IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 245–253

DOI: 10.1214/10-IMSCOLL724

Nonparametric estimation of residual quantiles in a conditional Koziol–Green model with dependent censoring

Noël Veraverbeke

Hasselt University

Abstract: This paper discusses nonparametric estimation of quantiles of the residual lifetime distribution. The underlying model is a generalized Koziol–Green model for censored data, which accomodates both dependent censoring and covariate information.

1. Introduction

Consider a fixed design regression model where for each design point (covariate) $x \in [0,1]$ there is a nonnegative response variable Y_x, called lifetime or failure time. As in the case in many clinical or industrial trials, Y_x is subject to random right censoring by a nonnegative censoring variable C_x. The observed random variables at the design point x are

$$Z_x = \min(Y_x, C_x) \text{ and } \delta_x = I(Y_x \leq C_x).$$

Let us denote by F_x, G_x and H_x the distribution functions of Y_x, C_x and Z_x respectively. The main goal is to estimate the distribution function $F_x(t) = P(Y_x \leq t)$ (and functionals of it) from independent data $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ at fixed design points $0 \leq x_1 \leq \ldots \leq x_n \leq 1$. Here $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$. Note that at the design points x_i we write Y_i, C_i, Z_i, δ_i instead of $Y_{x_i}, C_{x_i}, Z_{x_i}, \delta_{x_i}$.

The classical assumption of independence between Y_x and C_x leads to the well known product-limit estimator of Beran [1]), which is the extension of the estimator of Kaplan and Meier [10] to the covariate case. However the assumption of independence between lifetime and censoring time is not always satisfied in practice and we should rather work with a more general assumption about the association between Y_x and C_x.

As in Zheng and Klein [18], Rivest and Wells [15] and Braekers and Veraverbeke [16] we will work with an Archimedean copula model for Y_x and C_x. See Nelsen [12] for information on copulas. It means that, for each $x \in [0,1]$ we assume

$$P(Y_x > t_1, C_x > t_2) = \varphi_x^{-1}(\varphi_x(\bar{F}_x(t_1)) + \varphi_x(\bar{G}_x(t_2))) \tag{1.1}$$

for all t_1, t_2, where φ_x is a known generator function depending on x in a general way, and $\bar{F}_x = 1 - F_x$, $\bar{G}_x = 1 - G_x$. We recall that for each x, $\varphi_x : [0,1] \to [0, +\infty]$

Hasselt University, Belgium. e-mail: noel.veraverbeke@uhasselt.be
AMS 2000 subject classifications: Primary 62N01; secondary 62N02, 62G08.
Keywords and phrases: Archimedean copula, asymptotic normality, dependent censoring, fixed design, Koziol–Green model, quantiles, residual lifetime.

is a continuous, convex, strictly decreasing function with $\varphi_x(1) = 0$.
In the random right censorship model there is an extensive literature on an important submodel initiated by Koziol and Green [11]. It is a submodel obtained by imposing an extra assumption on the distribution functions F_x and G_x. In this way it is a type of informative censoring. In the case of independence between Y_x and C_x, the Koziol–Green assumption is

$$\bar{G}_x(t) = (\bar{F}_x(t))^{\beta_x} \tag{1.2}$$

for all $t \geq 0$, where $\beta_x > 0$ is some constant depending in a general way on x. This extra assumption leads to an estimator for the survival function that is more efficient than the Kaplan-Meier estimator. See Cheng and Lin [4] in the case without covariates and Veraverbeke and Cadarso Suarez [17] in the regression case.
In order to generalize (1.2) to the dependent censoring case, we recall that for continuous F_x, (1.2) is equivalent to

$$Z_x \text{ and } \delta_x \text{ are independent.} \tag{1.3}$$

Translating property (1.3) into the model (1.1) gives that it is equivalent to the assumption

$$\varphi_x(\bar{G}_x(t)) = \beta_x \varphi_x(\bar{F}_x(t)) \tag{1.4}$$

for all $t \geq 0$ and for some $\beta_x > 0$.

Let us consider condition (1.4) for some examples of Archimedean copula models. For the independence case ($\varphi_x(t) = -\log t$), (1.4) coincides with (1.2). For the Gumbel copula ($\varphi_x(t) = (-\log t)^\alpha, \alpha \geq 1$), condition (1.4) becomes $\bar{G}_x(t) = (\bar{F}_x(t))^{\beta_x^{1/\alpha}}$. For the Clayton copula ($\varphi_x(t) = t^{-\alpha} - 1, \alpha > 0$), (1.4) becomes $\bar{G}_x(t) = (1 + \beta_x(F_x(t)^{-\alpha} - 1))^{-1/\alpha}$. This becomes (1.2) as $\alpha \to 0$.
In this paper we focus on nonparametric estimation of the median (or any other quantile) of the conditional residual lifetime in the above model. The conditional residual lifetime distribution is defined as $F_x(y \mid t) = P(Y_x - t \leq y \mid Y_x > t)$, i.e. the distribution of the residual lifetime, conditional on survival upon a given time t and at a given value of the covariate x. For any distribution function F, we denote by T_F the right endpoint of the support of F. Then, for $0 < y < T_{F_x}$, we have that

$$F_x(y \mid t) = \frac{F_x(t+y) - F_x(t)}{1 - F_x(t)}.$$

We define, for $0 < p < 1$, the p-th quantile of $F_x(y \mid t)$:

$$\begin{aligned} Q_x(t) = F_x^{-1}(p \mid t) &= \inf\{y \mid F_x(y \mid t) \geq p\} \\ &= -t + F_x^{-1}(p + (1-p)F_x(t)) \end{aligned} \tag{1.5}$$

where for any $0 < q < 1$ we write $F_x^{-1}(q) = \inf\{y \mid F_x(y) \geq q\}$ for the q-th quantile of F_x.
The paper is organized as follows. In Section 2 we discuss estimation of F_x and F_x^{-1}. We deal with residual quantiles in Sections 3 and 4. Some concluding remarks are in Section 5.

2. Estimation of the conditional distribution function and quantile function

Estimation of $Q_x(t)$ on the basis of observations (Z_i, δ_i), $i = 1, \ldots, n$, will be done by replacing F_x and F_x^{-1} in (1.5) by corresponding empirical versions F_{xh} and F_{xh}^{-1} where F_{xh} is the estimator studied in Braekers and Veraverbeke [2] and Gaddah and Braekers [8]. The derivation of this estimator goes as follows. From (1.1) we have that $\varphi_x(\bar{H}_x(t)) = \varphi_x(\bar{F}_x(t)) + \varphi_x(\bar{G}_x(t))$. Combining this with assumption (1.4) gives $\varphi_x(\bar{H}_x(t)) = (1+\beta_x)\varphi_x(\bar{G}_x(t))$, or with $\gamma_x = \frac{1}{1+\beta_x} = P(\delta_x = 1)$:

$$\bar{F}_x(t) = \varphi_x^{-1}(\gamma_x(\varphi_x(\bar{H}_x(t))). \tag{2.1}$$

In order to estimate $\bar{F}_x(t)$ at some fixed $x \in]0,1[$, we will use the idea that observations (Z_i, δ_i) with x_i close to x give the largest contribution to the estimator. Therefore we will smooth in the neighborhood of x by using Gasser-Müller type weights defined by

$$w_{ni}(x; h_n) = \frac{1}{c_n(x; h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) \mathrm{d}z \quad (i = 1, \ldots, n)$$

where $c_n(x; h_n) = \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) \mathrm{d}z$, $x_0 = 0$, K is a known probability density function and $h = \{h_n\}$ is a positive bandwidth sequence, tending to 0 as $n \to \infty$. The estimator $F_{xh}(t)$ of $F_x(t)$ is now obtained by replacing γ_x and $H_x(t)$ in (2.1) by the following empirical versions

$$\gamma_{xh} = \sum_{i=1}^n w_{ni}(x; h_n)\delta_i$$
$$H_{xh}(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i \le t).$$

Hence the estimator is given by

$$\bar{F}_{xh}(t) = \varphi_x^{-1}(\gamma_{xh}\varphi_x(\bar{H}_{xh}(t))). \tag{2.2}$$

To formulate some results on this estimator we need to introduce some further notations and some regularity conditions.

First some notations: 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 $\bar{\Delta}_n = \max_{1\le i\le n}(x_i - x_{i-1})$ and for the kernel K we write $\|K\|_2^2 = \int_{-\infty}^{\infty} K^2(u)\,\mathrm{d}u$, $\mu_1^K = \int_{-\infty}^{\infty} u\, K(u)\,\mathrm{d}u$, $\mu_2^K = \int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u$.

On the design and on the kernel, we will assume the following regularity conditions:

(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 is Lipschitz of order 1.

The results also require typical smoothness conditions on the elements of the model. For a fixed $0 < T < T_{F_x}$,

(C3) $\dot{F}_x(t) = \frac{\partial}{\partial x} F_x(t)$, $\ddot{F}_x(t) = \frac{\partial^2}{\partial x^2} F_x(t)$ exist and are continuous in $(x,t) \in [0,1] \times [0, T]$

(C4) $\dot{\beta}_x = \frac{\partial}{\partial x}\beta_x$, $\ddot{\beta}_x = \frac{\partial^2}{\partial x^2}\beta_x$ exist and are continuous in $x \in [0,1]$.

The generator φ_x of the Archimedean copula has to satisfy

(C5) $\varphi'_x(v) = \frac{\partial}{\partial v}\varphi_x(v)$, $\varphi''_x(v) = \frac{\partial^2}{\partial v^2}\varphi_x(v)$ are Lipschitz continuous in the x-direction, $\varphi'''_x(v) = \frac{\partial^3}{\partial v^3}\varphi_x(v) \le 0$ exists and is continuous in $(x, v) \in [0, 1]\times]0, 1]$.

Below we will use asymptotic representations for the estimator F_{xh} and the corresponding quantile estimator F_{xh}^{-1}. The representation for F_{xh} in Lemma 1 is taken from Theorem 2 in Braekers and Veraverbeke [3]. The representation for $F_{xh}^{-1}(p_n)$ in Lemma 2 is formulated for random p_n, tending to a fixed p as $n \to \infty$ at a certain rate. The proof of Lemma 2 is not given since it parallels that of a similar result in Gijbels and Veraverbeke ([9], Theorem 2.1).

Lemma 1. Assume (C1)–(C5) in $[0, T]$ with $T < T_{F_x}$, $h_n \to 0$, $\frac{\log n}{nh_n} \to 0$, $\frac{nh_n^5}{\log n} = O(1)$. Then, for $t < T_{F_x}$,

$$F_{xh}(t) = F_x(t) + \sum_{i=1}^{n} w_{ni}(x; h_n) g_x(Z_i, \delta_i, t) + r_n(x, t)$$

where

$$\begin{aligned} g_x(Z_i, \delta_i, t) &= \frac{-\varphi_x(\bar{H}_x(t))}{\varphi'_x(\bar{F}_x(t))}\{I(\delta_i = 1) - \delta_x\} \\ &+ \gamma_x \frac{\varphi'_x(\bar{H}_x(t))}{\varphi'_x(\bar{F}_x(t))}\{I(Z_i \le t) - H_x(t)\} \end{aligned}$$

and, as $n \to \infty$,

$$\sup_{0\le t\le T} |r_n(x, t)| = O((nh_n)^{-1} \log n) \text{ a.s.}$$

Lemma 2. Assume (C1)–(C5) in $[0, T]$ with $T < T_{F_x}$, $h_n \to 0$, $\frac{\log n}{nh_n} = o(1)$, $\frac{nh_n^5}{\log n} = O(1)$. Assume that $F_x^{-1}(p) < T$ and that $f_x(F_x^{-1}(p)) > 0$, where $f_x = F'_x$.

If $\{p_n\}$ is a sequence of random variables $(0 < p_n < 1)$ with $p_n - p = O_P((nh_n)^{-1/2})$, then as $n \to \infty$,

$$F_{xh}^{-1}(p_n) = F_x^{-1}(p) + \frac{1}{f_x(F_x^{-1}(p))}(p_n - F_{xh}(F_x^{-1}(p))) + o_P((nh_n)^{-1/2}).$$

3. Estimation of quantiles of the conditional residual lifetime

From (1.5) it follows that the obvious estimator for $Q_x(t)$ is given by

$$Q_{xh}(t) = -t + F_{xh}^{-1}(p + (1 - p)F_{xh}(t)) \tag{3.1}$$

where F_{xh} is the estimator in (2.2).
Denote $q_x = p + (1 - p)F_x(t)$ and $q_{xh} = p + (1 - p)F_{xh}(t)$.
We have the following asymptotic normality result.

Theorem 1. Assume (C1) − (C5) in $[0, T]$ with $T < T_{F_x}$. Assume that $F_x^{-1}(q_x) < T$ and that $f_x(F_x^{-1}(q_x)) > 0$.

(a) If $nh_n^5 \to 0$ and $(\log n)^2/(nh_n) \to 0$:

$$(nh_n)^{1/2}(Q_{xh}(t) - Q_x(t)) \to N(0; \sigma_x^2(t))$$

(b) If $h_n = Cn^{-1/5}$ for some $C > 0$:

$$(nh_n)^{1/2}(Q_{xh}(t) - Q_x(t)) \to N(\beta_x(t); \sigma_x^2(t)).$$

Here

$$\begin{aligned}
\sigma_x^2(t) &= \frac{\|K\|_2^2}{f_x^2(F_x^{-1}(q_x))} \left\{ \frac{1-\gamma_x}{\gamma_x} \left[(1-p)\frac{\varphi_x(\bar{H}_x(t))}{\varphi_x'(\bar{F}_x(t))} - \frac{\varphi_x(\bar{H}_x(F_x^{-1}(q_x))}{\varphi_x'(\bar{F}_x(F_x^{-1}(q_x))} \right]^2 \right. \\
&+ \gamma_x^2 \left[(1-p)^2 \frac{\varphi_x'^2(\bar{H}_x(t))}{\varphi_x'^2(\bar{F}_x(t))} H_x(t)(1-H_x(t)) \right. \\
&\quad + \frac{\varphi_x'^2(\bar{H}_x(F_x^{-1}(q_x))}{\varphi_x'^2(\bar{F}_x(F_x^{-1}(q_x))} H_x(F_x^{-1}(q_x))(1-H_x(F_x^{-1}(q_x))) \\
&\quad \left.\left. -2(1-p) \ \frac{\varphi_x'(\bar{H}_x(t))}{\varphi_x'(\bar{F}_x(t))} \frac{\varphi_x'(\bar{H}_x(F_x^{-1}(q_x))}{\varphi_x'(\bar{F}_x(F_x^{-1}(q_x))} H_x(t)(1-H_x(F_x^{-1}(q_x)) \right] \right\}
\end{aligned}$$

$\beta_x(t) = (1-p)b_x(t) + b_x(F_x^{-1}(q_x))$
with

$$b_x(t) = \frac{1}{2} C^{5/2} \mu_2^K \left\{ \frac{-\varphi_x(\bar{H}_x(t))}{\varphi_x'(\bar{F}_x(t))} \ddot{\gamma}_x + \frac{\gamma_x \varphi_x'(\bar{H}_x(t))}{\varphi_x'(\bar{F}_x(t))} \ddot{H}_x(t) \right\}. \tag{3.2}$$

Proof. Using Lemma 2 first and then Lemma 1, we have that

$$Q_{xh}(t) - Q_x(t) = \frac{1}{f_x(F_{xh}^{-1}(q_x))}(q_{xh} - F_{xh}(F_x^{-1}(q_x))) + o_P((nh_n)^{-1/2})$$

$$= \frac{1}{f_x(F_x^{-1}(q_x))}[q_{xh} - q_x - (F_{xh}(F_x^{-1}(q_x)) - F_x(F_x^{-1}(q_x)))] + o_P((nh_n)^{-1/2})$$

$$= \frac{1}{f_x(F_x^{-1}(q_x))} \sum_{i=1}^{n} w_{ni}(x; h_n)[(1-p)g_x(Z_i, \delta_i, t) - g_x(Z_i, \delta_i, F_x^{-1}(q_x))]$$

$$+ o_P((nh_n)^{-1/2}).$$

From this asymptotic representation it is now standard to derive the asymptotic normality results. It also uses the expressions for covariance and bias functions as in Gaddah and Braekers [8].

Note. In the case of independent censoring we have that $\varphi_x(t) = -\log t$ and the expression for the asymptotic variance simplifies to

$$\begin{aligned}
&\frac{\|K\|_2^2}{f_x^2(F_x^{-1}(q_x))} \left\{ \frac{1-\gamma_x}{\gamma_x}(1-p)^2 \ln^2(1-p) \bar{F}_x^2(t) \right. \\
&\left. + \gamma_x^2 (1-p)^2 \bar{F}_x^{2-\frac{1}{\gamma_x}}(t) \left[\frac{H_x(F_x^{-1}(q_x))}{(1-p)^{1/\gamma_x}} - H_x(t) \right] \right\}
\end{aligned}$$

If there are no covariates this leads to a (corrected) formula in Csörgő [6]. And if there is no censoring ($\gamma_x = 1$), we also recognize the formula of Csörgő and Csörgő [?]:

$$\frac{p(1-p)\bar{F}(t)}{f^2(p+(1-p)\bar{F}(t))}.$$

4. Estimation of quantiles of the duration of old age

In many situations it is necessary to replace the t in $Q_x(t)$ by some estimator $\hat{t}$. The variable t is then considered as an unknown parameter, usually the starting point of "old age". For example, t could be defined through the proportion of retired people in the population under study, that is $t = F_x^{-1}(p_0)$ for some known p_0. The unknown t could then be estimated by $\hat{t} = F_{xh}^{-1}(p_0)$.
Let $\hat{t}$ be some general estimator for t and consider the estimator (3.1) with t replaced by $\hat{t}$:

$$Q_{xh}(\hat{t}) = -\hat{t} + F_{xh}^{-1}(p + (1-p)F_{xh}(\hat{t})).$$

The next theorem gives an asymptotic representation for $Q_{xh}(\hat{t}) - Q_x(t)$. It requires a stronger form of condition (C3):

(C3') $\dot{F}_x(t)$, $\ddot{F}_x(t)$, $F"_x(t) = \frac{\partial^2}{\partial t^2}F_x(t)$, $\dot{F}'_x(t) = \frac{\partial^2}{\partial x \partial t}F_x(t)$ exist and are continuous in $(x,t) \in [0,1] \times [0,T]$.

Theorem 2. Assume (C1) (C2) (C3') (C4) (C5) in $[0,T]$ with $T < T_{F_x}$, $F_x^{-1}(q_x) < T$, $f_x(F_x^{-1}(q_x)) > 0$. Assume $h_n \to 0$, $(\log n)^2/(nh_n) \to 0$, $\frac{nh_n^5}{\log n} = O(1)$.

Also assume that $\hat{t} - t = O_P((nh_n)^{-1/2})$. Then, as $n \to \infty$,

$$Q_{xh}(\hat{t}) - Q_x(t) = (-1 + (1-p)\tfrac{f_x(t)}{f_x(F_x^{-1}(q_x))})(\hat{t} - t)$$

$$+\tfrac{1}{f_x(F_x^{-1}(q_x))}\sum_{i=1}^{n} w_{ni}(x;h_n)\{(1-p)g_x(Z_i,\delta_i,t) - g_x(Z_i,\delta_i,F_x^{-1}(q_x))\}$$

$$+o_P((nh_n)^{-1/2}).$$

Proof. Denote $\hat{q}_{xh} = p + (1-p)F_{xh}(\hat{t})$. Then $\hat{q}_{xh} - q_x = (1-p)(F_{xh}(\hat{t}) - F_x(t)]$ and $Q_{xh}(\hat{t}) - Q_x(t) = -(\hat{t} - t) + (F_{xh}^{-1}(\hat{q}_{xh}) - F_x^{-1}(q_x))$.
Now write

$$\text{(4.1)} \qquad \begin{aligned} F_{xh}(\hat{t}) - F_{xh}(t) &= \{[F_{xh}(\hat{t}) - F_{xh}(t)] - [F_x(\hat{t}) - F_x(t)]\} \\ &\quad + \{F_{xh}(t) - F_x(t)\} + \{F_x(\hat{t}) - F_x(t)\}. \end{aligned}$$

To the first term on the right hand side we can apply a modulus of continuity result analogous to the one in Veraverbeke [16]. The proof in the present situation goes along the same lines and therefore it is not given here. It requires condition (C3'). To the second term in the right hand side of (4.1) we apply our Lemma 1 and to the third term we apply a first order Taylor expansion. This gives that

$$\hat{q}_{xh} - q_x = (1-p)\{f_x(t)(\hat{t} - t) + \sum_{i=1}^{n} w_{ni}(x;h_n)g_x(Z_i,\delta_i,t)\} + o_P((nh_n)^{-1/2}).$$

This, together with Lemma 2, leads to the asymptotic representation for $Q_{xh}(\hat{t}) - Q_x(t)$.

Example. If $t = F_x^{-1}(p_0)$ and $\hat{t} = F_{xh}^{-1}(p_0)$ for some known p_0, we can apply Lemma 2 to $\hat{t} - t$ and from Theorem 2 we obtain that

$$\begin{aligned} Q_{xh}(\hat{t}) - Q_x(t) &= \sum_{i=1}^n w_{ni}(x; h_n) \left\{ \frac{g_x(Z_i, \delta_i, F_x^{-1}(p_0))}{f_x(F_x^{-1}(p_0))} - \frac{g_x(Z_i, \delta_i, F_x^{-1}(q_x))}{f_x(F_x^{-1}(q_x))} \right\} \\ &+ o_P((nh_n)^{-1/2}). \end{aligned}$$

Bias and variance of the main term can be calculated and we obtain by standard arguments the following result.

Corollary. Let $t = F_x^{-1}(p_0)$, $\hat{t} = F_{xh}^{-1}(p_0)$, $q = p + (1-p)p_0$. Assume (C1) (C2) (C3') (C4) (C5) in $[0, T]$ with $T < T_{F_x}$, $h_n \to 0$, $F_x^{-1}(q) < T$, $f_x(F_x^{-1}(q)) > 0$, $f_x(F_x^{-1}(p_0)) > 0$.

(a) If $nh_n^5 \to 0$ and $(\log n)^2/(nh_n) \to 0$:

$$(nh_n)^{1/2}(Q_{xh}(\hat{t}) - Q_x(t)) \xrightarrow{d} N(0; \tilde{\sigma}_x^2(t))$$

(b) If $h_n = Cn^{-1/5}$ for some $C > 0$:

$$(nh_n)^{1/2}(Q_{xh}(\hat{t}) - Q_x(t)) \xrightarrow{d} N(\tilde{\beta}_x(t); \tilde{\sigma}_x^2(t))$$

Here

$$\begin{aligned} \tilde{\sigma}_x^2(t) &= \|K\|_2^2 \left\{ \tfrac{1-\gamma_x}{\gamma_x^2} \left[\tfrac{(1-p)\ln(1-p_0)}{f_x(F_x^{-1}(p_0))} - \tfrac{1-p)(1-p_0)\ln((1-p)(1-p_0))}{f_x(F_x^{-1}(q))} \right]^2 \right. \\ &+ \gamma_x^2 (1-p_0)^{2-\frac{1}{\gamma_x}} \left[\tfrac{H_x(F_x^{-1}(p_0))}{f_x^2(F_x^{-1}(p_0))} + \tfrac{(1-p)^{2-\frac{1}{\gamma_x}} H_x(F_x^{-1}(q))}{f_x^2(F_x^{-1}(q))} \right. \\ &\left. \left. - \tfrac{2(1-p)H_x(F_x^{-1}(p_0))}{f_x(F_x^{-1}(p_0)) f_x(F_x^{-1}(q))} \right] \right\} \\ \tilde{\beta}_x(t) &= \tfrac{b_x(F_x^{-1}(p_0))}{f_x(F_x^{-1}(p_0))} - \tfrac{b_x(F_x^{-1}(q))}{f_x(F_x^{-1}(q))}, \text{ with } b_x(t) \text{ as in (3.2).} \end{aligned}$$

5. Some concluding remarks

We developed asymptotic theory for nonparametric estimation of residual quantiles of the lifetime distribution in the Koziol–Green model of right random censorship. The possible dependence between responses and censoring times is modeled by a copula. There are several remarks in order before this can be applied to real data examples.

(1) The model assumes that the Archimedean copula is known and also that the generator depends on the covariate. We remark that, due to the censoring, it is not possible to estimate the generator φ_x using only the data (Z_i, δ_i), $i = 1, \dots, n$. As can be seen in Braekers and Veraverbeke [2], [3] and Gaddah and Braekers [8], a good suggestion is to choose a reasonable φ_x by looking at the graph of a dependence measure for Y_x and C_x. One could for example take Kendall's tau $(\tau(x))$, which is related to the generator via the simple formula $\tau(x) = 1 + 4\int(\varphi_x(t)/\varphi'_x(t))\,\mathrm{d}t$.

(2) The expressions for asymptotic bias and variance are explicit but require a lot of further estimation of unknown quantities. In order to avoid this, we suggest the following bootstrap procedure. For $i = 1, \ldots, n$ obtain Z_i^* from $H_{x_i g}(t)$ and independently, δ_i^* from a Bernoulli distribution with parameter $\gamma_{x_i g}$, where $H_{x_i g}(t)$ and $\gamma_{x_i g}$ are defined as in Section 2, but with a bandwidth $g = \{g_n\}$ that is typically asymptotically larger than $h = \{h_n\}$, i.e. $g_n/h_n \to \infty$ as $n \to \infty$. Next calculate $\gamma_{xhg}^* = \sum_{i=1}^n w_{ni}(x; h_n)\delta_i^*$ and $H_{xhg}^*(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i^* \leq t)$ and use $\overline{F}_{xhg}^*(t) = \varphi_x^{-1}(\gamma_{xhg}^* \varphi_x(\overline{H}_{xhg}^*(t))$ as a bootstrap version of $\overline{F}_{xh}(t)$.

(3) Also the choice of the bandwidth is an important practical issue. For this, we propose to use the above bootstrap scheme and to minimize asymptotic mean squared error expression over a large number of bootstrap samples.

(4) Alternative approaches to the copula model could be explored. For example one could assume conditional independence of Y and C, given that the (random) covariate X equals x. Residual quantiles could be defined and studied starting from Neocleous and Portnoy [13] and El Ghouch and Van Keilegom [5]. These authors developed non- and semiparametric estimators based on the nonparametric censored regression quantiles of Portnoy [14].

Acknowledgements

This research was supported by the IAP Research Network P6/03 of the Belgian Science Policy and by the Research Grant MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovación.

References

[1] Beran, R. (1981). *Nonparametric regression with randomly censored survival data.* Technical Report, Univ. California, Berkeley. MR

[2] Braekers, R. and Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *Canad. J. Statist.* **33** 429–447.

[3] Braekers, R. and Veraverbeke, N. (2008). A conditional Koziol–Green model under dependent censoring. *Statist. Probab. Letters* **78** 927–937.

[4] Cheng, P. E. and Lin, G. D. (1987). Maximum likelihood estimation of a survival function under the Koziol–Green proportional hazards model. *Statist. Probab. Letters.* **5** 75–80.

[5] El Gouch, A. and Van Keilegom, I. (2009). Local linear quantile regression with dependent censored data. *Statistica Sinica* **19** 1621–1640.

[6] Csörgő, S. (1987). Estimating percentile residual life under random censorship. In: *Contributions to Stochastics* (W. Sandler, ed.) 19–27. Physica-Verlag, Heidelberg.

[7] Csörgő, M. and Csörgő, S. (1987). Estimation of percentile residual life. *Oper. Res.* **35** 598–606.

[8] Gaddah, A. and Braekers, R. (2009). Weak convergence for the conditional distribution function in a Koziol–Green model under dependent censoring. *J. Statist. Planning Inf.* **139** 930–943.

[9] Gijbels, I. and Veraverbeke, N. (1988). Weak asymptotic representations for quantiles of the product-limit estimator. *J. Statist. Planning Inf.* **18** 151–160.

[10] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

[11] Koziol, J. A. and Green, S. B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63** 465–474.

[12] Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer-Verlag, New York.

[13] Neocleous, T. and Portnoy, S. (2009). Partially linear censored quantile regression. *Lifetime Data Analysis* **15** 357–378.

[14] Portnoy, S. (2003). Censored regression quantiles *J. Amer. Statist. Assoc.* **98** 1001–1012.

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

[16] Veraverbeke, N. (2006). Regression quantiles under dependent censoring. *Statistics* **40** 117–128.

[17] Veraverbeke, N. and Cadarso Suárez, C. (2000). Estimation of the conditional distribution in a conditional Koziol–Green model. *Test* **9** 97–122.

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