(T_1,T_2)^2] < \infty$, *and assume that there exist i.i.d. random variables* $(Z_i)$ *such that*

$$\sup_i |W_{in} - W_i| \leq A_n Z_i,$$

*where* $A_n = O_P(n^{-1/2})$ *and* $E[Z_i] < \infty$.

**Examples 1 to 3 (continued).** Such decomposition of the weights can be deduced from Theorem 2.1 in Gill (1983), which implies that, for $\hat{G}_j$ denoting a Kaplan-Meier estimator of the cdf $G_j$ of a censoring variable $C_j$, and denoting by $Y_{[j,n]}$ the largest observation for the variable $j$,

$$\sup_{t\leq Y_{[j,n]}} \left| \frac{h(t)\{\hat{G}_j(t) - G_j(t)\}}{1-\hat{G}_j(t)} \right| = O_P(n^{-1/2}),$$

provided that $\int h(t)^2 dL_{F_j,G_j}(t) < \infty$, where $L_{F_j,G_j}(t) = \int_0^t [(1-F_j(u))(1-G_j(u))^2]^{-1} dG_j(u)$, with $F_j$ denoting the cdf of the corresponding lifetime $T_j$ (here, the roles of $T_j$ and $C_j$ are reversed compared to Theorem 2.1 in Gill (1983)). Therefore, in Example 1, Assumption 4 is valid for

$$A_n^1 = \sup_{t\leq Y_{[j,n]}} \left\{ \frac{|\hat{G}(t) - G(t)|}{L_{F_1,G}^{1/2+\varepsilon}(t)[1-\hat{G}(t)]} \right\},$$

where $\varepsilon > 0$, and $Z_i^1 = \delta_{1i}[1-G(Y_{1i})]^{-1}$. Similarly, for Example 2, assuming that $\mathcal{C}_G(u,v) \geq u^{\alpha_1}v^{\alpha_2}$ and that its first order partial derivative are bounded, take

$$A_n^2 = \sup_{t\leq Y_{[j,n]}} \left\{ \sum_{j=1}^{2} \frac{|\hat{G}_j(t) - G_j(t)|}{L_{F_j,G_j}^{1/2+\varepsilon}(t)[1-\hat{G}_j(t)]} \right\},$$

and

$$Z_i^2 = \frac{\delta_{1i}\delta_{2i}}{\mathcal{C}_G(1-G_1(t_1),1-G_2(t_2))} \left\{ \frac{(1-G_1(t_1))^{1-\alpha_1} L_{F_1,G_1}^{1/2+\varepsilon}(t)}{(1-G_2(t_2))^{\alpha_2}} \right.$$
$$\left. + \frac{(1-G_2(t_2))^{1-\alpha_2} L_{F_2,G_2}^{1/2+\varepsilon}(t)}{(1-G_1(t_1))^{\alpha_1}} \right\}.$$

For Example 3, one can take $A_n^3$ similar to $A_n^1$, but replacing $F_1$ by $F^* = \mathbb{P}(A \leq t)$ in the definition of $L_{F^*,G}$. In each case, the condition $E[Z_i] < \infty$ is a condition on the tails of the distributions: indeed, considering Example 1 which is the simplest, $L_{F_1,G}(t) \leq [1-F_1(t)]^{-1}[1-G(t)]^{-1}$, which shows that $E[Z_i]$ is finite provided that the density of the censored lifetime $T_{1i}$ decreases fast enough, compared to $[1-F_1(t)][1-G(t)]$.

The main tool to obtain the weak convergence of the empirical process associated with $\mathfrak{C}_n^1$ is Theorem 4.1, establishing the asymptotic equivalence of the smoothed estimator $\hat{\mathbb{F}}_n^1(t_1,t_2)$ to the estimator $\mathbb{F}_n(t_1,t_2)$, up to some negligible remainder term. The proof of Theorem 4.1 is given in the Appendix section.

**Theorem 4.1** *Consider a symmetric kernel function $k$ with compact support such that $k \geq 0$ and $\int u^2 k(u)du < \infty$. Let $F(t_1,t_2)$ be twice differentiable distribution function with the second order derivatives uniformly bounded on $\mathbb{R}^2$ and let $h^2\sqrt{n} \to 0$. Under Assumptions 3 and 4, we have,*

$$\sqrt{n} \sup_{t_1,t_2 \in \mathbb{R}} |\hat{\mathbb{F}}_n^1(t_1,t_2) - \mathbb{F}_n(t_1,t_2)| \underset{\mathbb{P}}{\longrightarrow} 0. \tag{4.6}$$

**Corollary 4.2** *Theorem 4.1 implies the weak convergence of the process $\sqrt{n}(\hat{\mathbb{F}}_n^1(t_1,t_2) - F(t_1,t_2))$ to the tight gaussian limit process $\mathbb{G}_F(t_1,t_2)$ from the Assumption 1.*

**Corollary 4.3** *If the condition $h^2\sqrt{n} \to 0$ is not satisfied, it can be seen from the proof of Theorem 4.1 (see Appendix section), that*

$$\sup_{t_1,t_2 \in \mathbb{R}} |\hat{\mathbb{F}}_n^1(t_1,t_2) - \mathbb{F}_n(t_1,t_2)| = O_P(h^2).$$

Proving the convergence of $\mathfrak{C}_n^2$ follows a different path, and requires some assumptions on the function $\Phi$ involved in the transformation of the observations, and on the behavior of the partial derivatives of the copula function near the boundaries of $[0,1]^2$. These requirements (mainly same as those of Omelka et al. (2009)) are listed in Assumption 5. The restrictions on the copula function are quite reasonable, since most of classical copula families follow this property, see Appendix D in Omelka et al. (2009).

**Assumption 5** *Assume that $\mathfrak{C}$ is twice continuously differentiable on $]0,1[^2$, and that*

$$\begin{aligned}
\frac{\partial^2 \mathfrak{C}(u,v)}{\partial u^2} &= O\left(\frac{1}{u(1-u)}\right), \\
\frac{\partial^2 \mathfrak{C}(u,v)}{\partial v^2} &= O\left(\frac{1}{v(1-v)}\right), \\
\frac{\partial^2 \mathfrak{C}(u,v)}{\partial u \partial v} &= O\left(\frac{1}{\sqrt{uv(1-u)(1-v)}}\right).
\end{aligned}$$

*Moreover, assume that $\Phi'$ and $\Phi'^2/\Phi$ are bounded.*

We also add an invariance properties of the weights $W_{in}$ after transformation of the variables by $\Phi$. Like Assumption 2, this assumption automatically holds if the weights only depend on the ranks of the lifetimes (since, for $j = 1, 2$, the transformation $\Phi^{-1}(F_j(\cdot))$ is increasing).

**Assumption 6** *Let $W_{in}^{\Phi}$ denote the weights computed from the method (2.1) based on the sample*

$$(\Phi^{-1}(F_1(Y_{1i})), \Phi^{-1}(F_1(Y_{2i})), \delta_{1i}, \delta_{2i})_{1 \leq i \leq n}.$$

*Assume that $W_{in}^{\Phi} = W_{in}$.*

We now state the main result of this section.

**Theorem 4.4** *Under the conditions of Theorem 4.1 for the case $i = 1$, and under Assumption 5 and 6 for the case $i = 2$, the smoothed empirical copula processes*

$$\hat{\mathbb{Z}}_n^i(u, v) = \sqrt{n}(\hat{\mathfrak{C}}_n^i(u, v) - \mathfrak{C}(u, v)), \quad 0 \leq u, v \leq 1, \tag{4.7}$$

*converge weakly to a tight Gaussian process $\{\mathbb{Z}_{\mathfrak{C}}(u, v), 0 \leq u, v \leq 1\}$ in $l^{\infty}([0, 1]^2)$ for $i = 1, 2$.*

**Proof.** The result for the case $i = 1$ follows directly from our Theorem 4.1 and technics used by Fermanian et al. (2004) in the uncensored case. Indeed, the result obtained by Fermanian et al. (2004) is a consequence of their Lemma 7, the stochastic equicontinuity of the process (4.7) and its finite-dimensional convergence. Arguments, used in Lemma 7 can not be applied in the presence of censoring. Thus establishing that the supremum of the difference between the discrete estimator of the distribution function and its smoothed version is of the order of $o_P(n^{-1/2})$ constitutes the main difficulty, resolved by Theorem 4.1. The rest of the arguments are applicable directly to our case. The proof for the case $i = 2$ is postponed to the Appendix section. ■

The results of Theorem 4.4 and Theorem 3.2 can be applied to construct a nonparametric goodness-of-fit test of the hypothesis $H_0 : \mathfrak{C} \in \mathcal{C}_{\Theta}$ against $H_1 : \mathfrak{C} \notin \mathcal{C}_{\Theta}$, where $\mathcal{C}_{\Theta} = \{\mathfrak{C}_{\theta}, \theta \in \Theta\}$ is some parametric class of copulas. The test procedure, which will be explained in details in section 5.2.1, is analogous to the test, based on Deheuvels copula in the uncensored case and uses the limit distribution of the empirical process associated with the copula estimator. Other approach, which we will not develop here, is a goodness-of-fit test based on copula densities. In the incensored case, it was studied by Fermanian (2005). This method can also be adapted to our framework, using density estimators (4.4).

## 4.3 Uniform consistency of copula density estimators

In this section, we derive the uniform consistency of the copula density estimators (4.4) on a compact subset of $[0,1]^2$, which we denote by $\mathcal{C} = [\theta_1, \theta_2] \times [\theta_1', \theta_2']$, where $\theta_1 > 0$, $\theta_1' > 0$, and $\theta_2 < 1$, $\theta_2' < 1$. The maximum size of the compact $\mathcal{C}$ is restricted by the following assumption.

**Assumption 7** *Assume that, for all $\mathcal{Y} = [-\infty, t_1] \times [-\infty, t_2]$ where $t_1 < \tau_1$ and $t_2 < \tau_2$,*

$$\sup_{i:(Y_{1i},Y_{2i})\in\mathcal{Y}} |W_{in} - W_i| = o_P(\eta_n), \quad \eta_n = h^2 + \frac{[\log n]^{1/2}}{h\sqrt{n}}.$$

**Examples 1 to 3 (continued).** This assumption holds for all examples due to the uniform convergence of the Kaplan-Meier estimators (of the censoring variables) involved in the definition of $W_{in}$, on sets that do not contain the tail of the distribution. A regularity assumption on $\mathcal{C}_G$ (first derivatives uniformly bounded) must be added in Example 2. In this case, it is easy to check that $\sup_{i:(Y_{1i},Y_{2i})\in\mathcal{Y}} |W_{in} - W_i| = O_P(n^{-1/2})$.

The next assumption serve for controlling the denominator, appearing in the derivatives of kernel copula estimator $\hat{\mathfrak{C}}_n^1$.

**Assumption 8** *Assume that there exist a constant $c$ such that*

$$\inf_{x\in\mathcal{C}} f_i(F_i^{-1}(x)) > c, \quad i = 1, 2.$$

*Moreover, assume that the density $f$ of $(T_1, T_2)$ is twice continuously differentiable with partial derivatives up to order two uniformly bounded.*

We now state the main result of this section, which is proved in the Appendix section.

**Theorem 4.5** *Recall that*

$$\eta_n = h^2 + \frac{[\log n]^{1/2}}{h\sqrt{n}}.$$

- *Under Assumptions 7 and 8, for a kernel function $k$ satisfying the assumptions of Theorem 4.1, and for $h$ is such that $hn^\alpha \to 0$ for some $\alpha > 0$, and $nh^2[\log n]^{-1} \to \infty$, then*

$$\sup_{(u,v)\in\mathcal{C}} |\hat{c}_1(u,v) - c(u,v)| = O_P(\eta_n). \tag{4.8}$$

- *If $nh^{10/3} \to \infty$, assume that*

*1. the kernel function $k$ is four times continuously differentiable,*

*2. the function $(x, y) \to c(\Phi(x), \Phi(y))\Phi'(x)\Phi'(y)$ is $C^2$ on every compact set,*

*and assume that Assumption 5 holds. Then,*

$$\sup_{(u,v)\in\mathcal{C}} |\hat{c}_2(u, v) - c(u, v)| = O_P(\eta_n). \tag{4.9}$$

# 5 Simulation study and real data analysis

This section is divided into two main parts. In the first of them (section 5.1) we present the results of a simulation study of the estimators introduced in the paper. Their performance is evaluated on censored datasets, simulated using several parametric copula models. The second part (section 5.2) is devoted to real data applications. A goodness-of-fit test based on our estimators is defined in section 5.2.1. In section 5.2.2, we study a dataset where only one variable is censored. In section 5.2.3, our nonparametric copula estimation techniques are applied to a joint survival dataset from a Canadian insurer.

## 5.1 Simulation study

To investigate the finite sample behavior of our estimators we carried out simulation studies in different settings. We consider Model 1, illustrating our Example 1 (where only one variable is censored) and Model 2, corresponding to Example 2 with two censored variables and an assumption on the joint distribution of the censoring. We do not present results from Example 3 (censoring variables are linked through an additional variable), since they are quite similar to those of Example 1 and 2.

### 5.1.1 Simulation scheme

In each setting, we simulate 1000 bivariate samples of size $n = 200$ according to the simulation schemes described below.

**Distribution of lifetimes.**

- The marginal distribution of $T_1$ and $T_2$ are simulated according to Weibull distributions with the shape parameters $k_1 = 2, k_2 = 2.2$ and the scale parameters $\lambda_1 = 3.1, \lambda_2 = 4.1$.

- To model the dependence structure, we consider three Archimedean copula families: Clayton, Frank and Gumbel (see Table 1 for the expressions of corresponding copulas). For each family, we consider two values of the dependence parameter,

| Model | $\mathfrak{C}_\theta(u,v)$ |
|---|---|
| Clayton | $\max(u^{-\theta}+v^{-\theta}-1,0)^{-1/\theta}$ |
| Frank | $-\theta^{-1}\log\left(1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{(\exp(-\theta)-1)}\right)$ |
| Nelsen 4.2.20 | $[\log\left(\exp(u^{-\theta})+\exp(v^{-\theta})-e\right)]^{-1/\theta}$ |
| Joe | $1-[(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta]^{1/\theta}$ |
| Gumbel | $\exp[-\{(-\log u)^\theta+(-\log v)^\theta\}^{1/\theta}]$ |

Table 1: Archimedean copulas

corresponding to Kendall's $\tau$ coefficients equal to $\tau_1 = 0.25$ and $\tau_2 = 0.75$. These values are summarized in Table 2.

| Copula | $\tau = 0.25$ | $\tau = 0.75$ |
|---|---|---|
| Clayton | 0.66 | 6.00 |
| Frank | 2.30 | 14.00 |
| Gumbel | 1.33 | 4.00 |

Table 2: Values of copula parameters.

**Distribution of censoring variables.**

- Censoring variables are modelized by Pareto distributions, that is, their survival function is equal to

$$\mathbb{P}(C>t)=\begin{cases}\frac{1}{(t+1)^\lambda} & \text{if } t\geq 0\\ 1 & \text{if } t<0.\end{cases}$$

The values of the Pareto distribution parameters are chosen in order to achieve approximatively 25% of censored observations (for each censored marginal) in a first case, and 50% of censored observations in a second case.

- In case of Model 1, only one lifetime is censored. In Model 2, two lifetimes are censored by two independent censoring variables.

For each of the obtained censored datasets, we perform nonparametric copula estimation, using the three estimators introduced in the paper. To reduce the computational time, a fixed bandwidth is used to assess the performance of the smooth estimators ($h = 0.2$). A data-driven choice of bandwidth is discussed in the real data applications.

### 5.1.2 Results.

In order to evaluate the performance of the estimators, we compute two distances between the estimated copula function and the true underlying copula. We first consider a Kolmogorov-Smirnov distance (KS in the following), that is

$$d_{KS}(\hat{\mathfrak{C}}, \mathfrak{C}) = \sup_{u,v} |\hat{\mathfrak{C}}(u, v) - \mathfrak{C}(u, v)|,$$

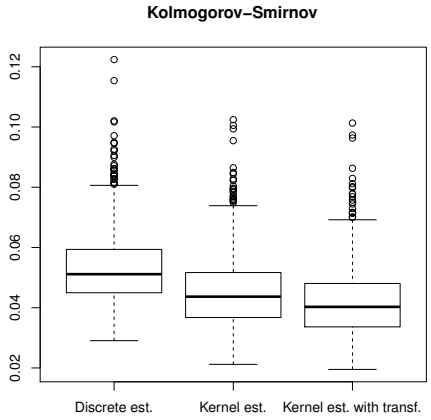
and a square-root of a quadratic integrated distance (RMSE in the following),

$$d_{RMSE}(\hat{\mathfrak{C}}, \mathfrak{C}) = \left[ \int (\hat{\mathfrak{C}}(u, v) - \mathfrak{C}(u, v))^2 du dv \right]^{1/2},$$

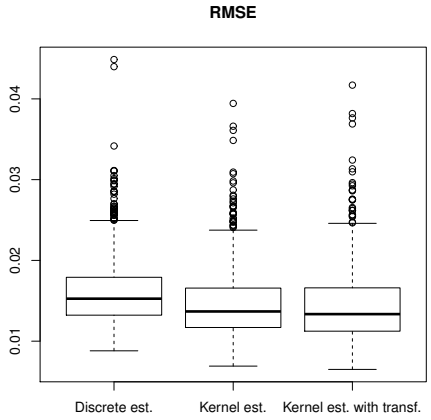
where $\hat{\mathfrak{C}}$ denotes one of the nonparametric copula estimators defined in the paper.

These distances calculated for 1000 replications are presented through boxplots on Figures 1 to 3. As the results are quite similar in several situations, we present here only the selected cases. Figure 1 for Frank copula permits to compare the errors of the estimation for different levels of censoring and values of Kendall's $\tau$ coefficient. As expected, estimation becomes less precise if we increase the percentage of censoring (compare (d) and (f) of Figure 1). The second observation is that, at fixed level of censoring, error decreases when the correlation between variables becomes stronger (compare (a),(b) of Figure 1 with (e),(f)). Figure 2 presents some results for Gumbel and Clayton copulas. Figure 3 illustrates Example 3, where two variables are censored. Here the situation is quite similar to the previous case.

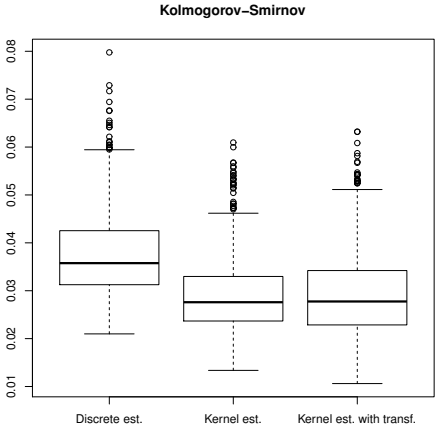
Most of the presented figures show that the performance of two kernel estimators is superior to that of the discrete estimator. These results are in accord with the results of Omelka et al. (2009) in the uncensored case, and are natural due to the fact that the data was simulated using copulas which are smooth on $]0, 1[^2$. While the performances of two kernel estimators (4.2) and (4.3) are quite close, the interest of considering the estimator (4.3) is that the transformation of initial variables makes it less sensitive to the marginal distributions of variables, than the estimator (4.2).



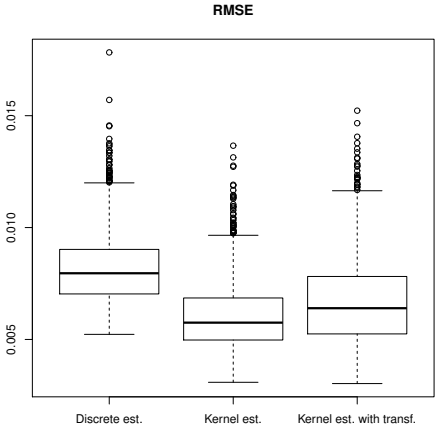

(a) Frank copula, 50% of censoring, $\tau = 0.25$



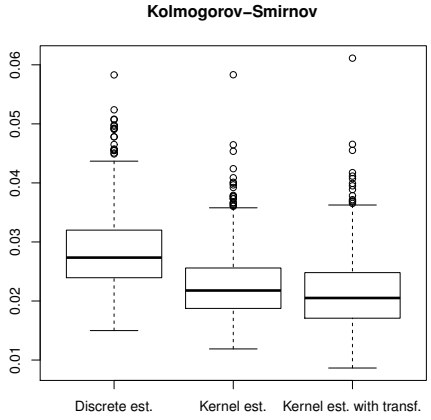

(b) Frank copula, 50% of censoring, $\tau = 0.25$



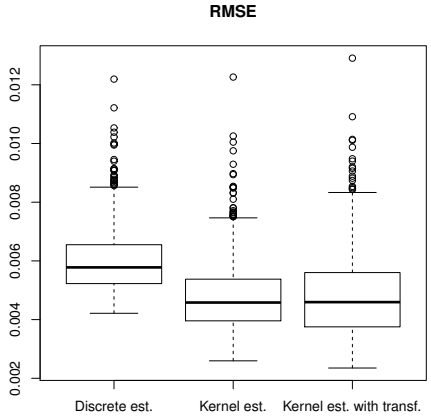

(c) Frank copula, 50% of censoring, $\tau = 0.75$




(d) Frank copula, 50% of censoring, $\tau = 0.75$

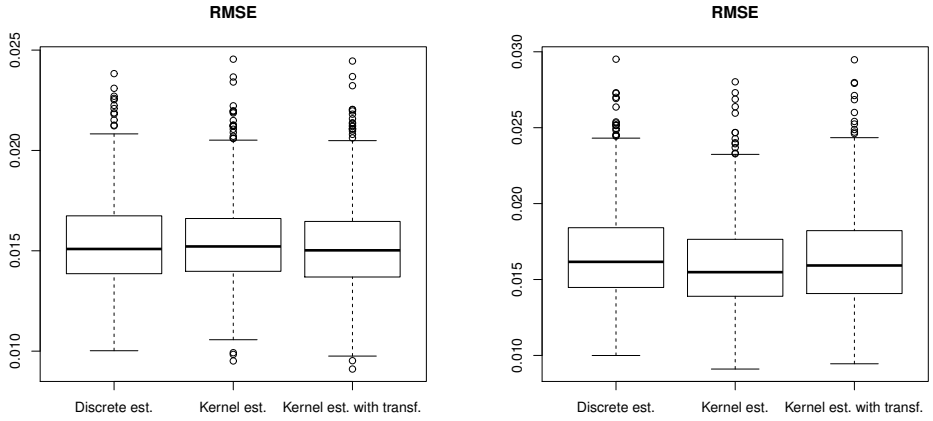



(e) Frank copula, 25% of censoring, $\tau = 0.75$

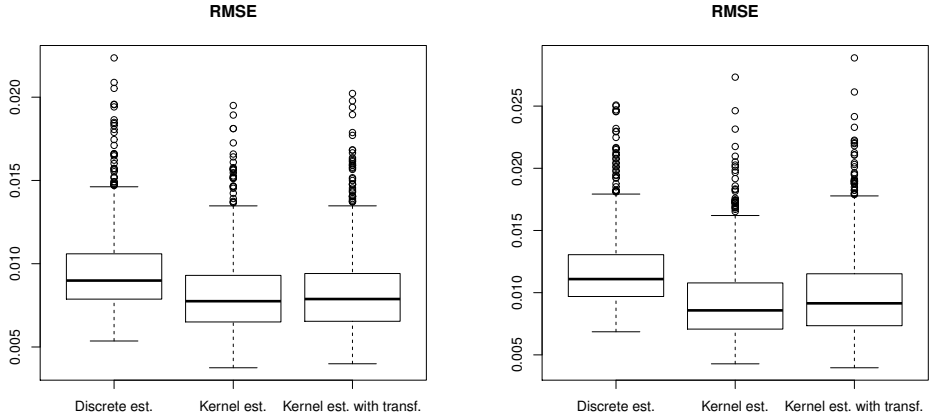



(f) Frank copula, 25% of censoring, $\tau = 0.75$

Figure 1: Model 1 (only one variable is censored): Frank copula

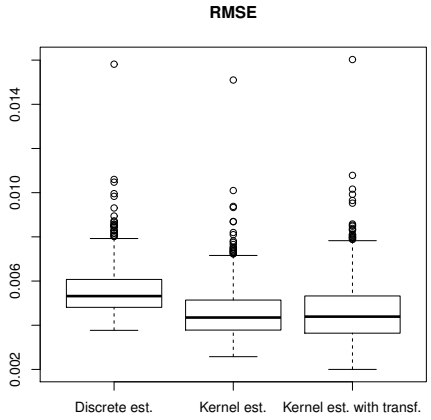



(a) Clayton copula, 25% of censoring, $\tau = 0.75$ (b) Clayton copula, 50% of censoring, $\tau = 0.75$

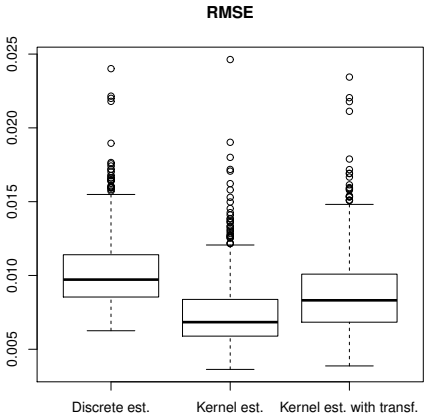



(c) Gumbel copula, 25% of censoring, $\tau = 0.75$ (d) Gumbel copula, 50% of censoring, $\tau = 0.75$

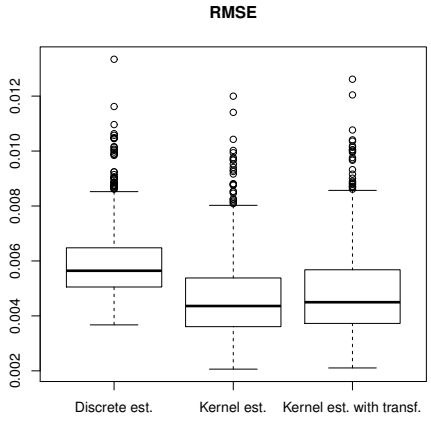
Figure 2: Model 1 (only one variables is censored): Clayton and Gumbel copulas



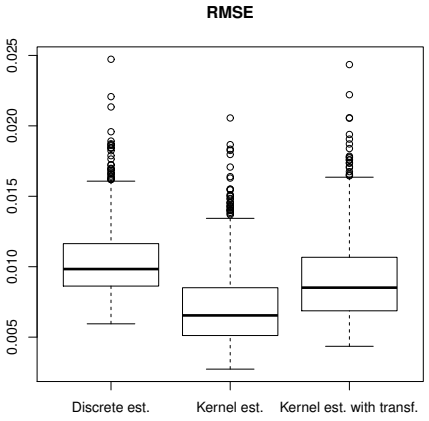

(a) Frank copula, 25% of censoring, $\tau = 0.75$



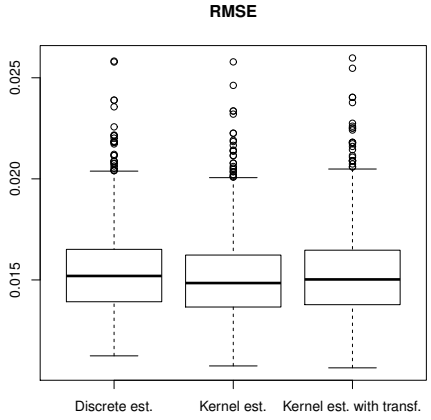

(b) Frank copula, 50% of censoring, $\tau = 0.75$




(c) Gumbel copula, 25% of censoring, $\tau = 0.75$




(d) Gumbel copula, 50% of censoring, $\tau = 0.75$




(e) Clayton copula, 25% of censoring, $\tau = 0.75$


(f) Clayton copula, 50% of censoring, $\tau = 0.75$

Figure 3: Model 2: two variables are censored

## 5.2 Real data applications

In this section we consider some applications of our results to two real data sets. We first present in section 5.2.1 a general method of goodness-of-fit testing of a parametric model, based on the nonparametric estimators defined in the previous sections, and a bootstrap procedure permitting to compute the $p-$values. Section 5.2.2 is devoted to a bivariate non life insurance data set, where one variable represents the indemnity to be paid for a claim and the other variable is the associated allocated loss adjustment expense. The second example is a life insurance data representing joint lifetimes of couples subscribed an insurance contract, which is studied in section 5.2.3.

### 5.2.1 Goodness-of-fit procedure based on the nonparametric copula estimators

Let $\mathfrak{C}_n$ be one of three nonparametric copula estimators defined previously, and $\mathfrak{C}_{\hat{\theta}}$ its parametric estimator under $H_0$. Here $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator of $\theta$ (obtained, for example, by the maximum likelihood method (see, Shih and Louis (1995)) or by using a relationship between Kendall's $\tau$ coefficient and $\theta$ (see Luciano et al. (2008))). Consider a Cramér-Von-Mises type of distance between the estimators $\mathfrak{C}_n$ and $\mathfrak{C}_{\hat{\theta}}$, defined by

$$d_n = n \int_0^1 \left(\mathfrak{C}_n(u,v) - \mathfrak{C}_{\hat{\theta}}(u,v)\right)^2 d\mathfrak{C}_n(u,v). \tag{5.1}$$

Other kind of distances between $\mathfrak{C}_n$ and $\mathfrak{C}_{\hat{\theta}}$ may be used instead. It follows from Theorem 3.2 that $d_n$ admits a weak limit under $H_0$, while, under $H_1$, $d_n$ tends to infinity with probability tending to one. Thus the critical region of the test is of the form $\mathcal{R} = \{d_n > d\}$. The limit law of $d_n$ can be computed using the exact form of $\mathbb{Z}_C$ and of a limit distribution for $\hat{\theta}$.

Alternatively, one can compute critical values and $p-$values via bootstrap. This is the path that we used in our examples, relying on the bootstrap procedure described below. To this aim, suppose that $\mathbb{G}_n$ is some estimator of the joint distribution function of the censoring variables, defining a true distribution function. In all the examples presented in this paper, such an estimator is available (up to some normalization of the Kaplan-Meier estimator of the censoring distribution, or to some affectation of the residual mass to infinity).

**Bootstrap procedure.**

For $b = 1, ..., B$, where $B$ denotes the number of bootstrap replications,

1. simulate $(T_{1i}^b, T_{2i}^b)_{1\leq i\leq n}$ according to the distribution defined by $\mathfrak{C}_{\hat{\theta}}$ and the marginal distributions $\mathbb{F}_{1n}$ and $\mathbb{F}_{2n}$;

2. simulate $(C_{1i}^b, C_{2i}^b)_{1\leq i\leq n}$ according to $\mathbb{G}_n$;

3. compute the $b-$th bootstrap sample $(Y_{1i}^b, Y_{2i}^b, \delta_{1i}^b, \delta_{2i}^b)_{1\leq i\leq n}$ based on the simulated variables;

4. using the $b-$th bootstrap sample, compute estimators $\hat{\theta}^b$ and $\mathfrak{C}_n^b$ and the corresponding distance $d_n^b$;

Then, use the vector $(d_n^b)_{1\leq b\leq n}$ to estimate the $p-$value.

### 5.2.2 Loss-ALAE data.

This bivariate dataset was provided by the US Insurance Services Office and studied previously by Denuit and Van Keilegom (2006) and Frees and Valdez (1998). It contains 1500 observations, composed of losses (indemnities to be paid by insurance company) and of allocated loss adjustment expenses (ALAE's), associated with each loss. ALAE's are additional costs which are be related to lawyer's fees or claim investigation expenses. Each contract has a specific policy limit $L$ (a maximal claim amount). If amount of the i-th claim exceeds the corresponding limit $L_i$, only $L_i$ is registered by the insurance company and the loss variable is censored. As the expensive claims are usually associated with greater settlement costs, large values of the loss variable are expected to be associated with large values of ALAE's. In the reinsurance practice it is important to model correctly this association. We refer to Denuit and Van Keilegom (2006) for more details. As it was mentioned in the last paper, although the data contains only 34 censored observations, they have a much higher mean than the complete data (217.941\$ versus 37.110\$), so that the estimation can be biased if censored observations are not taken into account.

In order to identify which parametric copula family is more adapted to modeling the dependance structure of the data, we performed a goodness-of-fit test (using the estimator (3.2)) for four families of archimedean copulas: Frank, Gumbel, Clayton, Joe (see Table 1 for the corresponding copula expressions). The results are given in Table 3. According to this test procedure Gumbel's copula outperforms the three other models. Denuit and Van Keilegom (2006) and Frees and Valdez (1998) also noticed that Gumbel's copula furnishes the best fit. Nevertheless our procedure clearly rejects Frank's and Clayton's models.

| Model | $\hat{\theta}$ | Test statistic | 95% quantile | 97.5 % quantile | 99 % quantile | p-value |
|---|---|---|---|---|---|---|
| Frank | 3.30 | $9,64 \times 10^{-5}$ | $1.10 \times 10^{-5}$ | $1.43 \times 10^{-5}$ | $1.88 \times 10^{-5}$ | $< 0.001$ |
| Gumbel | 1.50 | $2.13 \times 10^{-5}$ | $4.99 \times 10^{-5}$ | $5.61 \times 10^{-5}$ | $6.11 \times 10^{-5}$ | 0.851 |
| Clayton | 1.00 | $37.8 \times 10^{-5}$ | $1.97 \times 10^{-5}$ | $2.20 \times 10^{-5}$ | $2.58 \times 10^{-5}$ | $< 0.001$ |
| Joe | 1.90 | $6.71 \times 10^{-5}$ | $5.92 \times 10^{-5}$ | $6.49 \times 10^{-5}$ | $7.35 \times 10^{-5}$ | 0.019 |

Table 3: Loss-ALAE data:. goodness-of-fit for the considered copula models.

### 5.2.3 Canadian insurer's data.

This dataset belongs to a large Canadian insurer and contains joint lifetimes of members of the couples who subscribed an insurance contract[1]. 11,947 couples were observed between December, 29, 1988, and December 31, 1993. In our study, we eliminated the same-sex contracts and we kept only one contract for couples with more than one policy. The remaining sample concerns 11,454 observations. As most couples were still alive at the end of the observation period, the dataset contains a huge proportion of censored observations (98,2% with at least one censored lifetime). In the present approach, we neglected the left-truncation phenomenon as in Gribkova et al. (2013). For more details on this dataset we refer to Carriere (2000), Frees et al. (1996), Luciano et al. (2008) and Youn and Shemyakin (2001).

We recall that, if two variables $T_1$ and $T_2$ are coupled by $C(u, v)$, their joint survival function $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ can be written as $S(t_1, t_2) = \tilde{\mathcal{C}}(S_1(t_1), S_2(t_2))$, where $S_1, S_2$ are marginal survival functions and $\tilde{\mathcal{C}}$ is a copula function with

$$\tilde{\mathcal{C}}(u, v) = C(1 - u, 1 - v) + u + v - 1.$$

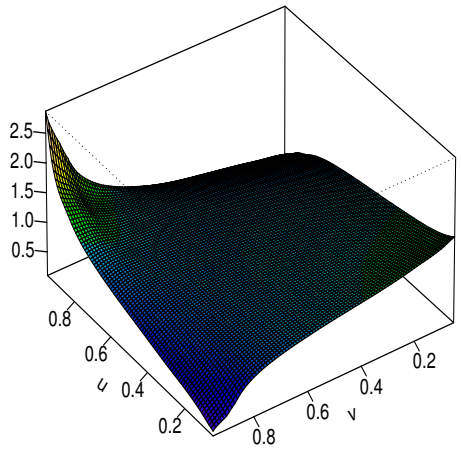
For easier comparison with former studies of this dataset, we present the results of the estimation of $\tilde{C}$ rather than $C$ itself.

**1. Copula density estimation.** Denote by $\tilde{c}$ the copula density associated with $\tilde{C}$. We use $\hat{c}^1(1-u, 1-v)$ and $\hat{c}^2(1-u, 1-v)$ according to the definition (4.4) to estimate $\tilde{c}$. To select the bandwidth, we propose the following heuristic criterion based on a reference copula $\mathfrak{C}_{ref}$. We select a bandwidth in a set $\mathcal{H}$ such that

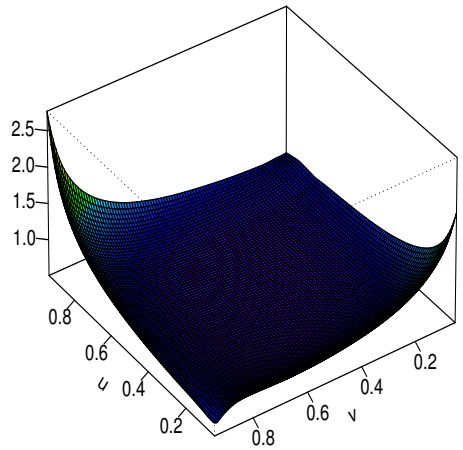
$$\hat{h}^j = \arg\min_{h \in \mathcal{H}} \int (\mathfrak{C}_n^j(u, v) - \mathfrak{C}_{ref}(u, v))^2 du dv, \quad j = 1, 2.$$

[1]The authors wish to thank the Society of Actuaries, through the courtesy of Edward J. Frees and Emiliano Valdez, for allowing use of the data in this paper

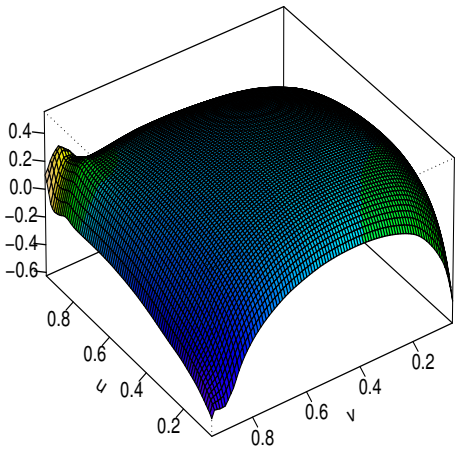
As a reference copula, we considered Frank's copula with parameter value specified in Table 4, due to the shape of the estimated density (which had similarities with Frank's copula). The two estimations of the survival copula density $\tilde{c}$ that we obtained are represented by Figure 4 (a) and (b). The difference between these two estimators is represented in Figure 4 (c). This difference is pronounced in the corners of the unit square. This is not surprising due to the fact that the estimator $\hat{\mathfrak{C}}_n^2$ is designed to improve estimation on the border of $[0,1]^2$.




(a) Estimator $\hat{c}^1(1-u,1-v),\ h=0.14$




(b) Estimator $\hat{c}^2(1-u,1-v),\ h=2$




(c) Difference between two density estimators

Figure 4: Nonparametric copula density estimation. Data from a Canadian insurer.

**2. Goodness-of-fit for semiparametric copula models.** We now apply the methodology proposed at the end of the section 4.2 to perform a goodness-of-fit test. Three Archimedean copula families are consdered: Clayton, Frank and Nelsen 4.2.20 (see Table 1 for the coresponding expressions). The test statistics is given by (5.1), where we

used the discrete estimator (3.2) for $\mathfrak{C}_n$. The estimated copula parameters and p-values of the test are presented in Table 4. They were calculated via the procedures described in Gribkova et al. (2013).

| Model | $\hat{\theta}$ | Test statistic | 95% quantile | 97.5 % quantile | 99 % quantile | p-value |
|---|---|---|---|---|---|---|
| Clayton | 4.89 | 0.00076 | 0.00208 | 0.00251 | 0.00310 | 0.391 |
| Frank | 11.41 | 0.00036 | 0.00092 | 0.00115 | 0.00143 | 0.416 |
| Nelsen 4.2.20 | 1.33 | 0.00116 | 0.00137 | 0.00160 | 0.00230 | 0.103 |

Table 4: Canadian data set: goodness-of-fit for the considered survival copula models.

# 6 Appendix

## 6.1 Proof of Lemma 3.3

Let $m$ be the total number of couples composed of doubly uncensored observations and let $(Y_{[j,1]}, \ldots, Y_{[j,m]})$ for $j = 1, 2$ be the order statistics, corresponding to these observations. Denote as $W_{[j,i]}$ the weight associated with $Y_{[j,i]}$. We recall that the weights attributed to the censored observations are equal to 0. Under Assumption 2, both $\mathfrak{C}_n$ and $\mathfrak{C}_n^*$ are defined on the same set of greed points, i.e.

$$(u_i, v_j) = \left( \frac{1}{n} \sum_{k=1}^{m} W_{[1,k]} \mathbb{1}_{k \leq i}, \frac{1}{n} \sum_{l=1}^{m} W_{[2,l]} \mathbb{1}_{l \leq j} \right), \quad 1 \leq i, j \leq m.$$

Then, similarily to Fermanian et al. (2004), with probability one,

$$\begin{aligned}
\mathfrak{C}_n(u_i, v_j) &= \mathbb{F}_n(Y_{[1,i]}, Y_{[2,j]}) \\
&= \mathbb{F}_n(F_1^{-1}F_1(Y_{[1,i]})), F_2^{-1}F_2(Y_{[2,j]})) \\
&= \mathbb{F}_n^*(F_1(Y_{[1,i]}), F_2(Y_{[2,j]})) \\
&= \mathbb{F}_n^*((\mathbb{F}_{1n}^*)^{-1}\mathbb{F}_{1n}^*(F_1(Y_{[1,i]})), (\mathbb{F}_{2n}^*)^{-1}\mathbb{F}_{2n}^*(F_2(Y_{[2,j]}))) \\
&= \mathfrak{C}_n^*(\mathbb{F}_{1n}^*(F_1(Y_{[1,i]})), \mathbb{F}_{2n}^*(F_2(Y_{[2,j]}))) \\
&= \mathfrak{C}_n^*(u_i, v_j).
\end{aligned}$$

## 6.2 Proof of Theorem 4.1

The proof is decomposed in two parts. First, we consider $t_1 < \tau_1$ and $t_2 < \tau_2$ and study the convergence for $(y_1, y_2) \in \mathcal{Y} = [-\infty, t_1] \times [-\infty, t_2]$. In this case, since the kernel function $k$ has compact support, one can consider that $W_{in}$ is bounded by some finite function (for $n$ large enough, the observations with values of $Y_{1i} \geq t_1 + \varepsilon$ or $Y_{2i} \geq t_2 + \varepsilon$ for any $\varepsilon > 0$ give a contribution zero to the sum). Next, we rely on a tightness argument to make $t_1 \to \tau_1$ and $t_2 \to \tau_2$.

**1. Convergence for** $(y_1, y_2) \in \mathcal{Y}$.

As stated before, since $k$ has compact support, for $n$ large enough, only terms with $(Y_{1i}, Y_{2i}) \in \mathcal{C}_\varepsilon$ contribute to $\hat{\mathbb{F}}_n^1(y_1, y_2)$ for $y_1 < t_1$ and $y_2 < t_2$. Therefore, we will assume, in this first part of the proof, that $(Y_{1i}, Y_{2i}) \in \mathcal{Y}_\varepsilon = [-\infty, t_1 + \varepsilon] \times [-\infty, t_2 + \varepsilon]$ for all $i$.

Defining $K_h(y) = \int_{-\infty}^{yh^{-1}} k(u)du$, observe that the class of functions

$$\mathcal{F}_1 = \{(y_1, y_2) \to \phi_{h,y_1,y_2}(t_1, t_2) = K_h(y_1 - t_1) K_h(y_2 - t_2), h \in [0, 1/4]\},$$

is Donsker from Lemma A.1 in Omelka et al. (2009). Therefore, it follows from (4.5) in Assumption 3 that

$$\begin{aligned}\hat{\mathbb{F}}_n^1(y_1, y_2) - \mathbb{F}_n(y_1, y_2) &= \frac{1}{n}\sum_{i=1}^n W_i \left[\phi_{h,y_1,y_2}(Y_{i1}, Y_{i2}) - \mathbb{1}_{Y_{i1} \leq y_1, Y_{i2} \leq y_2}\right] \\ &\quad + \frac{1}{n}\sum_{i=1}^n \eta^{\psi_{h,y_1,y_2}}(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i}) + R_n(y_1, y_2),\end{aligned}$$

where $\sup_{y_1,y_2} |R_n(y_1, y_2)| = o_P(n^{-1/2})$, and where we recall that

$$\psi_{h,y_1,y_2}(t_1, t_2) = \phi_{h,y_1,y_2}(t_1, t_2) - \mathbb{1}_{t_1 \leq y_1, t_2 \leq y_2}.$$

From the Donsker assumption on the functions $\eta^{\psi_{h,y_1,y_2}}$ (see Assumption 3), we get, from the asymptotic equicontinuity of Donsker classes and the fact that the $L^2-$norms of $\eta_i^{\psi_{h,y_1,y_2}}$ uniformly tend to zero,

$$\sup_{h,(y_1,y_2)\in\mathcal{Y}} \left| \frac{1}{n}\sum_{i=1}^n \eta^{\psi_{h,y_1,y_2}}(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i}) \right| = o_P(n^{-1/2}),$$

since, for all $h$ and all $(y_1, y_2)$, $E[\eta^{\psi_{h,y_1,y_2}}(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i})] = 0$.

Next, let $W$ be a random variable having the same distribution as the variables $(W_i)_{1 \leq i \leq n}$. Observe that, again, since $y_1 < t_1$ and $y_2 < t_2$, $W$ can be considered as almost surely bounded, and $W \times \mathcal{F}_1$ is a Donsker classes of functions, from a permanence

properties of Donsker classes, see van der Vaart and Wellner (1996) Example 2.10.10. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} W_i \left[\phi_{h,y_1,y_2}(Y_{i1},Y_{i2}) - \mathbb{1}_{Y_{i1}\leq y_1, Y_{i2}\leq y_2}\right] = \int\int [\phi_{h,y_1,y_2}(t_1,t_2) - \mathbb{1}_{t_1\leq y_1, t_2\leq y_2}] dF(t_1,t_2) + o_P(n^{-1/2}),$$

where the $o_P-$rate holds uniformly in $(y_1, y_2)$, and where we used that, for any function $\phi$ with finite first moment, $E[W\phi(Y_{1i}, Y_{2i})] = E[\phi(T_{1i}, T_{2i})]$. Using, again, a second order Taylor expansions and the differentiability assumptions on $F$, we get $\sup_{y_1<t_1,y_2<t_2} |\hat{\mathbb{F}}^1_n(y_1,y_2) - \mathbb{F}_n(y_1,y_2)| = o_P(n^{-1/2})$.

**2. Convergence in the right tail of the distribution.**

We apply Lemma 7.1 in Lopez and Saint-Pierre (2012) to the process

$$R_n(t_1,t_2) = n^{1/2}\left\{\hat{\mathbb{F}}^1_n(\tau_1,\tau_2) - \mathbb{F}_n(\tau_1,\tau_2) - \hat{\mathbb{F}}^1_n(t_1,t_2) + \mathbb{F}_n(t_1,t_2)\right\}.$$

This can be done by checking that $|R_n(t_1,t_2)| \leq M_n\Gamma_n(t_1,t_2)$, with

$$\begin{aligned} M_n &= n^{1/2}A_n, \\ \Gamma_n(t_1,t_2) &= \frac{1}{n}\sum_{i=1}^{n} Z_i \mathbb{I}_{Y_{1i}>t_1, Y_{2i}>t_2}. \end{aligned}$$

Assumption 4 ensures that $M_n$ and $\Gamma_n$ satisfy the conditions 2 to 5 of Lemma 7.1 in Lopez and Saint-Pierre (2012).

## 6.3 Proof of Theorem 4.4 (case $i = 2$)

Let $L_{1i} = \Phi^{-1}(F_1(T_{1i}))$ and $L_{2i} = \Phi^{-1}(F_2(T_{2i}))$. These random variables have marginal distribution $\Phi$, and have joint distribution function $F^{\Phi}(l_1,l_2) = \mathbb{P}(L_1 \leq l_1, L_2 \leq l_2) = C(\Phi(l_1),\Phi(l_2))$. Since the transformation $\Phi^{-1}(F_1(\cdot))$ is increasing, we can observe that the censoring model is equivalent (up to this transformation), to the model based on, for $j = 1, 2$,

$$\begin{aligned} M_{ji} &= \Phi^{-1}(F_j(Y_{ji})), \\ \delta^{\Phi}_{ji} &= \mathbb{1}_{L_{ji}\leq D_{ji}} = \delta_{ji}, \end{aligned}$$

where $D_{ji}$ are the transformed censoring variables, that is $D_{ji} = \Phi^{-1}(F_j(C_{ji}))$. We will denote by $\mathbb{F}^{\Phi}_n$ the estimator of the joint distribution of $(L_1, L_2)$ similar to $\mathbb{F}_n$ but based on

the transformed variables. From Assumption 6, the weights $W_{in}^{\Phi} = \delta_{1i}\delta_{2i}\hat{g}^{\Phi}(M_{1i}, M_{2i})$ based on the transformed model are the same as the weights $W_{in}$. Moreover, define $g^{\Phi}(l_1, l_2) = g(\Phi^{-1}(F_1(l_1)), \Phi^{-1}(F_1(l_2)))$.

Next, define $\hat{M}_{ji} = \Phi^{-1}(\mathbb{F}_{jn}(Y_{ji}))$. Let $f_{nj}(m) = \Phi^{-1}(\mathbb{F}_{jn}(F_j^{-1}(\Phi(m))))$. We have $\hat{M}_{ji} = f_{nj}(M_{ji})$. We can decompose

$$\begin{aligned}\hat{\mathfrak{C}}_n^2(\Phi(l_1), \Phi(l_1)) &= \frac{1}{n}\sum_{i=1} W_{in}\left[K_h(l_1 - \hat{M}_{1i})K_h(l_2 - \hat{M}_{2i}) - K_h(l_1 - M_{1i})K_h(l_2 - M_{2i})\right] \\ &\quad + \hat{\mathbb{F}}_n^{1,\Phi}(l_1, l_2), \qquad (6.1)\end{aligned}$$

where $\hat{\mathbb{F}}_n^{1,\Phi}$ denotes the estimator $\hat{\mathbb{F}}_n^1$ based on the transformed variables. From Theorem 4.1, we see that this second term in (6.1) satisfies

$$\hat{\mathbb{F}}_n^{1,\Phi}(l_1, l_2) = \mathbb{F}_n^{\Phi}(l_1, l_2) + R_n^{\Phi}(l_1, l_2),$$

with $\sup_{l_1,l_2} |R_n^{\Phi}(l_1, l_2)| = o_P(n^{-1/2})$. Indeed, the limit $F^{\Phi}$ of $\hat{\mathbb{F}}_n^{1,\Phi}$ is assumed to satisfy the assumptions of Theorem 4.1 as a consequence of Assumption 5. Moreover, $\mathbb{F}_n^{\Phi}(l_1, l_2) = \mathbb{F}_n^*(\Phi(l_1), \Phi(l_2))$.

Hence the proof consists of showing the negligibility of the first term in (6.1). The path of the proof is similar to the one of Theorem 4.1: we first consider $l_1 < t_1^{\Phi}$ and $l_2 < t_2^{\Phi}$ (with obvious extension of the notation $(t_1, t_2)$) which allows to consider bounded weights $W_i$. Next, we use a tightness argument to obtain convergence on the whole plane. This last part is exactly the same as in the proof of Theorem 4.1, we therefore only focus on the case $l_1 < t_1^{\Phi}$ and $l_2 < t_2^{\Phi}$.

Since $\Phi^{-1}(\mathbb{F}_{jn})$ are increasing, we can see that the functions in the bracket of the first term of (6.1) belong to a Donsker class of functions, from Lemma A.1 in Omelka et al. (2009). Therefore, this first term can be rewritten as

$$\begin{aligned}&\int\int \left[K_h\left(l_1 - f_{1n}(l)\right) - K_h\left(l_1 - l\right)\right] K_h\left(l_2 - l'\right) \frac{\hat{g}^{\Phi}(l, l')dF^{\Phi}(l, l')}{g^{\Phi}(l, l')} \\ &+ \int\int \left[K_h\left(l_2 - f_{2n}(l')\right) - K_h\left(l_2 - l'\right)\right] K_h\left(l_1 - l\right) \frac{\hat{g}^{\Phi}(l, l')dF^{\Phi}(l, l')}{g^{\Phi}(l, l')} \\ &+ \int\int \left[K_h\left(l_1 - f_{1n}(l)\right) - K_h\left(l_1 - l\right)\right]\left[K_h\left(l_2 - f_{2n}(l')\right) - K_h\left(l_2 - l'\right)\right] \frac{\hat{g}^{\Phi}(l, l')dF^{\Phi}(l, l')}{g^{\Phi}(l, l')} \\ &= \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,\end{aligned}$$

up to some remainder terms that are $o_P(n^{-1/2})$ uniformly in $(l_1, l_2)$, and where we used the fact that $E[\delta_{i1}\delta_{i2}g(Y_{1i}, Y_{2i})\phi(Y_{i1}, Y_{i2})] = E[\phi(T_{i1}, T_{i2})]$. The last term $\mathcal{T}_3$ is a second

order term. A Taylor expansion and the uniform convergence of $\mathbb{F}_{jn}$ for $j = 1, 2$ show that this term is $O_P(n^{-1})$ uniformly in $(l_1, l_2)$. The first two terms $\mathcal{T}_1$ and $\mathcal{T}_2$ can be studied in a similar way due to their symmetric definition, hence we only focus on $\mathcal{T}_1$.

To study $\mathcal{T}_1$, we first replace $\hat{g}$ by $g$, observing that, from a first order Taylor expansion,

$$\left| \int \int \left[ K_h \left( l_1 - f_{1n}(l) \right) - K_h \left( l_1 - l \right) \right] K_h \left( l_2 - l' \right) \frac{\{ \hat{g}^{\Phi}(l, l') - g^{\Phi}(l, l') \} dF^{\Phi}(l, l')}{g^{\Phi}(l, l')} \right|$$
$$\leq C_0 \sup_{i:(y_1,y_2) \in \mathcal{C}_\varepsilon} |\hat{g}(y_1, y_2) - g(y_1, y_2)| \sup_t |\mathbb{F}_{1n}(t) - F_1(t)| h^{-1},$$

for some absolute constant $C_0$. Indeed, since $dF^{\Phi}(l, l') = c(\Phi(l), \Phi(l'))\Phi'(l)\Phi'(l')dldl'$, we get $|f_{1n}(l) - l| dF^{\Phi}(l, l') \leq C_0' \sup_t |\mathbb{F}_{1n}(t) - F_1(t)|$. Since $nh^2 \to \infty$ and using Assumption 3, one obtains that

$$\mathcal{T}_1 = \int \left[ K_h \left( l_1 - f_{1n}(l) \right) - K_h \left( l_1 - l \right) \right] \int K_h \left( l_2 - l' \right) dF^{\Phi}(l, l') dw_1 + o_P(n^{-1/2}),$$

where the $o_P-$rate does not depend on $(l_1, l_2)$. Moreover,

$$\int K_h \left( l_2 - l' \right) dF^{\Phi}(l, l') = \int k(w_2) \partial_1 F^{\Phi}(l, l_2 + w_2 h) dw_2.$$

From a second order Taylor expansion of $\partial_1 F^{\Phi}(l, l_2 + w_2 h)$ and the boundedness of its derivatives (due to the presence of $\Phi$), we get $\int K_h \left( l_2 - l' \right) dF^{\Phi}(l, l') = \partial_1 F^{\Phi}(l, l_2) dl + O(h^2)$, where $O(h^2)-$rate does not depend on $l$. A Taylor expansion of $K_h$ leads to

$$\begin{aligned} \mathcal{T}_1 &= \frac{1}{h} \int k \left( \frac{l_1 - l}{h} \right) \left[ F_1(F_1^{-1}(\Phi(l))) - \mathbb{F}_{1n}(F_1^{-1}(\Phi(l))) \right] \partial_1 C(\Phi(l), \Phi(l_2)) dl \\ &+ o_P(n^{-1/2}). \end{aligned}$$

Performing a change of variables with $v = [l_1 - l]h^{-1}$, one gets,

$$\mathcal{T}_1 = \int k(v) \left[ \mathbb{F}_{1n}(F_1^{-1}(\Phi(l_1 + vh))) - F_1(F_1^{-1}(\Phi(l_1 + vh))) \right] \partial_1 C(\Phi(l_1 + vh), \Phi(l_2)) dv$$
$$+ o_P(n^{-1/2}). \tag{6.2}$$

Using the differentiability of $\Phi$, one can replace $F_1(F_1^{-1}(\Phi(l_1 + vh)))$ by $F_1(F_1^{-1}(\Phi(l_1))) + vh\Phi'(l_1) + O(h^2)$, uniformly in $l_1$.

Next, define $\hat{g}^*$ as the function such that $\hat{g}^*(M_{1i}, M_{2i}) = \hat{g}(M_{1i}, M_{2i})$, with an obvious extension of the definition of its limit $g^*$. The uniform convergence rate of $\hat{g}^*$ towards $g^*$

is a direct consequence of the uniform convergence rate of $\hat{g}$ towards $g$. We have

$$\begin{aligned}\mathbb{F}_{1n}(F_1^{-1}(\Phi(l_1+vh))) - \mathbb{F}_{1n}(F_1^{-1}(\Phi(l_1))) &= \frac{1}{n}\sum_{i=1}^{n}\delta_{i1}\delta_{i2}\hat{g}^*(M_{i1},M_{i2})\left[\mathbb{1}_{M_{1i}\leq l_1+vh}\right.\\ &\quad \left.-\mathbb{1}_{M_{1i}\leq l_1}\right]\\ &= \frac{1}{n}\sum_{i=1}^{n} f_n(\delta_{i1},\delta_{i2},M_{i1},M_{i2}),\end{aligned}$$

where $f_n$ belongs to a Donsker class $\mathcal{F}'$ (since the class of indicators function is Donsker, and that multiplication by a bounded Donsker class does not modify the Donsker property, see van der Vaart and Wellner (1996)), with $\|f_n\|_2 \to 0$. We then can write

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} f_n(\delta_{i1},\delta_{i2},M_{i1},M_{i2}) &= \int \hat{g}^*(u_1,u_2)\left[\mathbb{1}_{u_1\leq\phi(l_1+vh)} - \mathbb{1}_{u_1\leq\phi(l_1)}\right]\frac{dC(u_1,u_2)}{g^*(u_1,u_2)}\\ &\quad +o_P(n^{-1/2}),\end{aligned}$$

with the $o_P-$rate holding uniformly in $l_1$. In this equation, we can replace $\hat{g}^*$ by $g^*$ up to some residual term which is $O_P(n^{-1/2}h) = o_P(n^{-1/2})$. Indeed,

$$\begin{aligned}&\int\{\hat{g}^*(u_1,u_2)-g^*(u_1,u_2)\}\left[\mathbb{1}_{u_1\leq\phi(l_1+vh)} - \mathbb{1}_{u_1\leq\phi(l_1)}\right]\frac{dC(u_1,u_2)}{g^*(u_1,u_2)}\\ &\qquad = \int_{u_2}\int_{u_1=l_1}^{u_1=l_1+vh}[\hat{g}^*(u_1,u_2)-g^*(u_1,u_2)]\frac{dC(u_1,u_2)}{g^*(u_1,u_2)}\\ &\qquad\qquad \leq C_0 h\sup_{u_1,u_2}|\hat{g}^*(u_1,u_2)-g^*(u_1,u_2)|.\end{aligned}$$

Next, it follows from a Taylor expansion that

$$\mathbb{F}_{1n}(F_1^{-1}(\phi(l_1+vh))) - \mathbb{F}_{1n}(F_1^{-1}(\phi(l_1))) = vh\phi'(l_1)C(\phi(l_1),1) + o_P(n^{-1/2}).$$

This, combined with (6.2), the differentiability of $\partial_1 C$, and the fact that $\int vk(v)dv = 0$, shows that

$$\mathcal{T}_1 = -[\mathbb{F}_{1n}(F_1^{-1}(\phi(l_1))) - F_1(F_1^{-1}(\phi(l_1)))]\partial_1 C(\phi(l_1),\phi(l_2)) + o_P(n^{-1/2})$$

Since $\mathbb{F}_{1n}(F_1^{-1}(\phi(l_1))) = \mathbb{F}_n^*(\phi(l_1),\infty)$, we finally get

$$\hat{\mathfrak{C}}_n^2(u_1,u_2) = \mathbb{F}_n^*(u_1,u_2) - \mathbb{F}_n^*(u_1,\infty)\partial_1 C(u_1,u_2) - \mathbb{F}_n^*(\infty,u_2)\partial_2 C(u_1,u_2) + o_P(n^{-1/2}),$$

and the convergence of $\mathbb{F}_n^*(u_1,u_2)$ leads to the appropriate asymptotic distribution.

## 6.4 Proof of Theorem 4.5

**A. Convergence of $\hat{c}_1$.**

Let $f_{1n}, f_{2n}$ denote the derivatives of $\hat{\mathbb{F}}_{1n}, \hat{\mathbb{F}}_{2n}$. Decompose

$$\begin{aligned}\hat{c}_1(u,v) &= \frac{1}{nh^2}\sum_{i=1}^{n}\frac{W_i k\left(\frac{\hat{\mathbb{F}}_{1n}^{-1}(u)-Y_{1i}}{h}\right)k\left(\frac{\hat{\mathbb{F}}_{2n}^{-1}(v)-Y_{2i}}{h}\right)}{f_{1n}(\hat{\mathbb{F}}_{1n}^{-1}(u))f_{2n}(\hat{\mathbb{F}}_{2n}^{-1}(v))} \\ &\quad + \frac{1}{nh^2}\sum_{i=1}^{n}(W_{in}-W_i)\frac{k\left(\frac{\hat{\mathbb{F}}_{1n}^{-1}(u)-Y_{1i}}{h}\right)k\left(\frac{\hat{\mathbb{F}}_{2n}^{-1}(v)-Y_{2i}}{h}\right)}{f_{1n}(\hat{\mathbb{F}}_{1n}^{-1}(u))f_{2n}(\hat{\mathbb{F}}_{2n}^{-1}(v))} \\ &:= \frac{J_{1n}(u,v)}{f_{1n}(\hat{\mathbb{F}}_{1n}^{-1}(u))f_{2n}(\hat{\mathbb{F}}_{2n}^{-1}(v))} + \frac{J_{2n}(u,v)}{f_{1n}(\hat{\mathbb{F}}_{1n}^{-1}(u))f_{2n}(\hat{\mathbb{F}}_{2n}^{-1}(v))}. \qquad (6.3)\end{aligned}$$

Observe that, for $(u,v) \in \mathcal{C}$, $(\hat{\mathbb{F}}_{1n}^{-1}(u), \hat{\mathbb{F}}_{2n}^{-1}(v)) \in \mathcal{Y}$ with probability tending to one, where $\mathcal{Y} = [0,t_1]\times[0,t_2]$ for some $t_1 < \tau_1$ and $t_2 < \tau_2$. Indeed, recall that $\hat{\mathbb{F}}_{1n}^{-1}(u) = \inf\{f : \hat{\mathbb{F}}_{1n}(t) \geq u\}$. There exists some $\eta > 0$ such that $t_1 = F_1^{-1}(u) + \eta$ is strictly less than $\tau_1$. Moreover $\hat{\mathbb{F}}_{1n}(t_1) = F_1(t_1) + \varepsilon_n(t_1)$ where $\varepsilon_n(t_1)$ tends to zero, and $F_1(t_1) > u$, which shows that $\hat{\mathbb{F}}_{1n}^{-1}(u) \leq t_1$ with probability tending to one. Moreover, since $k$ has compact support, with probability tending to one, only the points $i$ such that $(Y_{1i}, Y_{2i}) \in \mathcal{Y}_\varepsilon$ give a non-zero contribution to the sum, where the definition of $\mathcal{Y}_\varepsilon$ can be found in the proof of Theorem 4.1.

The proof is then composed of several steps. We show first that the second term in (6.3) negligible. Indeed, following Assumption 7 and using previous observation, we require to bound the difference of the weights $(W_{in} - W_i)$ for the indexes $i$ corresponding to observations in $\mathcal{Y}_\varepsilon$. Then, using Assumption 8, we get

$$\sup_{(u,v)\in\mathcal{C}}|J_{2n}(u,v)| \leq O_P(\eta_n) \times \sup_{(y_1,y_2)\in\mathcal{Y}_\varepsilon}\left|\frac{1}{nh^2}\sum_{i=1}^{n}k\left(\frac{y_1-Y_{1i}}{h}\right)k\left(\frac{y_2-Y_{2i}}{h}\right)\right|,$$

where the supremum on the right-hand side is $O_P(1)$, by Theorem 4 in Einmahl and Mason (2005). Notice that the denominators in (6.3) are bounded away from zero. To see that, it suffices to write, for $j = 1, 2$,

$$f_{jn}(\hat{\mathbb{F}}_{jn}^{-1}(u)) = \sum_{i=1}^{n} W_i k\left(\frac{\hat{\mathbb{F}}_{jn}^{-1}(u) - Y_{1i}}{h}\right) + O_P(\eta_n). \qquad (6.4)$$

Thus, there exists $\varepsilon > 0$, such that, with probability tending to one

$$\inf_{u\in\mathcal{C}} f_{jn}(\hat{\mathbb{F}}_{jn}^{-1}(u)) = \inf_{u\in\mathcal{C}} f_j(\hat{\mathbb{F}}_{jn}^{-1}(u)) + O_P(\eta_n) \geq \inf_{x\in t_j+\varepsilon} f_j(x) + O_P(\eta_n) > \frac{c}{2},$$

and the second term in (6.3) is therefore negligible.

Let us study now the first term. It follows from Theorem 4 in Einmahl and Mason (2005) that

$$\sup_{(y_1,y_2)\in\mathcal{Y}_\varepsilon}\left|n^{-1}h^{-2}\sum_{i=1}^{n}W_i k\left(\frac{y_1-Y_{1i}}{h}\right)k\left(\frac{y_2-Y_{2i}}{h}\right)\right.$$
$$\left.-h^{-2}\int k\left(\frac{y_1-t_1}{h}\right)k\left(\frac{y_2-t_2}{h}\right)d\mathbb{P}_{(T_1,T_2)}(t_1,t_2)\right| \;=\; O_P\left(\frac{[\log n]^{1/2}}{h\sqrt{n}}\right),$$

where we used (2.2). Hence,

$$J_{1n}(u,v)=h^{-2}\int k\left(\frac{\hat{\mathbb{F}}_{1n}^{-1}(u)-t_1}{h}\right)k\left(\frac{\hat{\mathbb{F}}_{2n}^{-1}(v)-t_2}{h}\right)d\mathbb{P}_{(T_1,T_2)}(t_1,t_2)+O_P\left(\frac{[\log n]^{1/2}}{h\sqrt{n}}\right).$$

Using a second order Taylor expansion, and the fact that the derivatives of the density $f$ up to order 2 are uniformly bounded (Assumption 8), we get

$$\sup_{u,v}|J_{1n}(u,v)-f(\hat{\mathbb{F}}_{1n}^{-1}(u),\hat{\mathbb{F}}_{2n}^{-1}(v))|=O_P(\eta_n).$$

Again by Einmahl and Mason (2005), there exists $\varepsilon$ such as, for $j=1,2$, with probability tending to one,

$$\sup_{u:(u,v)\in\mathcal{C}}|f_{jn}(\hat{\mathbb{F}}_{jn}^{-1}(u))-f_j(\hat{\mathbb{F}}_{jn}^{-1}(u))|\leq\sup_{x<t_j+\varepsilon}|f_{jn}(x)-f_j(x)|=O_P(\eta_n).$$

Therefore, we have

$$\sup_{(u,v)\in\mathcal{C}}\left|\hat{c}_1(u,v)-\frac{f(\hat{\mathbb{F}}_{1n}^{-1}(u),\hat{\mathbb{F}}_{2n}^{-1}(v))}{f_1(\hat{\mathbb{F}}_{1n}^{-1}(u))f_2(\hat{\mathbb{F}}_{2}^{-1}(v))}\right|=O_P(\eta_n).$$

To conclude it remains to prove that

$$J_n:=\sup_{(u,v)\in\mathcal{C}}\left|\frac{f(\hat{\mathbb{F}}_{1n}^{-1}(u),\hat{\mathbb{F}}_{2n}^{-1}(v))}{f_1(\hat{\mathbb{F}}_{1n}^{-1}(u))f_2(\hat{\mathbb{F}}_{2}^{-1}(v))}-\frac{f(F_1^{-1}(u),F_2^{-1}(v))}{f_1(F_1^{-1}(u))f_2(F_2^{-1}(v))}\right|=O_P(\eta_n).$$

By the assumption of Theorem, the density of $(T_1,T_2)$ is twice continuously differentiable with bounded second derivatives and the marginal densities are bounded away from zero, so by using Taylor expansion, there exists a constant L, such as

$$\begin{aligned}J_n &= \sup_{u,v\in\mathcal{C}}\left|c(\hat{\mathbb{F}}_{1n}^{-1}(u),\hat{\mathbb{F}}_{2n}^{-1}(v))-c(F_1^{-1}(u),F_2^{-1}(v)\right| \\ &\leq L\left(\sup_u|\hat{\mathbb{F}}_{1n}^{-1}(u)-F_1^{-1}(u)|+\sup_v|\hat{\mathbb{F}}_{2n}^{-1}(v)-F_2^{-1}(v)|\right)\end{aligned}$$

We will show now that the bound from the previous inequality is of the order of $O_P(\eta_n)$. For an arbitrary constant $M$ for the first term (the second term is analogous) we have

$$\begin{aligned}
& P(\sup_u |\hat{\mathbb{F}}_{1n}^{-1}(u) - F_1^{-1}(u)| > M\eta_n) \\
\leq\ & P(\exists u : u > \mathbb{F}_{1n}(F_1^{-1}(u) + M\eta_n)) + P(\exists u : u < \mathbb{F}_{1n}(F_1^{-1}(u) - M\eta_n)) \\
\leq\ & P\left\{\sup_u |\mathbb{F}_{1n}(F_1^{-1}(u) + M\eta_n) - F_1(F_1^{-1}(u) + M\eta_n)| > \inf_u (F_1(F_1^{-1}(u) + M\eta_n) - u)\right\} \\
& + P\left\{\sup_u |\mathbb{F}_{1n}(F_1^{-1}(u) - M\eta_n) - F_1(F_1^{-1}(u) - M\eta_n)| > \inf_u (u - F_1(F_1^{-1}(u) - M\eta_n))\right\}.
\end{aligned}$$

By Corollary 4.3, $\sup_u |F_{1n}(u) - F_1(u)| = O_P(h^2)$, and, choosing $M$ sufficiently large, we obtain $P(\sup_u |\hat{\mathbb{F}}_{1n}^{-1}(u) - F_1^{-1}(u)| > M\eta_n) \to 0$.

**B. Convergence of $\hat{c}_2$.**

We have

$$\hat{c}_2(u,v) = \frac{\sum_{i=1}^n W_{in} k\left(\frac{\Phi^{-1}(u) - \Phi^{-1}[\mathbb{F}_{1n}(Y_{1i})]}{h}\right) k\left(\frac{\Phi^{-1}(v) - \Phi^{-1}[\mathbb{F}_{2n}(Y_{2i})]}{h}\right)}{nh^2 \Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v))},$$

Let us use a notation $y_{1i}^n(u) := \Phi^{-1}(u) - \Phi^{-1}[\mathbb{F}_{1n}(Y_{1i})]$ and $y_{1i}(u) := \Phi^{-1}(u) - \Phi^{-1}[F_1(Y_{1i})]$, with a similar definition for $y_{2i}^n$ and $y_{2i}$. Using a 4-th order Taylor expansion, we get

$$\begin{aligned}
\hat{c}_2(u,v) &= \frac{1}{nh^2 \Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v))} \sum_{i=1}^n W_{in} k\left(\frac{y_{1i}^n(u)}{h}\right) k\left(\frac{y_{2i}^n(v)}{h}\right) \\
&= \frac{1}{\Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v))} \left\{\frac{1}{nh^2} \sum_{i=1}^n W_{in} k\left(\frac{y_{1i}(u)}{h}\right) k\left(\frac{y_{2i}(v)}{h}\right) + I_n(u,v)\right\},
\end{aligned}$$

where

$$\begin{aligned}
I_n(u,v) &= \frac{1}{nh^2} \sum_{i=1}^n W_{in} k\left(\frac{y_{1i}^n(u)}{h}\right) k\left(\frac{y_{2i}^n(v)}{h}\right) - \frac{1}{nh^2} \sum_{i=1}^n W_{in} k\left(\frac{y_{1i}(u)}{h}\right) k\left(\frac{y_{2i}(v)}{h}\right) \\
&= \sum_{m=1}^3 \sum_{l=0}^m \sum_{i=1}^n \frac{W_{in}}{nh^2} k^{(l)}\left(\frac{y_{1i}(u)}{h}\right) k^{(m-l)}\left(\frac{y_{2i}(v)}{h}\right) \frac{(y_{1i}^n(u) - y_{1i}(u))^l (y_{2i}^n(v) - y_{2i}(v))^{m-l}}{m!\, h^m} \\
&+ \frac{1}{nh^2} \sum_{l=0}^4 \sum_{i=1}^n W_{in} k^{(l)}\left(\frac{\tilde{y}_{1i}(u)}{h}\right) k^{(4-l)}\left(\frac{\tilde{y}_{2i}(v)}{h}\right) \frac{(y_{1i}^n(u) - y_{1i}(u))^l (y_{2i}^n(v) - y_{2i}(v))^{4-l}}{4!\, h^4} \\
&:= \sum_{m=1}^3 \sum_{l=0}^m I_{ml}^n(u,v) + \sum_{l=0}^4 I_{4l}^n(u,v),
\end{aligned}$$

using the notation $k^{(l)}$ for the $l-$th derivative of $k$, and $\tilde{y}_{1i}(u)$ (resp. $\tilde{y}_{2i}(v)$) for some point between $y_{1i}(u)$ and $y_{1i}^n(u)$ (resp. $y_{2i}(v)$ and $y_{2i}^n(v)$).

We get

$$\sup_{(u,v)\in\mathcal{C}} |\hat{c}_2(u,v) - c(u,v)| \leq \sup_{x,y\in\mathcal{C}} \left| \frac{\sum_{i=1}^n W_{in} k\left(\frac{y_{1i}(u)}{h}\right) k\left(\frac{y_{2i}(v)}{h}\right)}{nh^2 \Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v))} - c(u,v) \right| + \sup_{(u,v)\in\mathcal{C}} |I_n(u,v)|. \quad (6.5)$$

The proof consists of showing separately the convergence rate of each term in this decomposition.

**1. Convergence of the first term in (6.5).**

We will show that this term converges at the rate $[\log n]^{1/2} n^{-1/2} h^{-1} + h^2$. Let us introduce the sets $\mathcal{U} = \{u : \exists v \ s.t. \ (u,v) \in \mathcal{C}\}$, and $\mathcal{V} = \{v : \exists u \ s.t. \ (u,v) \in \mathcal{C}\}$.

The first term in (6.5) can be written as

$$\sup_{(u,v)\in\mathcal{C}} \left| \frac{\left\{ \frac{1}{nh^2} \sum_{i=1}^n W_{in} k\left(\frac{y_{1i}(u)}{h}\right) k\left(\frac{y_{2i}(v)}{h}\right) - c(u,v)\Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v)) \right\}}{\Phi'(\Phi^{-1}(u))\Phi'(\Phi^{-1}(v))} \right|$$

$$\leq \sup_{x\in\Phi^{-1}(\mathcal{U}), y\in\Phi^{-1}(\mathcal{V})} \left| \left\{ \frac{1}{nh^2} \sum_{i=1}^n W_{in} k\left(\frac{x - \Phi^{-1}[F_1(Y_{1i})]}{h}\right) k\left(\frac{y - \Phi^{-1}[F_2(Y_{2i})]}{h}\right) \right.\right.$$

$$\left.\left. - \quad c(\Phi(x),\Phi(y))\Phi'(x)\Phi'(y) \right\} [\Phi'(x)\Phi'(y)]^{-1} \right|$$

Note that $c(\Phi(x),\Phi(y))\Phi'(x)\Phi'(y)$ is the density of the distribution function of random variables $(\Phi^{-1}(F_1(Y_1)), \Phi^{-1}(F_2(Y_2)))$, evaluated at the point $(x,y)$. As stated in the proof of A.2, one can use the compactness of the support of $k$ to deduce that only the points $i$ corresponding to observations in $\mathcal{Y}_\varepsilon$ contribute to the sum. Therefore, one can replace (up to some negligible term) $W_{in}$ by $W_i$ using Assumption 7. It then follows from Einmahl and Mason (2005) that the resulting quantity converges towards 0 at rate $[\log n]^{1/2} n^{-1/2} h^{-1} + h^2$, the rate $h^2$ coming from the convergence rate of the expectation (deduced from classical arguments on kernel estimators, and the regularity of $c(\Phi(x),\Phi(y))\Phi'(x)\Phi'(y)$).

**2. Convergence of $I_n(u,v)$.**

Let us now consider the second term of (6.5),

$$\sup_{(u,v)\in\mathcal{C}} |I_n(u,v)| \leq \sum_{m=1}^{3} \sum_{l=0}^{m} \sup_{(u,v)\in\mathcal{C}} |I_{ml}^n(u,v)| + \sum_{l=0}^{4} \sup_{(u,v)\in\mathcal{C}} |I_{4l}^n(u,v)|.$$

First observe that

$$\sup_{u\in\mathcal{U}} |y_{1i}(u) - y_{1i}^n(u)| \leq \sup_{x<t_1+\varepsilon} |\Phi^{-1}(F_1(x)) - \Phi^{-1}(F_{1n}(x))|,$$

for some $t_1$ such that $t_1 + \varepsilon < \tau_1$, since $1 \notin \mathcal{U}$. Since $\Phi^{-1}$ has a continuous bounded derivative on $\mathcal{T}_1$ from Assumption 5, $\sup_{u\in\mathcal{U}} |y_{1i}(u) - y_{1i}^n(u)|$ has the same convergence rate as $\sup_{t<t_j+\varepsilon} |F_j(t) - F_{jn}(t)| = O_P(n^{-1/2})$. The same is true considering $\sup_{v\in\mathcal{V}} |y_{2i}(v) - y_{2i}^n(v)|$. Therefore, one gets

$$\begin{aligned}\sup_{(u,v)\in\mathcal{C}} |I_{4l}^n(u,v)| &\leq \frac{1}{h^6} \sup_{(u,v)\in\mathcal{C}} |(y_{1i}^n(u) - y_{1i}(u))^l (y_{2i}^n(v) - y_{2i}(v))^{4-l}| \times O_{\mathbb{P}}(1) \\ &= \frac{1}{n^2 h^6} \times O_{\mathbb{P}}(1).\end{aligned}$$

For the terms up to the third order in the Taylor expansion, one obtains, for $m = 1, ..., 3$,

$$\begin{aligned}\sup_{(u,v)\in\mathcal{C}} |I_{ml}^n(u,v)| &\leq \frac{1}{m! h^m} \sup_{(u,v)\in\mathcal{C}} |(y_{1i}^n(u) - y_{1i}(u))^l (y_{2i}^n(v) - y_{2i}(v))^{m-l}| \\ &\quad \times \sup_{(u,v)\in\mathcal{C}} \left[ \frac{1}{nh^2} \sum_{i=1}^n W_{in} k^{(l)} \left( \frac{y_{1i}(u)}{h} \right) k^{(m-l)} \left( \frac{y_{2i}(v)}{h} \right) \right] \\ &= \frac{1}{n^{m/2} h^m} \times O_{\mathbb{P}}(1).\end{aligned}$$

The condition on $h$ implies that $\sup_{(u,v)\in\mathcal{C}} |I_n(u,v)| = O_{\mathbb{P}}(h^{-1} n^{-1/2})$.

## 6.5 Properties of the functions $\eta^\psi$ is the Examples

The aim of this section is to show that part 3 of Assumption 3 holds for the functions $\eta^\psi$ corresponding to the three standard examples we consider. We focus on Example 1, where $\eta^\psi$ has the simplest form, which can be found in Stute (1996). We then explain how these arguments may be extended to Examples 2 and 3.

In the case of Example 1, we have

$$\begin{aligned}\eta^\psi(Y_1, Y_2, \delta_1, \delta_2) &= \frac{(1-\delta_1) \int_{Y_1}^{\tau_1} \int_{-\infty}^{\infty} \psi(y_1, y_2) dF(y_1, y_2)}{1 - H(Y_1)} \\ &\quad - \int \frac{\mathbf{1}_{Y_1 \geq y_1} [1 - F(y_1)] \{ \int_{y_1}^{\infty} \int_{-\infty}^{\infty} \psi(t_1, t_2) dF(t_1, t_2) \} dG(y_1, y_2)}{[1 - H(y_1)]^2 [1 - G(y_1)]},\end{aligned}$$

where $H(y_1) = \mathbb{P}(Y_1 \geq y_1)$. From this expression, one can see that, in the case where $\psi$ is a nonnegative function, we can write

$$\eta^\psi(Y_1, Y_2, \delta_1, \delta_2) = \frac{(1-\delta_1) f_1(Y_1)}{1 - H(Y_1)} + f_2(Y_1),$$

where $f_1$ and $f_2$ are monotone functions. Moreover, if $\psi$ satisfies the requirements of part 2 of Assumption 3, $f_1$ and $f_2$ are bounded, and $f_1(Y_1) = 0$ for $Y_1 > t_1$. Then, one can see

that $\eta^{\psi}$ belongs to the class $\mathcal{F}_2 = (1-\delta_1)[1-H(Y_1)]^{-1}\mathbb{1}_{Y_1 \leq t_1} \times \mathcal{F}_1 + \mathcal{F}_1$, where $\mathcal{F}_1$ is the class of positive functions bounded by some absolute constant (this class is Donsker from Theorem 2.7.5 in van der Vaart and Wellner (1996)). Therefore, from Examples 2.10.7 and 2.10.10 in van der Vaart and Wellner (1996), $\mathcal{F}_2$ class is Donsker. The functions $\psi_{h,y_1,y_2}$ are not non-negative, but they are the sum of two non-negative functions. Moreover, $\psi \to \eta^{\psi}$ is linear, therefore the family of functions $\eta^{\psi_{h,y_1,y_2}}$ is Donsker from Example 2.10.7 in van der Vaart and Wellner (1996). For Examples 2 and 3, the arguments are similar, since $\eta^{\psi}$ can always be decomposed into a some of bounded monotonic terms.

To show that the expectation of $(\eta^{\psi_{h,y_1,y_2}})^2$ tends to zero, observe that, for $j = 1, 2$,

$$|f_j(Y_1)| \leq C \sup_{(y,z)\in\mathcal{Y}} \left| \int_y^{\tau_1} \int_{-\infty}^{\infty} \psi(t_1,t_2) dF(t_1,t_2) \right|,$$

for some constant $C$. Similar bounds can be found in Examples 2 and 3. Next, observe that, from Fubini's Theorem,

$$\int_y^{\tau_1} \int_{-\infty}^{\infty} \psi_{h,y_1,y_2}(t_1,t_2) dF(t_1,t_2) =$$

$$\int_{-y}^{\infty} \int_{-\infty}^{\infty} [F(y_1 - hw_1, y_2 - hw_2) - F(y_1,y_2)] k(w_1) k(w_2) dw_1 dw_2.$$

A Taylor expansion then shows that, for $(y_1, y_2) \in \mathcal{Y}$, $\|\eta^{\psi_{h,y_1,y_2}}\|_{\infty} \leq \tilde{C}h$, which tends to zero.

# References


Akritas, M. G. and Van Keilegom, I. (2003). Estimation of bivariate and marginal distributions with censored data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2):457–471.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2007). Copulas for Finance - A Reading Guide and Some Applications. *Social Science Research Network Working Paper Series.*

Carriere, J. (2000). Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, 2000(1):17–32.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula methods in finance.* Wiley Finance Series. John Wiley & Sons Ltd., Chichester.

Choros, B., Ibragimov, R., and Permiakova, E. (2010). Copula estimation. In F. Durante, W. Haerdle, P. J. and T. Rychlik, e., editors, *Workshop on Copula Theory and its Applications. Lecture Notes in Statistics- Proceedings*. Springer.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, 16(4):1475–1489.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65(6):274–292.

Denuit, M. P. O. and Van Keilegom, I. (2006). Bivariate archimedean copula models for censored data in non-life insurance. *Journal of Actuarial Practice*, 13:5–32.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3):1380–1403.

Embrechts, P., Lindskog, F., and McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S., editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. North-Holland.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.*, 95(1):119–152.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Fermanian, J.-D. and Scaillet, O. (2005). Some statistical pitfalls in copula modelling for financial applications. In Klein, E. e., editor, *Capital Formation, Governance and Banking*, pages 59–74. New York: Nova Science Publishers.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York.

Frees, E., Carriere, J., and Valdez, E. (1996). Annuity valuation with dependent mortality. *Journal of Risk and Insurance*, 63(2):229–261.

Frees, E. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.*, 88(423):1034–1043.

Gijbels, I. and Veraverbeke, N. (1991). Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator. *Ann. Statist.*, 19(3):1457–1470.

Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, 11(1):49–58.

Gribkova, S., Lopez, O., and Saint-Pierre, P. (2013). A simplified model for studying bivariate mortality under right-censoring. *J. Multivariate Anal.*, 115:181–192.

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Lopez, O. (2012). A generalization of Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications in model-checking for survival copula models. *Insurance: Mathematics and Economics*, 51(3):505 – 516.

Lopez, O. and Saint-Pierre, P. (2012). Bivariate censored regression relying on a new estimator of the joint distribution function. *J. Statist. Plann. Inference*, 142(8):2440–2453.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance Math. Econom.*, 43(2):234–244.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37(5B):3023–3058.

Patton, A. J. (2012). A review of copula models for economic time series. *J. Multivariate Anal.*, 110:4–18.

Pruitt, R. C. (1991). On negative mass assigned by the bivariate Kaplan-Meier estimator. *Ann. Statist.*, 19(1):443–453.

Satten, G. A. and Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.*, 55(3):207–210.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.

Sklar, A. (1959). Fonctions de répartition à $n$ dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.*, 45(1):89–103.

Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.*, 23(2):422–439.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.*, 23(4):461–471.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33(3):357–375.

van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.*, 24(2):596–627.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Wang, W. and Wells, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, 84(4):863–880.

Wang, W. and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.*, 95(449):62–76. With a comment by Edsel A. Peña and a rejoinder by the authors.

Youn, H. and Shemyakin, A. (2001). Statistical aspects of joint life insurance pricing. *1999 Proceedings of the Business and Statistics Section of the American Statistical Association*, pages 34–38.