A copula-graphic estimator for the conditional survival function under dependent censoring

A copula-graphic estimator for the conditional survival function under dependent censoring.

Roel Braekers and Noël Veraverbeke
Limburgs Universitair Centrum, Universitaire Campus
B-3590 Diepenbeek, Belgium.
roel.braekers@luc.ac.be, noel.veraverbeke@luc.ac.be

Abstract

In survival analysis, it is very common to assume that the lifetime variable and the censoring variable are independent. In this case, the product limit estimator is the standard non-parametric estimator for the distribution function of the lifetime variable. When the assumption of independence is not satisfied, Zheng and Klein (1995) proposed a copula-graphic estimator where the dependence between lifetime and censoring variable is described by a known copula. Rivest and Wells (2001) derived an explicit form for this estimator if the copula is Archimedean.

In this paper, we extend the estimator of Rivest and Wells (2001) to the fixed design regression case. For our copula-graphic estimator, we find an asymptotic representation and prove weak convergence to a Gaussian limit. We perform a sensitivity analysis to assess the influence of a misspecified copula function on the estimator. Furthermore we illustrate the estimation method with a dataset on survival of Atlantic halibut.

1 Introduction

At fixed design points $0 \leq x_1 \leq \ldots \leq x_n \leq 1$, we have nonnegative responses $Y_1, \ldots, Y_n$ such as survival times or failure times. These responses are independent random variables and the distribution function of the response Y_i at x_i will be denoted by $F_{x_i}(t) = P(Y_i \leq t)$.

In many clinical or industrial trials, the responses $Y_1, \ldots, Y_n$ are subject to random right censoring. For each response, there is a censoring variable C_i with conditional distribution function $G_{x_i}(t) = P(C_i \leq t)$. The observed random variables at design point x_i are in

fact Z_i and δ_i $(i = 1, \ldots, n)$, with

$$Z_i = \min(Y_i, C_i) \qquad \text{and} \qquad \delta_i = I(Y_i \leq C_i).$$

At a given fixed design value $x \in [0, 1]$, we write F_x, G_x, H_x for the distribution function of respectively the response Y_x, the censoring variable C_x and the observed variable $Z_x = \min(Y_x, C_x)$ at x. Also we will write $\delta_x = I(Y_x \leq C_x)$. Note that for the design variables x_i, we write $Y_i, C_i, Z_i, F_i, \ldots$ instead of $Y_{x_i}, C_{x_i}, Z_{x_i}, F_{x_i}, \ldots$.

In order to estimate uniquely the distribution function F_x from the observed data, we have to make an assumption about the dependence between the Y_i and C_i for each i (Tsiatis 1975). It is very common in survival analysis to assume independence between these random variables (conditional on the covariate). However we see that in some practical situations this assumption clearly does not hold. For example in medicine when the event of interest is death due to a given disease and the censoring event is death due to other diseases. In industrial testing, it may occur that some piece of equipment is taken away (is censored) because it shows some sign of future failure. Therefore a dependence model is used in which the dependence structure is given by specifying a copula for the joint distribution of Y_x and C_x. Assume that the joint survival function of the response Y_x and the censoring variable C_x at x can be written as

$$S_x(t_1, t_2) = P(Y_x > t_1, C_x > t_2) = \mathcal{C}_x(\bar{F}_x(t_1), \bar{G}_x(t_2))$$

where $\mathcal{C}_x$ is a known copula function depending in a general way on x and $\bar{F}_x(t)$ (resp. $\bar{G}_x(t)$) is the survival function of Y_x (resp. C_x) at x. Without covariates x, this idea was introduced by Zheng and Klein (1995). However their copula-graphic estimator had no closed form expression. Rivest and Wells (2001) got around this problem by focusing on the class of Archimedean copulas. In this work, we will extend their ideas to the fixed design regression case.

We assume that at a fixed design value $x \in [0, 1]$, the joint survival function is given by

$$S_x(t_1, t_2) = \varphi_x^{[-1]}(\varphi_x(\bar{F}_x(t_1)) + \varphi_x(\bar{G}_x(t_2))) \tag{1}$$

where, for each x, $\varphi_x : [0, 1] \to [0, +\infty]$ is a known continuous, convex, strictly decreasing function with $\varphi_x(1) = 0$. $\varphi_x^{[-1]}$ is the pseudo-inverse of φ_x, as defined in Nelsen (1999) and given by

$$\varphi_x^{[-1]}(s) = \begin{cases} \varphi_x^{-1}(s) & 0 \leq s \leq \varphi_x(0) \\ 0 & \varphi_x(0) \leq s \leq +\infty \end{cases}.$$

We note from (1) that,

$$1 - H_x(t) = \bar{H}_x(t) = S_x(t, t) = \varphi_x^{[-1]}(\varphi_x(\bar{F}_x(t)) + \varphi_x(\bar{G}_x(t))).$$

This relation will be used to find a conditional distribution estimator F_{xh} for F_x where $x \in]0,1[$ is a fixed design value. We organized this paper as follows. In Section 2, we define the distribution function estimator F_{xh} and show that it is an extension of the Beran estimator, as it was studied by Van Keilegom and Veraverbeke (1996, 1997a and 1997b). After specifying some assumptions in Section 3, we give for this estimator an asymptotic representation and a weak convergence result in Section 4. In Section 5, we perform a sensitivity analysis to investigate the bias that is introduced by a misspecification of the copula function. In Section 6 we apply the estimator to a practical situation in which we explore different choices for the generator function φ_x. In the Appendix we give the proofs of the results in Section 4.

2 Copula-graphic estimator

For a fixed design value $x \in]0,1[$, we derive an estimator for the distribution function $F_x(t)$. Since we only have observations at the design points $x_1, \ldots, x_n$, we use smoothing weights to give observations at a design point close to x a larger contribution in our estimator than observations at design points far away from x. In a fixed design regression it is natural to work with Gasser-Müller weights,

$$w_{ni}(x, h_n) = \frac{1}{c_n(x, h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x - z}{h_n}\right) dz \qquad (i = 1, \ldots, n), \tag{2}$$

$$c_n(x, h_n) = \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x - z}{h_n}\right) dz \tag{3}$$

where $x_0 = 0$, K is a known probability density function, called the kernel and $\{h_n\}$ is a sequence of positive constants, tending to zero as $n \to +\infty$, called the bandwidth sequence.

Let us assume that there are no ties in the observations. To find an estimator for $\bar{F}_x(t)$ (resp. $\bar{G}_x(t)$) at the design point x, we work as Rivest and Wells (2001) and look for the right continuous step function $\bar{F}_{xh}(t)$ (resp. $\bar{G}_{xh}(t)$) with $\bar{F}_{xh}(0) = 1$ (resp. $\bar{G}_{xh}(0) = 1$), which has jumps at the points Z_i with $\delta_i = 1$ (resp. $\delta_i = 0$) satisfying

$$\varphi_x^{[-1]}(\varphi_x(\bar{F}_{xh}(Z_i)) + \varphi_x(\bar{G}_{xh}(Z_i))) = \bar{H}_{xh}(Z_i)$$

where $\bar{H}_{xh}(t) = \sum_{i=1}^{n} w_{ni}(x, h_n) I(Z_i > t)$.

To get a closed form expression for $\bar{F}_{xh}$, we take a point Z_i with $\delta_i = 1$. The function $\bar{G}_{xh}$

has not jump in this point i.e. $\bar{G}_{xh}(Z_i^-) = \bar{G}_{xh}(Z_i)$, and the jump of $\bar{F}_{xh}$ at Z_i satisfies

$$\begin{aligned}\varphi_x(\bar{F}_{xh}(Z_i^-)) - \varphi_x(\bar{F}_{xh}(Z_i)) &= \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i)) \\ &= \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x,h_n)).\end{aligned}$$

Hence

$$\begin{aligned}\varphi_x(\bar{F}_{xh}(t)) &= -\sum_{Z_i \le t, \delta_i = 1} \varphi_x(\bar{F}_{xh}(Z_i^-)) - \varphi_x(\bar{F}_{xh}(Z_i)) \\ &= -\sum_{Z_i \le t, \delta_i = 1} \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x,h_n))\end{aligned}$$

and

$$\bar{F}_{xh}(t) = \varphi_x^{[-1]}\left(-\sum_{Z_i \le t, \delta_i = 1} \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x,h_n))\right). \tag{4}$$

Since the argument of $\varphi_x^{[-1]}$ is never larger then $\varphi_x(0)$, we can replace in (4), without any complications, the pseudo inverse $\varphi_x^{[-1]}$ by the inverse φ_x^{-1}. Furthermore we note that this estimator in general does not tend to 0 as $t \to +\infty$. In order to have a proper distribution estimator, we use the modification

$$\bar{F}_{xh}(t) = \varphi_x^{-1}\left(-\sum_{Z_i \le t, \delta_i = 1} \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x,h_n))\right) I(t < Z_{(n)}) \tag{5}$$

where $Z_{(n)}$ is the largest order statistic in the sample $Z_1, \ldots, Z_n$. When we take the independent copula ($\varphi_x(t) = -\log(t)$), we note that this estimator becomes equal to the Beran estimator given by

$$F_{xh}(t) = 1 - \left\{\prod_{Z_{(i)} \le t}\left(1 - \frac{w_{n(i)}(x,h_n)}{1 - \sum_{j=1}^{i-1} w_{n(j)}(x,h_n)}\right)^{\delta_{(i)}}\right\} I(t < Z_{(n)})$$

as it was studied by Van Keilegom and Veraverbeke (1996, 1997a and 1997b).

3 Regularity conditions

For the design points $x_1, \ldots, x_n$ we write $\underline{\Delta}_n = \min_{1 \le i \le n}(x_i - x_{i-1})$ and $\bar{\Delta}_n = \max_{1 \le i \le n}(x_i - x_{i-1})$. The notations $||K||_\infty = \sup_{u \in I\!R} K(u)$, $||K||_2^2 = \int_{-\infty}^{+\infty} K^2(u)du$, $\mu_1^K = \int_{-\infty}^{+\infty} uK(u)du$, $\mu_2^K = \int_{-\infty}^{+\infty} u^2 K(u)du$ will be used for the kernel K.

We use the following assumptions on the design and on the kernel.

(C1) $x_n \to 1$, $\bar{\Delta}_n = O(n^{-1})$, $\bar{\Delta}_n - \underline{\Delta}_n = o(n^{-1})$.

(C2) K is a probability density function with finite support $[-M, M]$ for some $M > 0$, $\mu_1^K = 0$ and K Lipschitz of order 1.

The assumption (C1) expresses that the chosen design points are asymptotically equidistant points, selected uniformly over the whole interval $[0, 1]$. This implies that, for $c_n(x, h_n)$ defined in (3), $c_n(x, h_n) = 1$ for n sufficiently large. Therefore we may take $c_n(x, h_n) = 1$ in all proofs of the asymptotic results.

If L is any (sub)distribution, then T_L denotes the right endpoint of its support ($T_L = \inf\{t : L(t) = L(+\infty)\}$). Here we have that $T_{H_x} \leq \min(T_{F_x}, T_{G_x})$ where we attain the equality in case $\varphi_x(0) = +\infty$. For $\varphi_x(0) < +\infty$, it depends on the function φ_x whether or not we have an equality. To obtain our results, we need some smoothness conditions on the functions $H_x(t) = P(Z_x \leq t)$ and $H_x^u(t) = P(Z_x \leq t, \delta_x = 1)$. For a fixed $T > 0$,

(C3) $\dot{L}_x(t) = \frac{\partial}{\partial x} L_x(t)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$

(C4) $L'_x(t) = \frac{\partial}{\partial t} L_x(t)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$

(C5) $\ddot{L}_x(t) = \frac{\partial^2}{\partial x^2} L_x(t)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$

(C6) $L''_x(t) = \frac{\partial^2}{\partial t^2} L_x(t)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$

(C7) $\dot{L}'_x(t) = \frac{\partial^2}{\partial x \partial t} L_x(t)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$.

The generator $\varphi_x(v)$ of the Archimedean copula needs to satisfy the following properties.

(C8) $\varphi'_x(v) = \frac{\partial}{\partial v} \varphi_x(v)$ and $\varphi''_x(v) = \frac{\partial^2}{\partial v^2} \varphi_x(v)$ are Lipschitz in the x-direction with a bounded Lipschitz constant, and $\varphi'''_x(v) = \frac{\partial^3}{\partial v^3} \varphi_x(v) \leq 0$ exists and is continuous in $(x, v) \in [0, 1] \times]0, 1]$.

These assumptions and the fact that φ_x is a generator for an Archimedean copula, give that $\varphi'_x(v)$ is monotone increasing with $\varphi'_x(v) < 0$ and $\varphi''_x(v)$ is monotone decreasing with $\varphi''_x(v) \geq 0$.

4 Asymptotic results

In this section we give some asymptotic results for the copula-graphic estimator $F_{xh}(t)$. We show an asymptotic representation for this estimator in Theorem 1 and prove in Theorem 2 the weak convergence of the related empirical process $(nh_n)^{-1/2}(F_{xh}(\cdot)-F_x(\cdot))$ in the space $l^\infty[0,T]$ of uniformly bounded functions on $[0,T]$, endowed with the uniform topology. The proofs of these theorems are put in the Appendix at the end of this paper. Before we show these results, we give a lemma about the distribution function F_x.

Lemma 1. If $H_x(t)$ and $H_x^u(t)$ satisfy (C4) in $[0,1]\times[0,T]$ with $T < T_{H_x}$ and $\varphi_x'(v)$ exists on $[0,1]\times]0,1]$, then under (1),

$$\bar{F}_x(t) = \varphi_x^{-1}\left(-\int_0^t \varphi_x'(\bar{H}_x(s))dH_x^u(s)\right).$$

Proof. Under (1) and with Tsiatis (1975), we get that

$$H_x^{u\prime}(t) = -\frac{\partial}{\partial t_1}S_x(t_1,t_2)\Big|_{t_1=t_2=t} = \frac{\varphi_x'(\bar{F}_x(t))F_x'(t)}{\varphi_x'(\bar{H}_x(t))}.$$

This leads to

$$\varphi_x^{-1}\left(-\int_0^t \varphi_x'(\bar{H}_x(s))dH_x^u(s)\right) = \varphi_x^{-1}\left(-\int_1^{\bar{F}_x(t)} \varphi_x'(w)dw\right) = \bar{F}_x(t).$$

Theorem 1. Assume (C1), (C2), $H_x(t)$ and $H_x^u(t)$ satisfy (C5), (C6) and (C7) in $[0,T]$ with $T < T_{H_x}$, φ_x satisfies (C8), $h_n \to 0$, $\dfrac{\log n}{nh_n} \to 0$, $\dfrac{nh_n^5}{\log n} = O(1)$. Then, under (1) as $n \to +\infty$,

$$F_{xh}(t) - F_x(t) = \sum_{i=1}^n w_{ni}(x,h_n)g_{tx}(Z_i,\delta_i) + R_n(t)$$

where

$$\begin{aligned} g_{tx}(Z_i,\delta_i) &= \frac{-1}{\varphi_x'(\bar{F}_x(t))}\left[\int_0^t \varphi_x''(\bar{H}_x(s))(I(Z_i\le s) - H_x(s))dH_x^u(s)\right. \\ &\quad - \varphi_x'(\bar{H}_x(t))(I(Z_i\le t,\delta_i=1) - H_x^u(t)) \\ &\quad \left. - \int_0^t \varphi_x''(\bar{H}_x(s))(I(Z_i\le s,\delta_i=1) - H_x^u(s))dH_x(s)\right] \end{aligned}$$

and $\sup\limits_{0\le t\le T}|R_n(t)| = O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s.

Theorem 2. Assume (C1), (C2), $H_x(t)$ and $H_x^u(t)$ satisfy (C5), (C6), (C7) in $[0,T]$ with $T < T_{H_x}$ and φ_x satisfies (C8).

(a) If $nh_n^5 \to 0$ and $\dfrac{(\log n)^3}{nh_n} \to 0$, then, under (1), as $n \to +\infty$,

$$(nh_n)^{-1/2}(F_{xh}(\cdot) - F_x(\cdot)) \to W(\cdot|x) \qquad \text{in } l^\infty[0,T]$$

(b) If $h_n = Cn^{-1/5}$ for some $C > 0$, then, under (1), as $n \to +\infty$,

$$(nh_n)^{-1/2}(F_{xh}(\cdot) - F_x(\cdot)) \to \widetilde{W}(\cdot|x) \qquad \text{in } l^\infty[0,T]$$

where $W(\cdot|x)$ and $\widetilde{W}(\cdot|x)$ are Gaussian processes with covariance function given by

$$\begin{aligned}\Gamma_x(t,s) = \frac{||K||_2^2}{\varphi_x'(\bar{F}_x(t))\varphi_x'(\bar{F}_x(s))} \Bigg\{ & \int_0^{\min(t,s)} \varphi_x'(\bar{H}_x(z))^2 dH_x^u(z) \\ + & \int_0^{\min(t,s)} (\varphi_x''(\bar{H}_x(w))\bar{H}_x(w) + \varphi_x'(\bar{H}_x(w))) \int_0^w \varphi_x''(\bar{H}_x(y)) dH_x^u(y) dH_x^u(w) \\ + & \int_0^{\min(t,s)} \varphi_x''(\bar{H}_x(w)) \int_w^{\max(t,s)} (\varphi_x''(\bar{H}_x(y))\bar{H}_x(y) + \varphi_x'(\bar{H}_x(y))) dH_x^u(y) dH_x^u(w) \\ - & \int_0^t (\varphi_x''(\bar{H}_x(y))\bar{H}_x(y) + \varphi_x'(\bar{H}_x(y))) dH_x^u(y) \int_0^s (\varphi_x''(\bar{H}_x(w))\bar{H}_x(w) + \varphi_x'(\bar{H}_x(w))) dH_x^u(w) \Bigg\}\end{aligned} \tag{6}$$

and for $\widetilde{W}(\cdot|x)$, mean function given by

$$b_{tx} = \frac{-C^{5/2}\mu_2^K}{2\varphi_x'(\bar{F}_x(t))} \int_0^t [\varphi_x''(\bar{H}_x(s))\ddot{H}_x(s) dH_x^u(s) - \varphi_x'(\bar{H}_x(s)) d\ddot{H}_x^u(s)].$$

Remark 1. Note that when lifetime and censoring time are independent ($\varphi_x(t) = -\log(t)$), we obtain the well-known formulas for the asymptotic mean and variance of the Beran estimator as in Van Keilegom and Veraverbeke (1997a).

Remark 2. In order to keep the presentation simple, we have chosen for a fixed design setting. The theory for random design points $X_1, \ldots, X_n$ is similar but leads to somewhat more complicated expressions for the bias b_{tx} and the covariance $\Gamma_x(t,s)$, involving the density of the design points and its first derivative. The weights in (2) are then replaced by the simpler Nadaraya-Watson weights. Also an extension to multidimensional covariates is possible at the cost of more complexity.

5 Sensitivity analysis

In the previous section we have shown several results for the conditional copula-graphic estimator. We note that each result depends strongly on the underlying dependence structure between the survival time and the censoring time, which is described by a known Archimedean copula function. In a real data analysis it is very hard to know this copula function. The design of the experiment will in most cases give some hints about the association pattern between survival time and censoring time. In this section we explore the bias of the copula-graphic estimator due to a misspecified copula function. We will also consider the misspecification bias in a more general situation where the underlying dependence structure between survival time and censoring time is described by a general copula function instead of an Archimedean copula.

To determine the misspecification bias in the copula-graphic estimator, we assume that the true joint survival function is given by (1). If we use the generator ϕ_x in the calculations of the conditional copula-graphic estimator, we estimate the survival function $\bar{F}_x^*(t)$ defined as

$$\bar{F}_x^*(t) = \phi_x^{-1}\left(-\int_0^t \phi_x'(\bar{H}_x(s))dH_x^u(s)\right). \tag{7}$$

Using (1), we can rewrite (7) as

$$\begin{aligned}\bar{F}_x^*(t) &= \phi_x^{[-1]}\left(-\int_0^t \frac{\phi_x'(\bar{H}_x(s))}{\varphi_x'(\bar{H}_x(s))}\varphi_x'(\bar{F}_x(s))dF_x(s)\right) \\ &= \phi_x^{[-1]}\left(-\int_0^t \frac{\phi_x'\left(\varphi_x^{[-1]}(\varphi_x(\bar{F}_x(s))+\varphi_x(\bar{G}_x(s)))\right)}{\varphi_x'\left(\varphi_x^{[-1]}(\varphi_x(\bar{F}_x(s))+\varphi_x(\bar{G}_x(s)))\right)}\varphi_x'(\bar{F}_x(s))dF_x(s)\right).\end{aligned} \tag{8}$$

This survival function is different from the marginal survival function $\bar{F}_x(t)$ except when $\phi_x(t) = a\varphi_x(t)$ with a a constant (see Lemma 1). In a situation without censoring ($\bar{G}_x(t) = 1$) we note that $H_x(t) = H_x^u(t) = F_x(t) = F_x^*(t)$ and the bias due to misspecification is here also zero. In general we see that the misspecified survival function $\bar{F}_x^*(t)$ not only depends on the survival function $\bar{F}_x(t)$, but also on the survival function $\bar{G}_x(t)$ of the censoring time. This means that the percentage of censoring influences the misspecification bias. Along the same lines as in Proposition 2 of Rivest and Wells (2001), we can show that this bias is increasing with the percentage of censoring if $\phi_x'(u)/\varphi_x'(u)$ is monotone in u.

For the situation of a general copula function, we proceed as in Rivest and Wells (2001) and define a generator function for this copula. Let us assume that the true joint survival

function is given by

$$S_x(t_1, t_2) = \mathcal{C}_x(\bar{F}_x(t_1), \bar{G}_x(t_2)) \tag{9}$$

where $\mathcal{C}_x(u, v)$ is a general copula function. We calculate the crude hazard rates of the uncensored observations, both under model (9) and also under the Archimedean model (1). Equating these two hazard rates, we obtain

$$\varphi_x'(\bar{F}_x(t)) = \mathcal{C}_x^1(\bar{F}_x(t), \bar{G}_x(t))\varphi_x'(S_x(t,t))$$

where $\mathcal{C}_x^1(u, v) = \frac{\partial}{\partial u}\mathcal{C}_x(u, v)$ is the first partial derivative of the copula function $\mathcal{C}_x$.

Let us now define a generator $\varphi_{x\mathcal{C}}$ for a general copula $\mathcal{C}_x$ as a positive decreasing function satisfying $\varphi_{x\mathcal{C}}(1) = 0$ and for which the derivative is given by

$$\varphi_{x\mathcal{C}}'(s) = \mathcal{C}_x^1(s, K_x(s))\varphi_{x\mathcal{C}}'(\mathcal{C}_x(s, K_x(s))) \tag{10}$$

where $K_x(s) = \bar{G}_x(\bar{F}_x^{-1}(s))$.

We note that for a general copula function $\mathcal{C}_x$, it is difficult to solve equation (10). Often we cannot find a closed form solution $\varphi_{x\mathcal{C}}$ for this differential equation because we do not know the function $K_x(s)$. One of the situations where we can find a solution for $\varphi_{x\mathcal{C}}$ in (10), is when $\mathcal{C}_x$ is an Archimedean copula with generator φ_x. In this case any generator of the form $a.\varphi_x$, with a a constant, is a solution of this equation. Another situation is when $F_x(t) = G_x(t)$ and $\mathcal{C}_x$ is symmetric (i.e. $\mathcal{C}_x(u, v) = \mathcal{C}_x(v, u)$, for all u, v). Integrating both sides of the equation, we can rewrite (10) as $\varphi_{x\mathcal{C}}(s) = \varphi_{x\mathcal{C}}(\mathcal{C}_x(s, s))/2$ which is equivalent to $\mathcal{C}_x(s, s) = \varphi_{x\mathcal{C}}^{-1}(2\varphi_{x\mathcal{C}}(s))$. A solution $\varphi_{x\mathcal{C}}$ of this equation is the generator of an Archimedean copula with the same diagonal section as $\mathcal{C}_x$. The construction of such an Archimedean copula was discussed in Sungur and Yang (1996). The misspecified survival function $\bar{F}_x^*(t)$ of (7) can now be rewritten as $\bar{F}_x^*(t) = \phi_x^{-1}(\frac{1}{2}\phi_x(\mathcal{C}_x(\bar{F}_x(t), \bar{F}_x(t))))$ and hence the maximal bias due to misspecification is

$$\max_{t\in\mathbb{R}} |\bar{F}_x^*(t) - \bar{F}_x(t)| = \max_{s\in[0,1]} \left|\phi_x^{-1}\left(\frac{1}{2}\phi_x(\mathcal{C}_x(s, s))\right) - s\right|.$$

We note that this maximal bias only depends on $\mathcal{C}_x$ and ϕ_x, not on the marginals $F_x(t)$ and $G_x(t)$. In Table 5.1 we have calculated the maximal bias due to misspecification for several combinations of the underlying copula function $\mathcal{C}_x$ and the misspecified generator function ϕ_x. For the underlying copula function $\mathcal{C}_x$ we consider the independence copula (uv), the Gumbel-Morgenstern copula $(uv+u(1-u)v(1-v))$, the Fréchet-Hoeffding lower bound copula $(\max(u + v - 1, 0))$ and the Fréchet-Hoeffding upper bound $(\min(u, v))$. For the misspecified generator function ϕ_x we take $\phi_x = -\log(t)$ from the independence copula and $\phi_x = 1-t$ from the Fréchet-Hoeffding lower bound. We see that the maximal

$\mathcal{C}_x(u,v)$	$\phi_x(t)$ $-\log(t)$	$1-t$
Independence	0	0.5
Gumbel-Morgenstern	0.06744	0.5
Fréchet-Hoeffding lower bound	0.5	0
Fréchet-Hoeffding upper bound	0.25	0.5

Table 5.1 : The calculated maximal bias due to misspecification for several combinations of underlying copula function $\mathcal{C}_x$ and misspecification generator ϕ_x.

bias due to misspecification is zero when the misspecified generator ϕ_x is a generator of the underlying $\mathcal{C}_x$, as was expected. For the misspecified generator $\phi_x(t) = 1-t$, we note that the maximal bias due to misspecification is here always 0.5 when the underlying copula is not the Fréchet-Hoeffding lower bound copula.

Remark. A possible attempt to design a method for selecting an appropriate copula from a parametric family $\{\varphi_{\theta,x}\}$ of Archimedean copula generators has been suggested by a referee and is inspired by Andersen et al (2001). The idea is to use $\varphi_{\widehat{\theta},x}$ where $\widehat{\theta}$ is a maximizer of the likelihood function

$$L(\theta) = \prod_{\delta_i=1} \frac{\varphi'_{\theta,x}(\bar{F}_x(Z_i))F'_x(Z_i)}{\varphi'_{\theta,x}(\bar{H}_x(Z_i))} \prod_{\delta_i=0} \frac{\varphi'_{\theta,x}(\bar{G}_x(Z_i))G'_x(Z_i)}{\varphi'_{\theta,x}(\bar{H}_x(Z_i))}$$

after F_x, F'_x, G_x, G'_x and H_x have been replaced by estimators obtained under the assumption of independence of Y_x and C_x.

6 Example: survival of Atlantic halibut

In this section, we apply the copula-graphic estimator on a practical data set about survival of Atlantic halibut, studied by Neilson, Waiwood and Smith (1989). An important issue was the survival time of the fish after it was caught and handled as in the commercial fishery. For this purpose they had installed special holding tanks on the research vessel in which they placed the fish. Each fish was followed until it died. However some fish, mainly large fish, were removed after 48 hours to make space for other experimental animals. So the time until death was censored by the time that the animal had spent in the holding tank. Also the fish that were alive at the end of the experiment, were treated as censored observations. The researchers recorded several covariates among which we

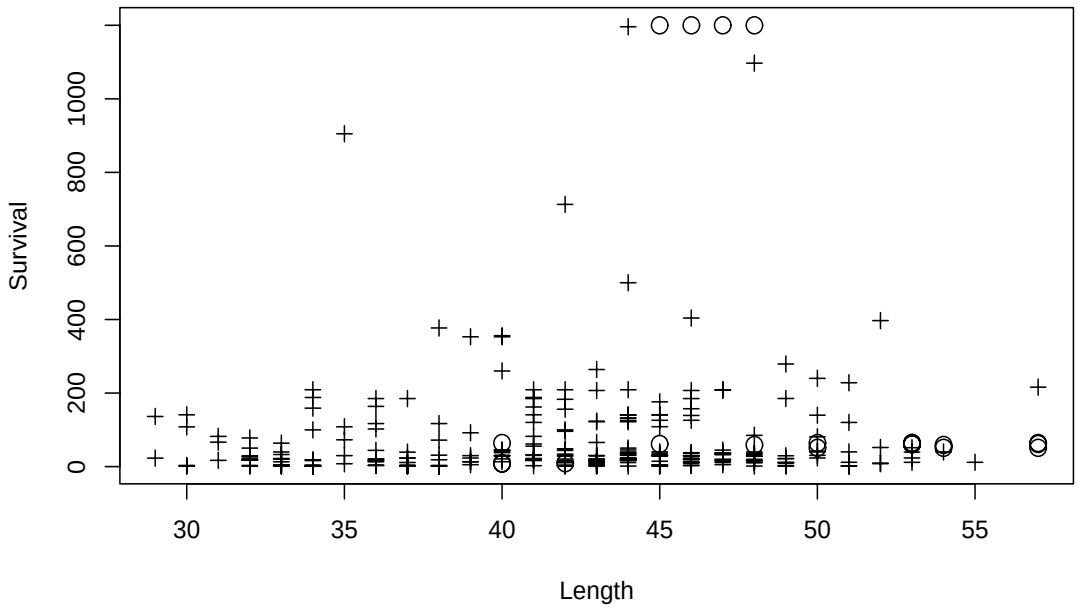

Figure 1 : Atlantic halibut data set: Survival times (in hours) versus fork length (in cm). Fish died in the holding tanks: +, fish removed from the holding tanks or alive at the end of the study: O.

focus on the fork length of the fish. In previous analyses of the data set, a significant effect of fork length on survival time had been found. In Figure 1 we show a scatter plot of the survival time versus the fork length of each animal, where we use + for uncensored observations and O for censored observations. The main causes of death for the fish were the stress of the new environment and also an infection caused by sick fishes in the tank. Therefore we believe that the survival time Y_x of a fish depends on the time that this fish has spent in the holding tank C_x, where the time in the holding tank has a negative influence on the survival time.

For these data, we construct the copula-graphic estimator for different choices of φ_x at fork lengths 32 cm and 53 cm, representing typical small fishes and typical large fishes. The four choices of the function φ_x that we will consider here, will lead each time to a different association for the dependence structure between the survival time and the time spent in the holding tank. This association can be measured in several ways. In this example, we take Kendall's τ which is defined in Nelsen (1999) as $\tau(x) = 1 + 4\int_0^1 \frac{\varphi_x(t)}{\varphi_x'(t)}dt$ and which has a range from -1 till 1. The association gets stronger when τ goes further away from zero.

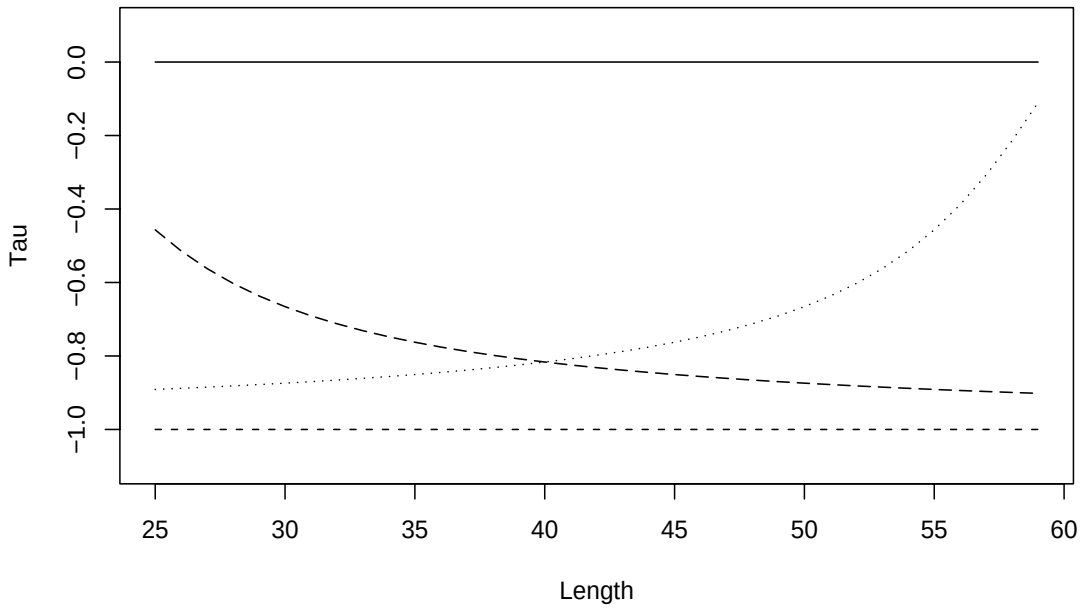

Figure 2 : Kendall's tau $\tau(x)$ for the different choices of generator φ_x. Independence (solid line), Fréchet - Hoeffding lower bound (dashed line), Frank family 1 (longdashed line) and Frank family 2 (dotted line).

The first choice is the independent copula ($\varphi_x(t) = -\log(t)$). This is the (possibly wrong) choice used in previous analyses of the data. In this case Kendall's tau $\tau(x)$ is equal to 0. The other choices of φ_x are such that they express that the time spent in the holding tank has a negative influence on the survival time. This is the notion of discordance, as defined formally in Nelsen (1999). The second choice of φ_x is the Fréchet-Hoeffding lower bound ($\varphi_x(t) = 1 - t$), which is the most extreme discordance that can be considered. Here $\tau(x) = -1$. For the next choices we take a generator function φ_x which really depends on the fork length x. Our third choice is the Frank family 1 copula given by

$$\varphi_x(t) = -\log\left(\frac{e^{(x-20)t} - 1}{e^{x-20} - 1}\right)$$

and the fourth choice is the Frank family 2 copula given by

$$\varphi_x(t) = -\log\left(\frac{e^{(60-x)t} - 1}{e^{60-x} - 1}\right).$$

From Figure 2, we see that for the Frank family 1 copula, Kendall's tau is a decreasing, negative function and hence gives a stronger discordant association for larger fishes than

for small fishes. For the fourth choice there is a stronger discordant association for small fishes.

In Figure 3, we show the copula-graphic estimates for the conditional survival function $\bar{F}_x(t)$ at fork lengths of 32 cm and 53 cm, and for bandwidths 20 and 40. The fork length is kept equal for plots in the same column and the bandwidth is equal for plots in the same row. In each of the four plots, we construct the copula-graphic estimator for the four choices of φ_x. We use in this data set the Gasser-Müller weights with a biquadratic kernel given by $K(z) = (15/16)(1 - z^2)^2 I(|z| \leq 1)$. As we saw in Figure 1, the covariate fork length of a fish is measured crudely on a scale of whole centimeters such that the observations form vertical lines on this plot. It is therefore possible to treat this covariate as fixed. It is also easy to see that our results for the copula-graphic estimator remain valid in the interval $[25, 60]$ for the covariate x, instead of the standard interval $[0, 1]$. The choices of the bandwidth are selected here for illustration purpose only. It is possible to set up a bandwidth selection criterium using, for example, the asymptotic mean squared error expression, but this would lead us into a field of research that we do not enter at this moment.

We note in Figure 3 that the estimates for the conditional survival function lie close together in each of the four plots and lie even almost on top of each other in the plots of the first column. This means that the choice of the generator function φ_x does not have a great influence on the survival time of small fishes. In the plots at the fork length of 53 cm we see that the copula-graphic estimates can be divided in two groups. The Fréchet-Hoeffding lower bound copula and the Frank family 1 copula give estimates that lie almost on top of each other but that are clearly different from the estimates of the independent copula and the Frank family 2 copula which form the second group. By this division in two groups, we see that this data set reacts differently to two different situations. The choices of φ_x in the first group have in common that they assume a large discordant association between survival time and time spent in the holding tank for larger fishes. In the second group, the choices of φ_x assume practically no discordant association for larger fishes. This influences the estimates for the survival function. With a φ_x from the first group, the estimated survival function for larger fish is higher than with a φ_x from the second group (in particular the φ_x that describes independence). This allows us to make some comments on the previous research that has been done on this data set. The researchers used at that time only the independent copula and ignored in this way that some of the fishes in their experiment did not die from the catch and handling but from stress caused by the living conditions in the holding tank. Therefore the estimate for the survival time that they found is an underestimate of the true survival time. By

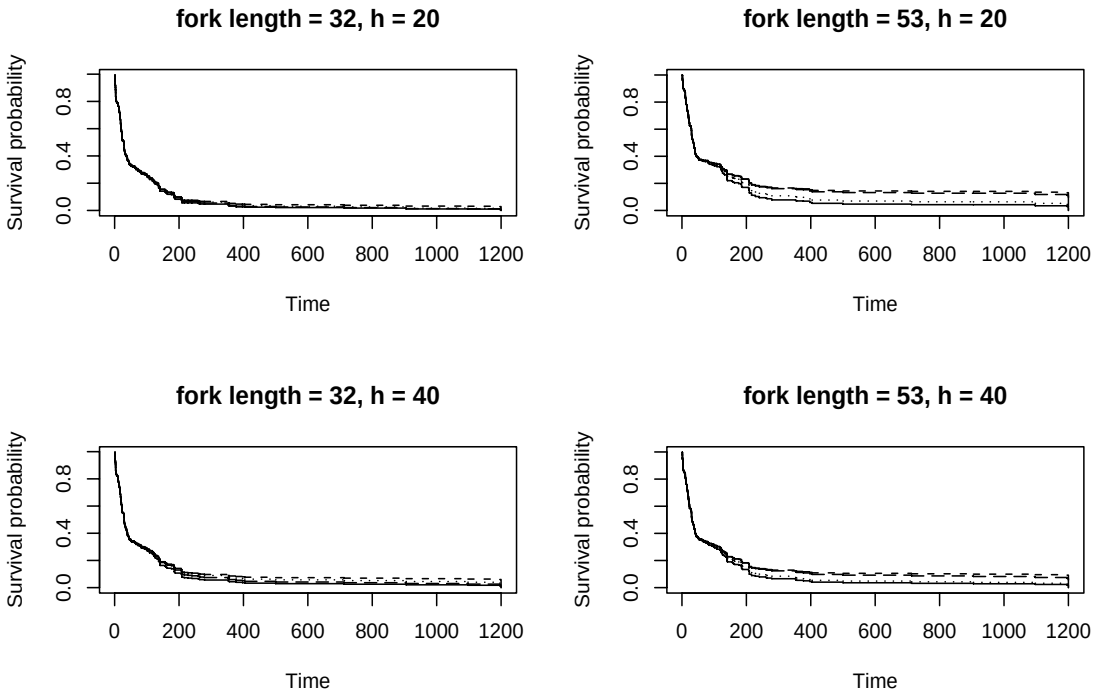

Figure 3 : Different copula-graphic estimates for the conditional survival function at lengths 32 cm and 53 cm and bandwidths 20 and 40. Independence (solid line), Fréchet - Hoeffding lower bound (dashed line), Frank family 1 (longdashed line) and Frank family 2 (dotted line).

the analysis in this example, we are able to show that for larger fishes their estimate is an underestimate and that the stress caused to a fish has to be taken into account in the estimate of the survival function. This is also positive news since fishes of this size are able to survive the catch and handling in commercial fishing better than the researchers originally expected. To finish this section, we note that the estimates for the survival function do not change much when we use a different bandwidth in the calculations.

Appendix

In this section we prove the asymptotic results of Section 4. We will make frequent use of several results of Van Keilegom and Veraverbeke (1997b), who dealt with the special case of independence ($\varphi_x(t) = -\log(t)$). Results based on observable quantities like $H_x(t)$ and $H_x^u(t)$ will be taken over from their work, since they do not depend on the underlying dependence structure of the model.

We start with the asymptotic representation for the conditional copula-graphic estimator.

Proof of Theorem 1. Based on Lemma 1, we can write for $t < T_{H_{xh}}$,

$$F_{xh}(t) - F_x(t) = \left[-\varphi_x^{-1} \left(- \sum_{Z_i \leq t, \delta_i = 1} \varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x, h_n)) \right) \right.$$
$$\left. + \ \varphi_x^{-1} \left(- \sum_{Z_i \leq t, \delta_i = 1} \varphi_x'(\bar{H}_{xh}(Z_i)) w_{ni}(x, h_n) \right) \right]$$
$$- \ \left[\varphi_x^{-1} \left(- \int_0^t \varphi_x'(\bar{H}_{xh}(s)) dH_{xh}^u(s) \right) - \varphi_x^{-1} \left(- \int_0^t \varphi_x'(\bar{H}_x(s)) dH_x^u(s) \right) \right].$$

Applying a first order Taylor expansion on the first term and a second order Taylor expansion on the second term, we get

$$\begin{aligned} F_{xh}(t) - F_x(t) &= \frac{-1}{\varphi_x'(\bar{F}_x(t))} \left[- \int_0^t \varphi_x'(\bar{H}_{xh}(s)) dH_{xh}^u(s) + \int_0^t \varphi_x'(\bar{H}_x(s)) dH_x^u(s) \right] \\ &\quad + R_{n1}(t) + R_{n2}(t) \end{aligned}$$

where

$$\begin{aligned} R_{n1}(t) &= \frac{\varphi_x''(\varphi_x^{-1}(\varepsilon_1))}{2\varphi_x'(\varphi_x^{-1}(\varepsilon_1))^3} \left[- \int_0^t \varphi_x'(\bar{H}_{xh}(s)) dH_{xh}^u(s) + \int_0^t \varphi_x'(\bar{H}_x(s)) dH_x^u(s) \right]^2 \\ R_{n2}(t) &= \frac{-1}{\varphi_x'(\varphi_x^{-1}(\varepsilon_2))} \left[- \sum_{Z_i \leq t, \delta_i = 1} (\varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x, h_n))) \right. \\ &\quad \left. + \sum_{Z_i \leq t, \delta_i = 1} \varphi_x'(\bar{H}_{xh}(Z_i)) w_{ni}(x, h_n) \right] \end{aligned}$$

with ε_1 between $-\int_0^t \varphi_x'(\bar{H}_{xh}(s)) dH_{xh}^u(s)$ and $-\int_0^t \varphi_x'(\bar{H}_x(s)) dH_x^u(s)$, and ε_2 between $-\sum_{Z_i \leq t, \delta_i = 1} (\varphi_x(\bar{H}_{xh}(Z_i^-)) - \varphi_x(\bar{H}_{xh}(Z_i^-) - w_{ni}(x, h_n)))$ and $-\sum_{Z_i \leq t, \delta_i = 1} \varphi_x'(\bar{H}_{xh}(Z_i)) w_{ni}(x, h_n)$.

Furthermore, for $t < T_{H_{xh}}$:

$$- \int_0^t \varphi_x'(\bar{H}_{xh}(s)) dH_{xh}^u(s) + \int_0^t \varphi_x'(\bar{H}_x(s)) dH_x^u(s) =$$
$$- \int_0^t (\varphi_x'(\bar{H}_{xh}(s)) - \varphi_x'(\bar{H}_x(s))) dH_x^u(s) - \int_0^t \varphi_x'(\bar{H}_x(s)) d(H_{xh}^u(s) - H_x^u(s))$$
$$- \int_0^t (\varphi_x'(\bar{H}_{xh}(s)) - \varphi_x'(\bar{H}_x(s))) d(H_{xh}^u(s) - H_x^u(s)).$$

On the integrand of the first term, we use a second order Taylor expansion and the second term can be rewritten by partial integration. So we get

$$
\begin{aligned}
&-\int_0^t \varphi'_x(\bar{H}_{xh}(s))dH^u_{xh}(s) + \int_0^t \varphi'_x(\bar{H}_x(s))dH^u_x(s) = \\
&\int_0^t \varphi''_x(\bar{H}_x(s))(H_{xh}(s) - H_x(s))dH^u_x(s) - \varphi'_x(\bar{H}_x(t))(H^u_{xh}(t) - H^u_x(t)) \\
&-\int_0^t \varphi''_x(\bar{H}_x(s))(H^u_{xh}(s) - H^u_x(s))dH_x(s) + R_{n3}(t) + R_{n4}(t) \qquad (11)
\end{aligned}
$$

where

$$
\begin{aligned}
R_{n3}(t) &= -\int_0^t \frac{\varphi'''_x(\varepsilon_3)}{2}(H_{xh}(s) - H_x(s))^2 dH^u_x(s) \\
R_{n4}(t) &= -\int_0^t (\varphi'_x(\bar{H}_{xh}(s)) - \varphi'_x(\bar{H}_x(s)))d(H^u_{xh}(s) - H^u_x(s))
\end{aligned}
$$

with ε_3 between $\bar{H}_{xh}(s)$ and $\bar{H}_x(s)$.

Since $H_x(T) < 1$ and $H_{xh}(T) \to H_x(T)$ a.s. (Lemma A.2. of Van Keilegom and Veraverbeke (1997b)), we may suppose that $T < T_{H_{xh}}$. For $R_{n3}(t)$ we have

$$
\begin{aligned}
\sup_{0\le t\le T}|R_{n3}(t)| &\le \frac{1}{2}\sup_{0\le t\le T}(H_{xh}(t) - H_x(t))^2 \max(\sup_{0\le t\le T}|\varphi'''_x(\bar{H}_{xh}(t))|, \sup_{0\le t\le T}|\varphi'''_x(\bar{H}_x(t))|) \\
&= O((nh_n)^{-1}\log n) \qquad \text{a.s.}
\end{aligned}
$$

by applying Lemma A.4. of Van Keilegom and Veraverbeke (1997b). By Lemma 2 below, we see that $\sup_{0\le t\le T}|R_{n4}(t)| = O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s.

From (11), Lemma A.4. of Van Keilegom and Veraverbeke (1997b) and the bounds on $R_{n3}(t)$ and $R_{n4}(t)$, we get

$$
\sup_{0\le t\le T}\left|-\int_0^t \varphi'_x(\bar{H}_{xh}(s))dH^u_{xh}(s) + \int_0^t \varphi'_x(\bar{H}_x(s))dH^u_x(s)\right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \text{ a.s.}
$$

This leads to $\sup_{0\le t\le T}|R_{n1}(t)| = O((nh_n)^{-1}\log n)$ a.s. Furthermore in Lemma 3 below, we show that $\sup_{0\le t\le T}|R_{n2}(t)| = O((nh_n)^{-1})$ a.s. which finishes the proof of this theorem.

We still have to prove the two lemmas used above.

Lemma 2. Under the conditions of Theorem 1, as $n \to +\infty$,

$$\sup_{0\le t\le T}\left|-\int_0^t(\varphi_x'(\bar{H}_{xh}(s))-\varphi_x'(\bar{H}_x(s)))d(H_{xh}^u(s)-H_x^u(s))\right| = O((nh_n)^{-3/4}(\log n)^{3/4}) \text{ a.s.}$$

Proof. Divide $[0,T]$ into $k_n = O((nh_n)^{1/2}(\log n)^{-1/2})$ subintervals $[t_i, t_{i+1}]$ of length $O((nh_n)^{-1/2}(\log n)^{1/2})$. We can find, as in the proof of Lemma 2 of Lo and Singh (1985), that

$$\begin{aligned}
&\sup_{0\le t\le T}\left|-\int_0^t(\varphi_x'(\bar{H}_{xh}(s))-\varphi_x'(\bar{H}_x(s)))d(H_{xh}^u(s)-H_x^u(s))\right| \\
&\le 2\max_{1\le i\le k_n}\sup_{t_i\le y\le t_{i+1}}|\varphi_x'(\bar{H}_{xh}(y))-\varphi_x'(\bar{H}_x(y))-\varphi_x'(\bar{H}_{xh}(t_i))+\varphi_x'(\bar{H}_x(t_i))| \\
&+k_n\sup_{0\le t\le T}|\varphi_x'(\bar{H}_{xh}(t))-\varphi_x'(\bar{H}_x(t))|\max_{1\le i\le k_n}|H_{xh}^u(t_{i+1})-H_x^u(t_{i+1})-H_{xh}^u(t_i)+H_x^u(t_i)| \\
&\le 2\max_{1\le i\le k_n}\sup_{t_i\le y\le t_{i+1}}\varphi_x''(\bar{H}_x(t_{i+1}))|H_{xh}^u(y)-H_x^u(y)-H_{xh}^u(t_i)+H_x^u(t_i)| \\
&+k_n\sup_{0\le t\le T}|\varphi_x'(\bar{H}_{xh}(t))-\varphi_x'(\bar{H}_x(t))|\max_{1\le i\le k_n}|H_{xh}^u(t_{i+1})-H_x^u(t_{i+1})-H_{xh}^u(t_i)+H_x^u(t_i)| \\
&+O((nh_n)^{-1}\log n).
\end{aligned}$$

In the last inequality we used a second order Taylor expansion and Lemma A.4. of Van Keilegom and Veraverbeke (1997b). To estimate the first term we further divide each $[t_i, t_{i+1}]$ into $a_n = O((nh_n)^{1/4}(\log n)^{-1/4})$ subintervals $[t_{ij}, t_{i,j+1}]$ of length $O((nh_n)^{-3/4}(\log n)^{3/4})$. By using Berstein's inequality, we can show that this term is bounded a.s. by $C\max_{1\le i\le k_n}\max_{0\le j\le a_n-1}|H_{xh}(t_{i,j+1})-H_x(t_{i,j+1})-H_{xh}(t_i)+H_x(t_i)|+$ $O((nh_n)^{-3/4}(\log n)^{3/4})$, for some constant $C>0$. Applying Lemma A.5 and Corollary A.1 of Van Keilegom and Veraverbeke (1997b) gives that this term is $O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. The second term is treated similarly and leads to the same order.

Lemma 3. Assume (C1), (C2), $H_x(t)$ satisfies (C3) in $[0,T]$ with $T < T_{H_x}$, $h_n \to 0$, $\dfrac{\log n}{nh_n} \to 0$, φ_x satisfies (C8). Then as $n \to +\infty$,

$$\begin{aligned}
&\sup_{0\le t\le T}\left|-\sum_{Z_i\le t,\delta_i=1}(\varphi_x(\bar{H}_{xh}(Z_i^-))-\varphi_x(\bar{H}_{xh}(Z_i^-)-w_{ni}(x,h_n))-\varphi_x'(\bar{H}_{xh}(Z_i))w_{ni}(x,h_n))\right| \\
&= O((nh_n)^{-1}) \text{ a.s.}
\end{aligned}$$

Proof. Because $H_x(T) < 1$ and $H_{xh}(T) \to H_x(T)$ a.s. (Lemma A.2. Van Keilegom and Veraverbeke (1997b)), we may suppose that $T < T_{H_{xh}}$. If $t < T$, then after applying a

second order Taylor expansion, we get

$$
\begin{aligned}
&-\sum_{Z_i\le t,\delta_i=1}(\varphi_x(\bar H_{xh}(Z_i^-))-\varphi_x(\bar H_{xh}(Z_i^-)-w_{ni}(x,h_n))-\varphi_x'(\bar H_{xh}(Z_i))w_{ni}(x,h_n))\\
&=-\frac{1}{2}\sum_{Z_i\le t,\delta_i=1}\varphi_x''(\varepsilon_i)w_{ni}^2(x,h_n)
\end{aligned}
$$

with ε_i between $\bar H_{xh}(Z_i)$ and $\bar H_{xh}(Z_i)+w_{ni}(x,h_n)$.
Hence

$$
\begin{aligned}
&\sup_{0\le t\le T}\left|-\sum_{Z_i\le t,\delta_i=1}(\varphi_x(\bar H_{xh}(Z_i^-))-\varphi_x(\bar H_{xh}(Z_i^-)-w_{ni}(x,h_n))-\varphi_x'(\bar H_{xh}(Z_i))w_{ni}(x,h_n))\right|\\
&\quad\le\frac{1}{2}\varphi_x''(\bar H(T))\sum_{i=1}^n w_{ni}^2(x,h_n)=O((nh_n)^{-1})\qquad\text{a.s.}
\end{aligned}
$$

Before we prove the weak convergence result, we give two lemmas about the asymptotic bias and variance of the conditional copula-graphic estimator.

Lemma 4. Assume (C1), (C2), $H_x(t)$ and $H_x^u(t)$ satisfy (C3) and (C5) in $[0,T]$ with $T<T_{H_x}$ and φ_x satisfies (C8), $h_n\to 0$. Then, as $n\to+\infty$

$$
\begin{aligned}
\sup_{0\le t\le T}\Bigg|\sum_{i=1}^n w_{ni}(x,h_n)Eg_{tx}(Z_i,\delta_i)+\frac{\mu_2^K h_n^2}{2\varphi_x'(\bar F_x(t))}\Bigg(&\int_0^t\varphi_x''(\bar H_x(s))\ddot H_x(s)dH_x^u(s)\\
&-\int_0^t\varphi_x'(\bar H_x(s))d\ddot H_x^u(s)\Bigg)\Bigg|=o(h_n^2)+O(n^{-1}).
\end{aligned}
$$

Proof. For fixed $t\le T$,

$$
\begin{aligned}
&\sum_{i=1}^n w_{ni}(x,h_n)Eg_{tx}(Z_i,\delta_i)=\frac{-1}{\varphi_x'(\bar F_x(t))}\times\\
&\left(\int_0^t\varphi_x''(\bar H_x(s))(EH_{xh}(s)-H_x(s))dH_x^u(s)-\int_0^t\varphi_x'(\bar H_x(s))d(EH_{xh}^u(s)-H_x^u(s))\right)
\end{aligned}
$$

By Lemma A.1.b of Van Keilegom and Veraverbeke (1997b), we get the result.

Lemma 5. Assume (C1), (C2), $H_x(t)$ and $H_x^u(t)$ satisfy (C3) in $[0,T]$ with $T<T_{H_x}$ and φ_x satisfies (C8), $h_n\to 0$, $nh_n\to+\infty$. Then, as $n\to+\infty$

$$
\sup_{0\le t\le T}\left|\sum_{i=1}^n w_{ni}^2(x,h_n)\mathrm{Cov}(g_{tx}(Z_i,\delta_i),g_{ts}(Z_i,\delta_i))-\frac{1}{nh_n}\Gamma_x(t,s)\right|=o((nh_n)^{-1})
$$

where $\Gamma_x(t,s)$ is given by (6).

Proof. Some straightforward calculations show that

$$\mathrm{Cov}(g_{tx}(Z_i,\delta_i), g_{ts}(Z_i,\delta_i)) = \frac{1}{\varphi'_x(\bar F_x(t))\varphi'_x(\bar F_x(s))}\left\{ \int_0^{\min(t,s)} \varphi'_x(\bar H_x(z))^2 dH^u_{x_i}(z) \right.$$

$$+ \int_0^{\min(t,s)} \int_0^{w} \varphi''_x(\bar H_x(y)) dH^u_x(y) [\varphi''_x(\bar H_x(w))\bar H_{x_i}(w) dH^u_x(w) + \varphi'_x(\bar H_x(w)) dH^u_{x_i}(w)]$$

$$+ \int_0^{\min(t,s)} \varphi''_x(\bar H_x(w)) \int_w^{\max(t,s)} [\varphi''_x(\bar H_x(y))\bar H_{x_i}(y) dH^u_x(y) + \varphi'_x(\bar H_x(y)) dH^u_{x_i}(y)] dH^u_x(w)$$

$$- \int_0^t [\varphi''_x(\bar H_x(y))\bar H_{x_i}(y) dH^u_x(y) + \varphi'_x(\bar H_x(y)) dH^u_{x_i}(y)] \times$$

$$\left. \int_0^s [\varphi''_x(\bar H_x(w))\bar H_{x_i}(w) dH^u_x(w) + \varphi'_x(\bar H_x(w)) dH^u_{x_i}(w)] \right\}$$

from which the result follows via standard calculations of asymptotic variances in a fixed design regression situation.

Proof of Theorem 2. From Theorem 1 and Lemma 4, we find

$$F_{xh}(t) - F_x(t) = \sum_{i=1}^n w_{ni}(x,h_n)\xi_{tx}(Z_i,\delta_i) + h_n^2 \bar b_{tx} + \bar R_n(t)$$

where $\xi_{tx}(Z_i,\delta_i) = g_{tx}(Z_i,\delta_i) - Eg_{tx}(Z_i,\delta_i)$, $\sup_{0\le t\le T}|\bar R_n(t)| = O((nh_n)^{-3/4}(\log n)^{3/4}) + o(h_n^2)$ a.s. and $\bar b_{tx} = \dfrac{-\mu_2^K}{2\varphi'_x(\bar F_x(t))} \displaystyle\int_0^t [\varphi''_x(\bar H_x(s))\ddot{\bar H}_x(s) dH^u_x(s) - \varphi'_x(\bar H_x(s)) d\ddot{\bar H}^u_x(s)]$. The bias $(nh_n)^{1/2}h_n^2\bar b_{tx}$ is $o(1)$ under conditions (a) and equals b_{tx} under conditions (b). Hence it suffices to prove the weak convergence of $W_{hx}(\cdot) = (nh_n)^{1/2}\sum_{i=1}^n w_{ni}(x,h_n)\xi_{\cdot x}(Z_i,\delta_i)$ to the Gaussian process $W(\cdot|x)$ with mean zero and covariance function $\Gamma_x(t,s)$.

This will be done in two steps. First we show the convergence of the finite dimensional distributions. Next we verify the asymptotic tightness by Theorem 2.11.9 (Bracketing central limit theorem) of van der Vaart and Wellner (1996).

Convergence of the finite dimensional distributions is that for any $q = 1, 2, \ldots$ and any $0 \le t_1 \le \ldots \le t_q \le T$: $(W_{hx}(t_1), W_{hx}(t_2), \ldots, W_{hx}(t_q)) \xrightarrow{D} N(0, \Gamma_x(t_i,t_j))$. Since $W_{hx}(t_i) = \sum_{k=1}^n W_{nki}$ where $W_{nki} = (nh_n)^{1/2} w_{nk}(x,h_n)\xi_{t_ix}(Z_k,\delta_k)$, it suffices to check that (see e.g. Araujo and Giné (1980)),

$$\lim_{n\to+\infty}\sum_{k=1}^n E(W_{nki}W_{nkj}) = \Gamma_x(t_i,t_j) \qquad (1\le i,j\le q)$$

$$\lim_{n\to+\infty}\sum_{k=1}^{n}\int_{\{|W_{nk}|>\varepsilon\}}|W_{nk}|^2dP=0$$

for every $\varepsilon>0$, where $|W_{nk}|^2=\sum_{i=1}^{q}W_{nki}^2$. Now, applying Lemma 5,

$$\sum_{k=1}^{n}E(W_{nki}W_{nkj})=(nh_n)\sum_{k=1}^{n}w_{nk}^2(x,h_n)\mathrm{Cov}(g_{t_ix}(Z_k,\delta_k),g_{t_jx}(Z_k,\delta_k))=\Gamma_x(t_i,t_j)+o(1).$$

Since the functions $\xi_{t_ix}(Z_k,\delta_k)$ are uniformly bounded, it follows that $\max_{1\le k\le n}|W_{nk}|=O((nh_n)^{-1/2})$ a.s. and $\sum_{k=1}^{n}|W_{nk}|^2=O(1)$ a.s., and hence,

$$\sum_{k=1}^{n}\int_{\{|W_{nk}|>\varepsilon\}}|W_{nk}|^2dP\le O(1)P(\max_{1\le k\le n}|W_{nk}|>\varepsilon)=o(1).$$

To prove the asymptotic tightness, we denote the process $W_{hx}(t)$ as $W_{hx}(t)=\sum_{i=1}^{n}Z_{ni}(t)$ where $Z_{ni}(t)=(nh_n)^{1/2}w_{ni}(x,h_n)\xi_{tx}(Z_i,\delta_i)$.

To verify the three conditions of Theorem 2.11.9 of van der Vaart and Wellner (1996), we put on $\mathcal{F}=[0,T]$, the semimetric

$$\begin{aligned}\rho(t,t') &= \max\left\{\left|\frac{-1}{\varphi_x'(\bar F_x(t))}+\frac{1}{\varphi_x'(\bar F_x(t'))}\right|,|\varphi_x'(\bar H_x(t))-\varphi_x'(\bar H_x(t'))|,\right.\\ &\qquad \left.|H_x(t)-H_x(t')|,\sup_{x'\in[0,1]}\sqrt{|H_{x'}^u(t)-H_{x'}^u(t')|}\right\}.\end{aligned}$$

In the third condition, we need the bracketing number $N_{[\ ]}(\varepsilon,\mathcal{F},L_2^n)$. This number is defined as the minimal number of sets in a partition of $\mathcal{F}=[0,T]=\bigcup_j\mathcal{F}_{\varepsilon j}$ such that for every set $\mathcal{F}_{\varepsilon j}$:

$$\sum_{i=1}^{n}E\left[\sup_{t,t'\in\mathcal{F}_{\varepsilon j}}|Z_{ni}(t)-Z_{ni}(t')|^2\right]\le\varepsilon^2.$$

Let us divide $\mathcal{F}=[0,T]$ into subintervals $0=t_0\le t_1\le\ldots\le t_q=T$ where $\rho(t,t')\le C\varepsilon$ for all $t,t'\in[t_{j-1},t_j],j=1,\ldots,q$ with C some constant which we will determine further on. For the partition $\mathcal{F}=[0,t_1]\bigcup\bigcup_{j=2}^{q}]t_{j-1},t_j]$, we find after some tedious calculations that

$$\begin{aligned}&|Z_{ni}(t)-Z_{ni}(t')|\le(nh_n)^{1/2}w_{ni}(x,h_n)\left(-\frac{\varphi_x''(\bar H_x(T))}{\varphi_x'(1)}|H_x^u(t)-H_x^u(t')|\right.\\ &+(\varphi_x''(\bar H_x(T))-2\varphi_x'(\bar H_x(T)))\left|\frac{-1}{\varphi_x'(\bar F_x(t))}+\frac{1}{\varphi_x'(\bar F_x(t'))}\right|+|\varphi_x'(\bar H_x(t))-\varphi_x'(\bar H_x(t'))|\\ &-\varphi_x'(\bar H_x(T))(|I(Z_i\le t,\delta_i=1)-I(Z_i\le t',\delta=1)|+|H_{x_i}^u(t)-H_{x_i}^u(t')|)\\ &\left.+\frac{\varphi_x'(\bar H_x(T))}{\varphi_x'(1)}|H_x(t)-H_x(t')|\right)\end{aligned}\tag{12}$$

So

$$\begin{aligned}\sup_{t,t'\in\mathcal{F}_{\varepsilon j}} |Z_{ni}(t) - Z_{ni}(t')|^2 &\leq (nh_n)w_{ni}^2(x,h_n)\{C_1(C\varepsilon)^2 \\ &+ C_2(C\varepsilon)|I(Z_i \leq t_j, \delta_i = 1) - I(Z_i \leq t_{j-1}, \delta = 1)| \\ &+ \varphi_x'(\bar{H}_x(T))^2|I(Z_i \leq t_j, \delta_i = 1) - I(Z_i \leq t_{j-1}, \delta = 1)|^2\}\end{aligned}$$

where C_1, C_2 are constants, uniquely determined by the right hand side of (12). For the appropriate choice of C, this leads to

$$\sum_{i=1}^{n} E\left[\sup_{t,t'\in\mathcal{F}_{\varepsilon j}} |Z_{ni}(t) - Z_{ni}(t')|^2\right] \leq \varepsilon^2.$$

Hence the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)$ is equal to $O(\varepsilon^{-1})$ and we get

$$\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)}d\varepsilon = \int_0^{\delta_n} \sqrt{\log O(\varepsilon^{-1})}d\varepsilon \to 0$$

when $\delta_n \to 0$.

We do not need to verify the second condition of Theorem 2.11.9 in van der Vaart and Wellner (1996), since our partition of $\mathcal{F} = [0, T]$ is independent of n. As last condition we have to check whether for all $\eta > 0$,

$$\sum_{i=1}^{n} E\left[\sup_{0\leq t\leq T} |Z_{ni}(t)| I\left(\sup_{0\leq t\leq T} |Z_{ni}(t)| > \eta\right)\right] \to 0 \text{ as } n \to +\infty.$$

Since $\xi_{tx}(Z_i, \delta_i)$ is bounded uniformly and $\max_{1\leq i\leq n} w_{ni}(x, h_n) = O((nh_n)^{-1})$ a.s., we get that $\sup_{0\leq t\leq T} |Z_{ni}(t)| = O((nh_n)^{-1/2})$ a.s., which is always smaller than η for n sufficiently large. So the first condition is also satisfied. By Theorem 2.11.9 of van der Vaart and Wellner (1996), we have that $W_{hx}(\cdot) \to W(\cdot|x)$ in $l^\infty[0, T]$.

Acknowledgement

This work was supported by the Ministry of the Flemish Community (Project BIL00/28, International Scientific and Technological Cooperation) and by the IAP research network nr P5/24 of the Belgian State (Belgian Science Policy).
The authors are grateful to the associate editor and two referees for their constructive remarks on the paper.

References

P. K. Andersen, C. Ekstrøm, J. P. Klein, Y. Shu and M. -J. Zhang (2001) Testing the goodness of fit of a copula based on right censored data, *Research report*, Department of Biostatistics, University of Copenhagen. (http://www.biostat.ku.dk/research-reports/2001/rr-01-09.ps)

A. Araujo and E. Giné, (1980) *The central limit theorem for real and Banach valued random variables*, Wiley, New York.

J. P. Klein and M. L. Moeschberger, (1997) *Survival analysis: Techniques for censored and truncated data*, Springer-Verlag, New York.

S.-H. Lo and K. Singh, (1986) The product-limit estimator and the bootstrap: Some asymptotic representations, *Probab. Theory Related Fields.* **71**, 455-465.

J. D. Neilson, K. G. Waiwood and S. J. Smith, (1989) Survival of Atlantic halibut (Hippoglossus hippoglossus) caught by longline and otter trawl gear, *Canadian Journal of Fisheries and Aquatic Sciences.* **46**, 887-897.

R. B. Nelsen, (1999) *An introduction to copulas*, Springer-Verlag, New York.

S. J. Smith, K. G. Waiwood and J. D. Neilson, (1994) Survival Analysis for size regulation of Atlantic Halibut, *Case studies in biometry*, ed. N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse, Wiley, New York, 125-144.

E. A. Sungur and Y. Yang, (1996) Diagonal copulas of archimedean class, *Commun. Statist.-Theory Meth.* **25**, 1659-1676.

A. Tsiatis, (1975) A nonidentifiability aspect of the problem of competing risks, *Proc. Nat. Acad. Sci. USA.* **72**, 20-22.

L. Rivest and M. T. Wells, (2001) A martingale approach to the copula-graphic estimator for the survival function under dependent censoring, *J. Multivariate Analysis.* **79**, 138-155.

A. W. van der Vaart and J. A. Wellner, (1996) *Weak convergence and empirical processes*, Springer-Verlag, New York.

I. Van Keilegom and N. Veraverbeke, (1996) Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles, *Commun. Statist.-Theory Meth.* **25**, 2251-2265.

I. Van Keilegom and N. Veraverbeke, (1997a) Weak convergence of the bootstrapped conditional Kaplan-Meier process ant its quantile process, *Commun. Statist.-Theory Meth.* **26**, 853-869.

I. Van Keilegom and N. Veraverbeke, (1997b) Estimation and bootstrap with censored data in fixed design nonparametric regression, *Ann. Inst. Statist. Math.* **49**, 467-491.

M. Zheng and J. P. Klein, (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika.* **82**, 127-138.