Bivariate censored regression relying on a new estimator of the joint distribution function

Olivier Lopez, Philippe Saint-Pierre

▶ **To cite this version:**

Olivier Lopez, Philippe Saint-Pierre. Bivariate censored regression relying on a new estimator of the joint distribution function. 2011. <hal-00584379>

HAL Id: hal-00584379
https://hal.archives-ouvertes.fr/hal-00584379

Submitted on 8 Apr 2011

Bivariate censored regression relying on a new estimator of the joint distribution function

Olivier LOPEZ[1]   Philippe SAINT-PIERRE[2]

April 8, 2011

Abstract

In this paper we study a class of $M-$estimators in a regression model under bivariate random censoring and provide a set of sufficient conditions that ensure asymptotic $n^{1/2}-$convergence. The cornerstone of our approach is a new estimator of the joint distribution function of the censored lifetimes. A copula approach is used to modelize the dependence structure between the bivariate censoring times. The resulting estimators present the advantage of being easily computable. A simulation study enlighten the finite sample behaviour of this technique.

Key words: Bivariate censoring, $M-$estimation, regression modeling, copula functions, Kaplan-Meier estimator, i.i.d. representations.

Short title: Bivariate censored regression

[1] Laboratoire de Statistique Théorique et Appliquée, Université Paris VI, 175 rue du Chevaleret, 75013 Paris, France, E-mail: olivier.lopez0@upmc.fr.

[2] Laboratoire de Statistique Théorique et Appliquée, Université Paris VI, 175 rue du Chevaleret, 75013 Paris, France, E-mail: philippe.saint_pierre@upmc.fr.

1 Introduction

M-estimators arise in many situations in classical regression analysis, like for instance least square estimator for mean-regression or (pseudo-) maximum likelihood estimators. In this paper, we focus on extending the definition of a general class of M-estimators to the framework of a regression model where response and covariate both are randomly right-censored. The aim is to adapt procedures used in the uncensored case so that they stay consistent in this new setting, and to derive their asymptotic normality under mild conditions. Until now, this question has been essentially considered in the particular case where only the response is censored, where two main approaches have been considered. The synthetic data approach (see Koul et al., 1981; Delecroix et al., 2008) consists of defining new variables estimated from the data. The second approach consists of modifying classical procedures by introducing some weights on the observations to compensate censoring. These weights are related to a nonparametric estimator of the distribution of the variables involved in the regression model (see e.g. Stute, 1993). In the present paper, we extend this technique to the case where censoring does not only strike the response, but also the covariate. Therefore, defining the weights that are adapted to our problem relies on a new estimator of the joint distribution of bivariate censored lifetimes.

The estimator of the bivariate survival distribution that we introduce is different from the ones traditionally used. The main reason is that our estimator is mainly designed for a regression analysis purpose. The difference between this present estimator and the existing ones stands in the fact that we do not only focus on the estimation of distribution function of the failure times. Indeed, in view of the M-estimators that we consider, we are interested in the more general problem of estimating quantities such as $E[\phi_\theta(Y^{(1)}, Y^{(2)})]$, where $(Y^{(1)}, Y^{(2)})$ denotes the censored failure times, and ϕ_θ is some function, with θ belonging to a parametric set. Therefore we focus on the estimation of the distribution of $(Y^{(1)}, Y^{(2)})$ and not only on the distribution function. Indeed, estimating $F(y^{(1)}, y^{(2)}) = \mathbb{P}(Y^{(1)} \le y^{(1)}, Y^{(2)} \le y^{(2)})$ only corresponds to a special case, taking $\phi_{y^{(1)}, y^{(2)}}(Y^{(1)}, Y^{(2)}) = \mathbf{1}_{Y^{(1)} \le y^{(1)}, Y^{(2)} \le y^{(2)}}$. Moreover, previously existing estimators of F do not correspond to a positive distribution, making them inoperative for our purpose of estimating expectations of a family of functions ϕ_θ. By construction, our new estimator is not affected by this drawback.

Aiming to estimate only the function F, Campbell and Földes (1982) provided an estimator which is usually not a proper survival function due to nonmonotonicity. Moreover the estimator is not equivariant under reversal of coordinates because it depends on

a choice of the decomposition of the bivariate survival probability. Hanley and Parnes (1983) studied a nonparametric maximum likelihood estimation (NPMLE) using Efron's self-consistency algorithm and the EM algorithm. However, this estimator may be inconsistent for continuous data (Tsai et al., 1986) and is not unique for samples taken from absolutely continuous distributions. Tsai et al. (1986) suggested an estimation method based on conditional survival function using Beran's estimator (Beran, 1981). The estimator is consistent but its almost sure consistency is slow. Moreover, this estimator is not automatically symmetric in the sense that is not equivariant under reversal of coordinates. On the other hand, the multivariate product limit estimator proposed by Dabrowska (1988) is based on a rewriting of the multivariate distribution function which does not introduce dissymetry between the two coordinates. Nevertheless, the resulting estimator of Dabrowska assigns negative mass to some points in the plane (Pruitt, 1991b). See also Prentice and Cai (1992) for a related estimator based on Peano series, and Gill et al. (1993) for some improvements of this results by application of functional delta method. Another approach consists of the use of nonparametric smoothing techniques to get a bivariate survival curve estimator (Pruitt, 1991a). A product limit estimator is used to impute values for singly censored observations and then the generalized maximum likelihood ideas are applied to complete the estimator. The main problem arising from the implicit definition of the estimator. Moreover, its practical performance has been shown to be less satisfactory than expected (van der Laan, 1996). Another approach based on NPMLE is proposed by van der Laan (1996). However, the NPMLE does not directly work in the particular framework of bivariate censoring, because of difficulties to assign the masses of the censoring in an appropriate way. Therefore, the technique of van der Laan (1996) requires to slightly modify the data, introducing an interval censoring methodology. The estimator is asymptotically efficient for these modified data. However, the convergence rate depends on some parameter involved in the modification of the data, and does not achieve the expected $n^{-1/2}$ rate. The same problem of low convergence rate appears in Akritas and Van Keilegom (2003). In opposition to the other existing techniques, the estimators proposed by Akritas and Van Keilegom (2003) present the advantage of being regular distribution functions. Nevertheless, they require the censoring to be an absolutely continuous random vector, and their extension to multivariate censoring seems difficult.

Compared to existing approaches, our new estimator of the distribution function, designed in view of applying it to a regression framework defines a true probability distribu-

tion, achieves $n^{1/2}-$consistency, and can be easily extended to a multivariate framework. Based on this new tool, we provide a general uniform central limit theorem, which can be seen as two-dimensional extensions of the results of Stute (1996) and Akritas and Van Keilegom (2001) in the univariate censoring case. As a by-product of this Theorem, we obtain the uniform consistency of an estimator of the multivariate distribution function on the whole space, under suitable conditions. We point out that, for the other existing approaches, uniform consistency has only been proved until now on a compact set strictly included in the support of the distribution of $(Y^{(1)}, Y^{(2)})$. Our procedure works under a reasonable identifiability assumption which relies on a copula modelization of the dependence between the multivariate censoring times.

The rest of the paper is organized as follows. The section 2 introduce our estimator of the bivariate distribution. In section 3, we provide a general result on asymptotic i.i.d. representations related to this new estimator. Our Theorem 3.3 is the key result of this section, since it permits to obtain an extension to bivariate censoring of an uniform version of the Central Limit Theorem. Theorem 3.3 is then used in section 4 to derive general results on $M-$estimators in a regression framework. These results are then applied to least-square estimation and generalized linear modeling. Section 5 is devoted to some simulation study of the finite sample size behaviour of the procedure. Technical arguments are postponed to section 7.

2 Estimation

2.1 Observations and formulation of the problem

Let us first introduce some general notations and define the bivariate censoring model that we will consider throughout this paper. In our framework, we are interested in infering on the unknown distribution μ of a random vector $T = (T^{(1)}, T^{(2)})$, whose two components are randomly right-censored. This means that, instead of observing i.i.d. copies of the random vector T, observations are made of i.i.d. copies of $Y = (Y^{(1)}, Y^{(2)}) = (\inf(T^{(1)}, C^{(1)}), \inf(T^{(2)}, C^{(2)}))$, and $\Delta = (\Delta^{(1)}, \Delta^{(2)}) = (\mathbf{1}_{T^{(1)} \leq C^{(1)}}, \mathbf{1}_{T^{(2)} \leq C^{(2)}})$, where $C = (C^{(1)}, C^{(2)})$ is the censoring random vector. To avoid dissymmetry problems caused by the presence of ties, we will assume throughout this paper that

$$\mathbb{P}(T^{(i)} = C^{(i)}) = 0, \text{ for } i = 1, 2. \tag{2.1}$$

4

One of the technical advantages of this assumption stands in the fact that it will avoid us to deal with denominators too close to zero (in the tail of the considered distributions) which appear naturally in the statistical study of censored lifetimes.

An M-estimator is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} M_n(\theta), \tag{2.2}$$

where $\Theta \subset \mathbb{R}^k$. In uncensored regression analysis, the traditional least square estimator used for mean-regression is obtained for $M_n(\theta) = \sum_{j=1}^n (Y_j^{(1)} - r(\theta, Y_j^{(2)}))^2$, where r is a known function. More generally, estimators in the GLM model (see e.g. McCullagh and Nelder, 1983) can also be expressed as $M-$estimators. In the following, we will consider M-estimators defined by a functional M_n of the following type,

$$M_n(\theta) = \int \phi_\theta(y^{(1)}, y^{(2)}) d\hat{\mu}(y^{(1)}, y^{(2)}) = \int \phi_\theta(y) d\hat{\mu}(y),$$

where $\hat{\mu}$ will denote a new estimator of the distribution of T under bivariate random censoring. Indeed, in absence of censoring, a classical way to proceed is to use the empirical distribution as a measure $\hat{\mu}$. In our framework, this estimator is unavailable because the lifetimes are not fully observed. Therefore, it is crucial to define a proper $\hat{\mu}$. In the following, the target of $M_n(\theta)$ will be denoted as $M(\theta)$, that is

$$M(\theta) = \int \phi_\theta(y^{(1)}, y^{(2)}) d\mu(y^{(1)}, y^{(2)}).$$

The convergence of estimators of the type (2.2) will be obtained by studying the convergence of M_n towards M. Therefore, in view of the definition of M_n, it becomes crucial to derive asymptotic properties of intergrals with respect to $\hat{\mu}$.

2.2 Estimating the distribution μ

2.2.1 General method to define convergent estimators of μ

In absence of censoring, a simple way to estimate the distribution of the data is to consider the empirical distribution, which simply puts the weight n^{-1} at each point $T_j \in \mathbb{R}^2$. Under our framework, the issue is to define an appropriate weight which compensates the presence of censoring. Under univariate random censoring, Kaplan-Meier approach consists of defining a probability measure which puts mass only at the uncensored observations, and allows more weight to large observations in order to compensate the lack of

observations in the tail of the distribution (Satten and Datta, 2001). Based on the same idea, the estimator we propose is of the following form,

$$\hat{\mu} = \sum_{j=1}^{n} \Delta_i^{(1)} \Delta_i^{(2)} W_j(Y_1, \Delta_1, ..., Y_n, \Delta_n) \delta_{\{Y_j^{(1)}, Y_j^{(2)}\}}, \tag{2.3}$$

where $\delta_{\{a,b\}}$ denotes the Dirac mass at point (a, b). This means that we only put mass at the observations where the two components of Y are uncensored, while the weight W_j is here to take account for bivariate censoring. Defining the weight function W_j depends on the identifiability conditions we put on the model. More precisely, we consider weights of the following form,

$$W_j(Y_1, \Delta_1, ..., Y_n, \Delta_n) = \frac{1}{n} \hat{f}(Y_j),$$

where $\hat{f}$ is some consistent estimator (computed from the whole data set) of some appropriate function f that satisfies the following equation,

$$E\left[\Delta^{(1)} \Delta^{(2)} f(Y) \phi(Y^{(1)}, Y^{(2)})\right] = E[\phi(T^{(1)}, T^{(2)})], \tag{2.4}$$

for any measurable function ϕ with finite expectation. A simple calculation shows that $f(y) = E[\Delta^{(1)} \Delta^{(2)} | T = y]^{-1}$. This function can not be computed in practice, since it depends on the law of the unobserved variables T. However it can be estimated in various situations, provided that one adds some identifiability assumptions that specify the dependence structure between T and C.

2.2.2 Application under a simple identifiability assumption

In this section, we develop the approach of section 2.2.1 in a specific framework. A simple and classical assumption used for bivariate censoring inference consists of considering that T and C are independent (see e.g. Dabrowska, 1988). It can be compared to the case of the univariate censoring, where this assumption of independence is needed to obtain convergence of the Kaplan-Meier estimator (Kaplan and Meier, 1958; Stute and Wang, 1993). Under this assumption, the function f introduced in the previous subsection is

$$f(y) = \mathbb{P}(y^{(1)} \le C^{(1)}, y^{(2)} \le C^{(2)})^{-1}.$$

To rewrite this function in a more simple way, we add some assumption on the dependence structure between $C^{(1)}$ and $C^{(2)}$. Define, for $i = 1, 2$, $G_i(x) = \mathbb{P}(x \le C^{(i)})$, $F_i(x) = \mathbb{P}(x \le T^{(i)})$, and $H_i(x) = \mathbb{P}(x \le Y^{(i)})$. We now set the identifiability assumption that we will use throughout this paper.

Identifiability Assumption 1 *Assume that T and C are independent, and that there exists some known copula function $\mathcal{C}$ such that*

$$\mathbb{P}(y^{(1)} \leq C^{(1)}, y^{(2)} \leq C^{(2)}) = \mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)})).$$

As a simple practical example, we have the particular case where $\mathcal{C}(x_1, x_2) = x_1 x_2$, which corresponds to the case where $C^{(1)}$ and $C^{(2)}$ are independent.

Identifiability Assumption 1 has been initially introduced by Wang and Wells (1997). Compared to our approach, Wang and Wells noticed that, under this simple assumption, $\mathbb{P}(T^{(1)} \geq y^{(1)}, T^{(2)} \geq y^{(2)}) = H(y^{(1)}, y^{(2)}) \mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)}))^{-1}$, where $H(y^{(1)}, y^{(2)}) = \mathbb{P}(Y^{(1)} \geq y^{(1)}, Y^{(2)} \geq y^{(2)})$. Replacing H by the empirical distribution function of the observed random vector Y, and G_i by its Kaplan-Meier estimator for $i = 1, 2$, Wang and wells (1997) deduces an estimator of F, which is relatively simple but does not happen to be the distribution function of a probability measure, and is therefore not suited to our framework.

Under Identifiability Assumption 1, computation of f only requires the knowledge of the survival functions G_i for $i = 1, 2$. An important feature of this framework, is that functions G_i are easy to estimate, since they are survival functions of a single random variable $C^{(i)}$. Moreover, from (2.1), and reversing the roles of $T^{(i)}$ and $C^{(i)}$, $C^{(i)}$ can be seen as a censored variable in a univariate censoring model. Hence, G_i can be estimated using Kaplan-Meier estimator, that is for $i = 1, 2$

$$\hat{G}_i(t^{(i)}) = \prod_{Y_j^{(i)} \leq t^{(i)}} \left(1 - \frac{d\hat{H}_{0,i}(Y_j^{(i)})}{\hat{H}_i(Y_j^{(i)})} \right),$$

where $\hat{H}_{0,i}(t) = \sum_{k=1}^n [1 - \Delta_k^{(i)}] \mathbf{1}_{Y_k^{(i)} \geq t}$, and $\hat{H}_i(t) = \sum_{k=1}^n \mathbf{1}_{Y_k^{(i)} \geq t}$. Convergence of these Kaplan-Meier estimators requires the independence of $T^{(i)}$ and $C^{(i)}$ which is included in our assumption.

Finally, our estimator under Identifiability Assumption 1 becomes

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n \frac{\Delta_j^{(1)} \Delta_j^{(2)}}{\mathcal{C}(\hat{G}_1(Y_j^{(1)}), \hat{G}_2(Y_j^{(2)}))} \delta_{\{Y_j^{(1)}, Y_j^{(2)}\}}, \tag{2.5}$$

and for all function $\phi \in \mathcal{F}$,

$$\hat{\mu}(\phi) = \int \phi(y) d\hat{\mu}(y) = \frac{1}{n} \sum_{j=1}^n \frac{\Delta_j^{(1)} \Delta_j^{(2)} \phi(Y_j^{(1)}, Y_j^{(2)})}{\mathcal{C}(\hat{G}_1(Y_j^{(1)}), \hat{G}_2(Y_j^{(2)}))}.$$

One can notice that the estimator $\hat{\mu}$ can be seen as a generalization of Kaplan-Meier estimator, extended to bivariate censoring. Indeed, if one assume that only the first component of T is censored, computation of $\hat{\mu}$ corresponds to the same estimator of the joint distribution function proposed by Stute (1993) (with the uncensored variable acting as the covariate in Stute (1993)). This can be seen by observing that the jump of Kaplan-Meier estimator of the distribution function of $T^{(1)}$ at observation j is $\Delta_j^{(1)}\hat{G}_1(T_j)^{-1}n^{-1}$ (Satten and Datta, 2001). Let us also mention that the generalization of this estimator to the case of multivariate censoring (in dimension $d > 2$) is straightforward, this time using a $d-$dimensional copula function. The theoretical results that we prove below extend easily to this case.

3 Main results

3.1 A general result to obtain i.i.d. representations

In this section we consider an estimator of the distribution of $(T^{(1)}, T^{(2)})$ of the following form,

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n}\Delta_j^{(1)}\Delta_j^{(2)}\hat{f}(Y_j)\delta_{\{Y_j^{(1)}, Y_j^{(2)}\}}, \tag{3.1}$$

where $\delta_{(x_1,x_2)}$ denotes the Dirac mass at point (x_1, x_2). Here, $\hat{f}$ is a random function computed from the whole sample. We give here general conditions for the measure $\hat{\mu}$ to be a $n^{1/2}-$consistent estimator of μ.

The Assumption 1 below is composed of two parts. First, it states the existence of the function f corresponding to equation (2.4). As pointed out previously, exact knowledge of the function f is generally impossible, except if we specify the law of the censoring. Hence it is relevant to consider $\hat{\mu}$ defined by (3.1). The second part of Assumption 1 states that the random function $\hat{f}$ used in the computation of $\hat{\mu}$ converges uniformly towards f.

Assumption 1 *There exists some function f such as*

$$E\left[\Delta^{(1)}\Delta^{(2)}f(Y)\phi(Y^{(1)}, Y^{(2)})\right] = E\left[\phi(T^{(1)}, T^{(2)})\right], \tag{3.2}$$

for all measurable nonnegative function ϕ. Moreover,

$$\sup_{y \in \mathcal{Y}_\tau}|\hat{f}(y) - f(y)| = o_P(1),$$

where $\mathcal{Y}_\tau = \{y^{(1)} \le \tau^{(1)}, y^{(2)} \le \tau^{(2)}\}$ for $\tau^{(i)} < \inf\{y : \mathbb{P}(Y^{(i)} > y) = 0\}$, $i = 1, 2$.

Assumption 2 below furnishes more conditions on the convergence of $\hat{f}$. One of the main difficulties in studying quantities such as $\hat{\mu}(\phi)$ stands in the fact that the terms in (3.1) are non-i.i.d. since $\hat{f}$ is computed from the whole sample. The following Assumption 2 assumes that there exists some asymptotic i.i.d. representation of $\hat{f}$.

Assumption 2 *The following representation holds for function $\hat{f}$,*

$$\hat{f}(y) - f(y) = \frac{1}{n} \sum_{j=1}^{n} \eta(Y_j, \Delta_j; y) + R(y),$$

with $\sup_{y \in \mathcal{Y}_\tau} |R(y)| = o_P(n^{-1/2})$ and $E[\eta(Y_j, \Delta_j; y)] = 0$, with η being uniformly bounded on $\mathcal{Y}_\tau$. Moreover, let

$$f^*(y) = \frac{1}{n} \sum_{j=1}^{n} \eta(Y_j, \Delta_j; y),$$

and assume that there exists a Donsker class of functions $\mathcal{G}$ such as $f^* \in \mathcal{G}$ for all n with probability tending to one.

The interest of considering Donsker classes (see van der Vaart and Wellner, 1996, for a complete definition) is that these classes of functions satisfy an uniform Central Limit Theorem property. As we will see in the following, some natural and classical Donsker classes appear in the derivations of our main results, making the use of their asymptotic properties a valuable tool.

With at hand the function f, it becomes natural to compare $\hat{\mu}$ with the following measure,

$$\tilde{\mu} = \frac{1}{n} \sum_{j=1}^{n} \Delta_j^{(1)} \Delta_j^{(2)} f(Y_i) \delta_{\{Y_j^{(1)}, Y_j^{(2)}\}}.$$

Clearly, for any function ϕ, $\tilde{\mu}(\phi)$ tends to $\mu(\phi)$ from Assumption 1 and the classical strong law of large numbers that applies to this sum of i.i.d. quantities. The next result is a general Theorem that permits to derive asymptotic i.i.d. representations for estimators such as (3.1). This general result is an important tool in our approach, since it will allow us to obtain representations for the estimator (2.5) by simply checking the two simple Assumptions 1 to 2.

Theorem 3.1 *Let $\mathcal{F}$ be a Donsker class of functions with bounded envelope Φ satisfying $\Phi(y^{(1)}, y^{(2)}) \equiv 0$ if $y^{(1)} > \tau^{(1)}$ or $y^{(2)} > \tau^{(2)}$ for $\tau^{(i)} < \inf\{y : \mathbb{P}(Y^{(i)} > y) = 0\}$, $i = 1, 2$. Then, under Assumptions 1 to 2,*

$$\hat{\mu}(\phi) - \tilde{\mu}(\phi) = \frac{1}{n} \sum_{j=1}^{n} \int \eta(Y_j, \Delta_j, t) \phi(t) f(t)^{-1} d\mu(t) + R_n(\phi),$$

with $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = o_P(n^{-1/2})$.

It is easy to deduce from Theorem 3.1 a weak convergence result (at rate $n^{1/2}$) of $\hat{\mu}(\phi)$ by simply applying the Central Limit Theorem to the main centered i.i.d. term.

Proof. By definition,

$$\hat{\mu}(\phi) - \tilde{\mu}(\phi) = \frac{1}{n} \sum_{j=1}^{n} \Delta_j^{(1)} \Delta_j^{(2)} [\hat{f}(Y_j) - f(Y_j)] \phi(Y_j) = \frac{1}{n} \sum_{j=1}^{n} \Delta_j^{(1)} \Delta_j^{(2)} f^*(Y_j) \phi(Y_j) + R_{1n}(\phi),$$

where we used the decomposition in Assumption 2, and where

$$R_{1n}(\phi) = \frac{1}{n} \sum_{j=1}^{n} R(Y_j) \phi(Y_j).$$

The remainder term in negligible, since

$$\sup_{\phi \in \mathcal{F}} |R_{1n}(\phi)| \leq \sup_{y \in \mathcal{Y}_\tau} |R(y)| \left(\frac{1}{n} \sum_{j=1}^{n} \Phi(Y_j) \right) = o_P(n^{-1/2}).$$

Since $\sup_{y \in \mathcal{Y}_\tau} |f^*(y)| = o_P(1)$ (this is a consequence of Assumption 1 and 2), and since $f^* \mathcal{F} \subset \mathcal{GF}$ is a Donsker class (see Example 2.10.9 in van der Vaart and Wellner, 1996), we can use the asymptotic equicontinuity property of Donsker classes (see e.g. van der Vaart and Wellner, 1996) and Assumption 1 to obtain

$$\hat{\mu}(\phi) - \tilde{\mu}(\phi) = \frac{1}{n} \sum_{j=1}^{n} \int \eta(Y_j, \Delta_j, t) \phi(t) f(t)^{-1} d\mu(t) + R_n(\phi),$$

where $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = o_P(n^{-1/2})$. ∎

3.2 Application to our framework

In this section, we apply Theorem 3.1 to the estimator (2.5). We assume that Identifiability Assumption 1 holds. As a consequence, we recall that equation (3.2) in Assumption 1 holds for

$$f(y) = \mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)}))^{-1},$$

and the estimated version consists of replacing the marginal survival functions by their Kaplan-Meier estimators, that is

$$\hat{f}(y) = \mathcal{C}(\hat{G}_1(y^{(1)}), \hat{G}_2(y^{(2)}))^{-1}.$$

The key argument consists of applying Theorem 3.1 to obtain our asymptotic representation in Theorem 3.2 below. This Theorem only applies for classes of functions that are zero outside some compact subset strictly included in the support of Y. The next step consists of using tightness arguments to derive our Theorem 3.3, which is valid on the whole support of Y up to some additional moment conditions.

The only additional assumption that is needed to obtained the result of Theorem 3.2 concerns the regularity of the copula function $\mathcal{C}$.

Assumption 3 *The function* $(x_1, x_2) \rightarrow \mathcal{C}(x_1, x_2)$ *is twice continuously differentiable, with uniformly bounded first order and second order derivatives, and* $\mathcal{C}(x_1, x_2) \neq 0$ *for* $x_1 \neq 0$ *and* $x_2 \neq 0$. *Moreover, we will denote* $\partial_1 \mathcal{C}(x_1, x_2)$ *(resp.* $\partial_2 \mathcal{C}_2(x_1, x_2)$*) the partial derivative of* $\mathcal{C}$ *with respect to* x_1 *(resp.* x_2*).*

With at hand this Assumption, it is easy to check the uniform convergence of $\hat{f}$ in Assumption 1. Indeed,

$$|\hat{f}(y)^{-1} - f(y)^{-1}| \leq \|D\mathcal{C}\|_\infty \left(\sup_{y \in \mathcal{Y}_\tau} |\hat{G}_1(y^{(1)}) - G_1(y^{(1)})| + \sup_{y \in \mathcal{Y}_\tau} |\hat{G}_2(y^{(2)}) - G_2(y^{(2)})| \right),$$
(3.3)

where $D\mathcal{C}$ denotes the differential of function $\mathcal{C}$ which is assumed to be bounded. Moreover, the two supremum in (3.3) tend to zero from the uniform consistency of Kaplan-Meier estimator (Stute and Wang, 1993). Then, it is straightforward to deduce that $\sup_{y \in \mathcal{Y}_\tau} |\hat{f}(y) - f(y)| = o_P(1)$ since $\hat{f}(y)^{-1}$ and $f(y)^{-1}$ are almost surely bounded away from zero for $y \in \mathcal{Y}_\tau$.

A most delicate task is to obtain a suitable representation as in Assumption 2. This is done first by using the differentiability assumptions on $\mathcal{C}$. Indeed, since, for $i = 1, 2$, $\sup_{y \in \mathcal{Y}_\tau} |\hat{G}_i(y^{(i)}) - G_i(y^{(i)})| = O_P(n^{-1/2})$ (see Gill, 1983), and using Assumption 3, we deduce that

$$\hat{f}(y) - f(y) = -\sum_{i=1,2} \frac{\partial_i \mathcal{C}(G_1(y^{(1)}), G(y^{(2)}))}{\mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)}))^2} [\hat{G}_i(y^{(i)}) - G_i(y^{(i)})] + R(y),$$

where $\sup_{y \in \mathcal{Y}_\tau} |R(y)| = o_P(n^{-1/2})$, and ∂_i denotes the partial derivative with respect to the $i-$th component. Next, one can use the asymptotic representation of Kaplan-Meier

11

estimator, provided by Stute (1996) or Gijbels and Veraverbeke (1991), that is

$$
\hat{G}_i(y) - G_i(y) \;=\; \frac{1}{n}\sum_{j=1}^{n}\left\{ \frac{(1-\Delta_j^{(i)})G_i(Y_j^{(i)}\vee y)}{H_i(Y_j^{(i)})} - \int \frac{\mathbf{1}_{Y_j^{(i)}\geq u}G_i(u\vee y)dF_i(u)}{H_i(u)F_i(u)} \right\}
$$

$$
+ \left\{ \frac{\Delta_j^{(i)}\mathbf{1}_{Y_j^{(i)}>y}}{F_i(Y_j^{(i)})} - G_i(y) \right\} + R_i(y), \tag{3.4}
$$

with $\sup_{y\in\mathcal{Y}_\tau}|R_i(y)| = o_P(n^{-1/2})$. By straightforward algebra, one can see that the ex-pecation of each of the brackets is zero. Then, the representation of Assumption 2 holds for

$$
\eta(Y_j,\Delta_j;y) \;=\; -\sum_{i=1,2}\frac{\partial_i\mathcal{C}(G_1(y^{(1)}),G_2(y^{(2)}))}{\mathcal{C}(G_1(y^{(1)}),G_2(y^{(2)}))^2}\left[\left\{\frac{\Delta_j^{(i)}\mathbf{1}_{Y_j^{(i)}>y}}{F_i(Y_j^{(i)})} - G_i(y)\right\}\right.
$$

$$
\left. + \left\{\frac{(1-\Delta_j^{(i)})G_i(Y_j^{(i)}\vee y)}{H_i(Y_j^{(i)})} - \int\frac{\mathbf{1}_{Y_j^{(i)}\geq u}G_i(u\vee y)dF_i(u)}{H_i(u)F_i(u)}\right\}\right]. \tag{3.5}
$$

By Lemma 7.3, this function η satisfies the additional requirements present in Assumption 2.

We now state can our representation theorem for the estimator (2.5), applying Theorem 3.1.

Theorem 3.2 *Let $\mathcal{F}$ be a Donsker class of functions with bounded envelope Φ satisfying $\Phi(y^{(1)},y^{(2)}) \equiv 0$ if $y^{(1)} > \tau^{(1)}$ or $y^{(2)} > \tau^{(2)}$ for $\tau^{(i)} < \inf\{y : \mathbb{P}(Y^{(i)} > y) = 0\}$, $i = 1,2$. Then, for $\hat{\mu}$ defined by (2.5), under Identifiability Assumption 1 and Assumption 3,*

$$
\hat{\mu}(\phi) - \tilde{\mu}(\phi) = \frac{1}{n}\sum_{j=1}^{n}\int \eta(Y_j,\Delta_j,t)\phi(t)\mathcal{C}(G_1(t^{(1)}),G_2(t^{(2)}))d\mu(t) + R_n(\phi),
$$

with $\sup_{\phi\in\mathcal{F}}|R_n(\phi)| = o_P(n^{-1/2})$, and where η is defined in (3.5).

In a regression framework, this result is not sufficient, since it only applies to function which vanish at the vincinity of the tail of the distribution. Fortunately, it is possible to extend the result of Theorem 3.2 to functions taking non-zero values in the whole support of Y using similar tightness arguments as for the consistency of Kaplan-Meier estimator on the whole line (Gill, 1983).

To obtain such kind of result, one has to add some assumptions on the tail of the marginal distribution. We will also require to add some restrictions on the copula function $\mathcal{C}$.

Assumption 4 *Assume that*

$$\mathcal{C}(x_1, x_2) \geq x_1^{\alpha_1} x_2^{\alpha_2}.$$

Assumption 5 *$\mathcal{F}$ is a Donsker class of locally bounded functions with positive locally bounded envelope function Φ such as*

$$\int \frac{\Phi^2(y)d\mu(y)}{\mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)}))} < \infty, \tag{3.6}$$

and, for some $\varepsilon > 0$ arbitrary small,

$$\int \Phi(y) \left[\frac{G_1^{1-\alpha_1}(y^{(1)})\mathcal{K}_1^{1/2+\varepsilon}(y^{(1)})}{G_2^{\alpha_2}(y^{(2)})} + \frac{G_2^{1-\alpha_2}(y^{(2)})\mathcal{K}_2^{1/2+\varepsilon}(y^{(2)})}{G_1^{\alpha_1}(y^{(1)})} \right] d\mu(y) < \infty \tag{3.7}$$

where for $i = 1, 2$,

$$\mathcal{K}_i(u) = - \int_0^u \frac{dG_i(t)}{G_i(t)^2 F_i(t)}.$$

Condition (3.6) corresponds to a finite asymptotic variance. Condition (3.7) is a technical assumption needed to obtain the tightness of the process. This assumption is of the same flavour as the condition (1.6) in Stute (1995), which holds for a large class of distributions of lifetimes and censoring. To illustrate this assumption, let us take the same particular example similar to the one considered by Stute (1995). Assume that

$$F_i(u) \sim cG_i(u)^{\beta_i}.$$

for u in the tail of the distribution of $T^{(1)}$. Using some straightforward computations, one can see that $\mathcal{K}_i(u) \leq F_i(u)^{-1}G_i(u)^{-1}$. Hence, condition (3.7) holds if

$$\int \frac{1}{G_2(y^{(2)})^{\alpha_2}} \int \frac{\Phi(y^{(1)}, y^{(2)})dF_1(y^{(1)}|y^{(2)})}{F_1(y^{(1)})^{\gamma_1}} < \infty,$$

where

$$\gamma_1 = \frac{\beta_1/2 + \varepsilon(1+\beta_1) + \alpha_1 - 1/2}{\beta_1},$$

and if a similar condition also holds if we reverse the roles of $y^{(1)}$ and $y^{(2)}$. This condition seems reasonable for a large number of lifetimes distribution. For example, if the lifetimes have sub-exponential tails, the condition will hold for any polynomial function Φ, provided that $\gamma_i < 1$ for $i = 1, 2$.

We now state the main result of this section.

Theorem 3.3 *Let $\mathcal{F}$ be a class of functions satisfying Assumption 5. Then, under Identifiability Assumption 1 and under Assumptions 3 and 4,*

$$\hat{\mu}(\phi) - \tilde{\mu}(\phi) = \frac{1}{n}\sum_{j=1}^{n}\int \eta(Y_j, \Delta_j, t)\phi(t)\mathcal{C}(G_1(t^{(1)}), G_2(t^{(2)}))d\mu(t) + R_n(\phi),$$

with $\sup_{\phi \in \mathcal{F}}|R_n(\phi)| = o_P(n^{-1/2})$, and where η is defined in (3.5).

Proof. The proof can be obtained from Theorem 3.2 and by checking conditions of Lemma 7.1. Indeed, we apply Lemma 7.1 for the process $P_n(t, \phi) = n^{1/2}\int \phi(y)\mathbf{1}_{y \in \mathcal{I}_t}d[\hat{\mu} - \tilde{\mu}](y)$. The weak convergence on compact subsets comes from Theorem 3.2. The first condition in Lemma 7.1 can be easily deduced from the first moment condition in Assumption 5, which is the condition needed to obtain a finite asymptotic variance. The second condition comes from Lemma 7.2, taking $M_n = n^{1/2}\mathcal{M}_n$ (which satisfies condition 3 in Lemma 7.1 using the result of Lemma 7.2), and

$$\Gamma_n(\tau) = \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j^{(1)}\delta_j^{(2)}\Phi(Y_j)\mathbf{1}_{Y_j \in \mathcal{I}_\tau}}{\mathcal{C}(G_1(Y_j^{(1)}), G_2(Y_j^{(2)}))}\left[\frac{G_1^{1-\alpha_1}(Y_j^{(1)})\mathcal{K}_1^{1/2+\varepsilon}(Y_j^{(1)})}{G_2^{\alpha_2}(Y_j^{(2)})} + \frac{G_2^{1-\alpha_2}(Y_j^{(2)})\mathcal{K}_2^{1/2+\varepsilon}(Y_j^{(2)})}{G_1^{\alpha_1}(Y_j^{(1)})}\right].$$

Conditions 4 and 5 come from the law of large numbers and the second moment condition in Assumption 5. ∎

4 Applications

In the following, we will consider that Identifiabilty Assumption 1 and assumptions 3 and 4 hold.

4.1 A class of M-estimators

In this section, we provide a general result on M-estimation in view to apply it to various regression frameworks. We consider an estimator

$$\hat{\theta} = \arg\min_{\theta \in \Theta} M_n(\theta), \tag{4.1}$$

where $\Theta \subset \mathbb{R}^k$, and M_n is a functional based on the estimator $\hat{\mu}$ introduced in the previous sections. More specifically, we consider M_n of the following form,

$$M_n(\theta) = \int \phi_\theta(y)d\hat{\mu}(y),$$

where $\{\phi_\theta, \theta \in \Theta\}$ is a parametric family of functions. Based on the results obtained for the estimation of $\hat{\mu}$, and standard conditions on the envelope of this class of functions, we provide a general result that ensures the asymptotic normality of $\hat{\theta}$.

We now state the main result of this section.

Proposition 4.1 *Let $\{\phi_\theta : \theta \in \Theta\}$, $\{\nabla_\theta\phi_\theta : \theta \in \Theta\}$, $\{\nabla_\theta^2\phi_\theta : \theta \in \Theta\}$, and assume that Assumption 5 hold for these three classes of functions.*

Let $\hat{\theta}$ be an estimator of the type (4.1), and assume that $\theta_0 = \arg\min_{\theta \in \Theta} M(\theta)$ is an interior point of Θ, where

$$M(\theta) = \int \phi_\theta(y) d\mu(y),$$

and Θ is a compact subset of $\mathbb{R}^k$.

Assume that, for any θ_1, $\theta_2 \in \Theta$, and any t,

$$|\phi_{\theta_1}(t) - \phi_{\theta_2}(y)| \leq \psi(y)\|\theta_1 - \theta_2\|^a, \tag{4.2}$$

for some $a > 0$ and where $\|\cdot\|$ denotes a norm on $\mathbb{R}^k$, and ψ satisfies the condition (3.6).

Moreover, assume that M is twice differentiable at point θ_0 with Hessian matrix Γ of full rank.

Then, under Assumptions 3, 4 and 5 ,

$$n^{1/2}(\hat{\theta}-\theta_0) = -\Gamma^{-1}\left(\frac{1}{n^{1/2}}\int \nabla_\theta\phi_{\theta_0}(t)d(\tilde{\mu}-\mu)(t) + \frac{1}{n^{1/2}}\sum_{j=1}^n \int \frac{\eta(Y_j,\Delta_j,t)\nabla_\theta\phi_{\theta_0}(t)}{\mathcal{C}(G_1(t^{(1)}),G_2(t^{(2)}))}d\mu(t)\right)+o_p(1).$$

Proof. To prove consistency of $\hat{\theta}$, it suffices to show that $\sup_\theta |M_n(\theta)-M(\theta)| = o_P(1)$. This can be easily done by applying Theorem 3.3 to the class of functions $\{\phi_\theta : \theta \in \Theta\}$. Indeed, we get

$$M_n(\theta) = \int \phi_\theta(t)d\tilde{\mu}(t) + \frac{1}{n}\sum_{j=1}^n \int \frac{\eta(Y_j,\Delta_j,t)\phi_\theta(t)}{\mathcal{C}(G_1(t^{(1)}),G_2(t^{(2)}))}d\mu(t) + R_n(\theta),$$

where $\sup_{\theta \in \Theta} |R_n(\theta)| = o_P(n^{-1/2})$. We have $\sup_{\theta \in \Theta} |\int \phi_\theta(t)d(\tilde{\mu}-\mu)(t)| = o_P(1)$. Indeed, this integral can be written as

$$\frac{1}{n}\sum_{j=1}^n \frac{\Delta_j^{(1)}\Delta_j^{(2)}\phi_\theta(Y_j)}{\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)}))} - E\left[\frac{\Delta_j^{(1)}\Delta_j^{(2)}\phi_\theta(Y_j)}{\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)}))}\right].$$

This is a centered empirical sum of functions belonging to the same Donsker class (they are of the form $g\mathcal{F}$ and condition (3.6) ensures the square integrability of their envelope), and

therefore it converges uniformly towards zero. A similar property can easily be derived for the second term in the i.i.d. expansion (which is already centered). This shows the consistency of $\hat{\theta}$.

Next, asymptotic normality of $\hat{\theta}$ is obtained using differentiation of M_n on $o_P(1)-$neighborhoods of θ_0. By definition of $\hat{\theta}$, and since M_n is differentiable with respect to θ, we have $\nabla_\theta M_n(\hat{\theta}) = 0$. Moreover, using a second order Taylor expansion for $\hat{\theta}$ in a neighborhood of θ_0,

$$\nabla_\theta M_n(\hat{\theta}) = \nabla_\theta M_n(\theta_0) + (\hat{\theta} - \theta_0)^T \nabla_\theta^2 M_n(\tilde{\theta}),$$

where $\tilde{\theta} = \theta_0 + o_P(1)$, by consistency of $\hat{\theta}$. Moreover, applying again Theorem 3.3, this time to the class of function $\{\nabla_\theta^2 \phi_\theta : \theta \in \Theta\}$, we get $\sup_{\theta \in \Theta} |\nabla_\theta^2 M_n(\theta) - \nabla_\theta^2 M(\theta)| = o_P(1)$. Hence, since $\nabla_\theta^2 M(\theta)$ is invertible in the neighborhood of θ_0, $\nabla_\theta^2 M_n(\tilde{\theta})$ is invertible on a set of probability tending to one. Then, we can write

$$\hat{\theta} - \theta_0 = -(\nabla_\theta^2 M(\theta_0) + o_P(1))^{-1} \nabla_\theta M_n(\theta_0).$$

The result follows from the weak convergence of $\nabla_\theta M_n(\theta_0)$, obtained by applying Theorem 3.3 to the function $\nabla_\theta \phi_{\theta_0}$.

∎

As a by-product of Proposition 4.1, one can easily deduce asymptotic normality of $\hat{\theta}$, by applying the Central Limit Theorem to the main term in the i.i.d. expansion. The limit covariance matrix can then be estimated from the data by replacing each unknown quantity in its expression by its empirical counterpart.

4.2   Examples

Example 1: Mean regression

In a mean-regression model, we assume that

$$T^{(2)} = r(\theta_0, T^{(1)}) + \varepsilon, \tag{4.3}$$

where $E[\varepsilon|T^{(1)}] = 0$. Note that the results that we provide are also valid if we replace $T^{(2)}$ by $\log T^{(2)}$, which is sometimes useful to obtain variables with support on $\mathbb{R}$ instead of $\mathbb{R}^+$, and therefore permits to consider symmetric unbounded distributions for ε. With at hand our estimator $\hat{\mu}$, we can extend the definition of the least-square estimator of θ_0. This leads to an estimator of the form (4.1), with

$$\phi_\theta(t) = (t^{(2)} - r(\theta, t^{(1)}))^2. \tag{4.4}$$

Assumption 5 reduces to assumptions on the tail of the marginal distribution and on the function r. In view of the discussion below Assumption 5, this Assumption will hold in the particular case where the lifetime distributions are sub-exponential, and for a polynomial function r.

Example 2: Generalized Linear Model

In a Generalized Linear Model (McCullagh and Nelder, 1983), one assumes that the conditional distribution of $T^{(2)}$ given $T^{(1)}$ belongs to some exponential family, allowing to perform maximum likelihood estimation. Therefore, one assumes that the conditional distribution of $T^{(2)}$ given $T^{(1)} = t^{(1)}$ admits a density with respect to a dominating measure of the following form,

$$f^\theta_{T^{(2)}|T^{(1)}=t^{(1)}}(t^{(2)}) = \exp\left([t^{(2)}(\theta_{01} + \theta_{02}t^{(1)}) - b(\theta_{01} + \theta_{02}t^{(1)})]a(\kappa)^{-1}\right),$$

where b is a known function, and a is a function depending on a dispersion parameter κ. Under random censoring, one could either derive a new likelihood contrast which can be numerically difficult to handle, or use the new estimator that we propose, to replace empirical sums used in the uncensored case by weighted sums. In this second approach, the estimator of θ_0 is of the form (4.1) with

$$\phi_\theta(t) = -\left([t^{(2)}(\theta_{01} + \theta_{02}t^{(1)}) - b(\theta_{01} + \theta_{02}t^{(1)})]\right).$$

Since there is a link between the canonical parameter of an exponential family and the expectation, we get

$$g(E(T^{(2)}|T^{(1)})) = \theta_{01} + \theta_{02}T^{(1)},$$

for some link function g.

5 Simulations

In the following we presente a small simulation study to evaluate the finite sample behavior of our estimator. To begin with, a mean regression analysis is performed. The following model is considered

$$E(T_2|T_1 = t_1) = a + b \times t_1$$

with $a = 1$ and $b = 3$. We consider three different ways for simulating survival and censoring times.

- Case 1: The distribution of the survival times T_1 and T_2 are normal variables such as $\mathcal{L}(T_1) \sim \mathcal{N}(0, \sigma_1^2)$ and $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{N}(a + b \times t_1, \sigma_2^2)$. The censoring times C_1 and C_2 are supposed to be exponentially distributed ($C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$) such as $E(C_1) = \lambda_1$ and $E(C_1) = \lambda_2$.

- Case 2: The distribution of the survival times T_1 is exponential such as $\mathcal{L}(T_1) \sim \mathcal{E}(\lambda)$. The survival times T_2 satisfy $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{W}(k, a + b \times t_1)$ where $\mathcal{W}$ denotes the weibull distribution. The parametrization of the weibull distribution is such that its means is $E(T_2|T_1 = t_1) = (a + bt_1)\Gamma(1 + \frac{1}{k})$. The censoring times C_1 and C_2 are such that $C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$.

- Case 3: The survival times is such as $\mathcal{L}(T_1) \sim \mathcal{E}(\gamma)$ and $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{P}(a + b \times t_1)$ such as $E(T_2|T_1 = t_1) = \frac{a + bt_1}{a + bt_1 - 1}$ where $\mathcal{P}$ denotes the Pareto distribution. The censoring times are still exponentially distributed: $C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$.

We consider 1000 datasets each with n=50, 100, 300 and 500 observations for the normal, Weibull and Pareto distributions. The results displayed in Table 1 are obtained for $\sigma_1^2 = 1$, $\sigma_2^2 = 0.5$, $\lambda = 0.5$, $k = 1.5$ and $\gamma = 1$. The values of λ_1 and λ_2 are chosen to reach 15% or 30% of censored observations for the variables T_1 and T_2. The estimations $\hat{a}$ and $\hat{b}$ obtained using equation (4.1) (with the function ϕ_θ defined in (4.4)) are compared to the true value of a and b using root mean squared error (rmse). For comparison purpose, the error obtained for uncensored observation are also given.

[Table 1]

The second application involves the following generalized linear model,

$$g(E(T_2|T_1 = t_1)) = a + b \times t_1$$

where g is a link function. In the simulations we fixed $a = 1$ and $b = 3$. The following cases are considered.

- Case 1: The distribution of the survival times T_1 and T_2 are normal variables such as $\mathcal{L}(T_1) \sim \mathcal{N}(0, \sigma_1^2)$ and $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{N}(a + b \times t_1, \sigma_2^2)$. The censoring times C_1 and C_2 are supposed to be exponentially distributed ($C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$). We consider the canonical link function $g = Id$. This case correspond to the gaussian linear model.

18

- Case 2: The distribution of the survival times are exponential such as $\mathcal{L}(T_1) \sim \mathcal{E}(\lambda)$ and $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{E}(a + b \times t_1)$. The censoring times C_1 and C_2 are such that $C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$. The canonical link $g(u) = u^{-1}$ is considered.

- Case 3: The survival times is such that $\mathcal{L}(T_1) \sim \mathcal{E}(\gamma)$ and $\mathcal{L}(T_2|T_1 = t_1) \sim \mathcal{G}(k, \frac{a+b \times t_1}{k})$ where $\mathcal{G}$ denotes the Gamma distribution. The parametrization of the Gamma distribution is such that its means is $E(T_2|T_1 = t_1) = (a + bt_1)$. The censoring times are still exponentially distributed: $C_1 \sim \mathcal{E}(\lambda_1)$ and $C_2 \sim \mathcal{E}(\lambda_2)$. The canonical link $g(u) = u^{-1}$ is used.

In each case, 1000 datasets with n=50, 100, 300 and 500 observations are considered. The values of λ_1 and λ_2 are chosen to reach 15% or 30% of censored observations for the variables T_1 and T_2. The results obtained for $\sigma_1^2 = 1$, $\sigma_2^2 = 0.5$, $\lambda = 0.5$ and $k = 2$ are given in Table 2. The rmse obtained with the proposed estimators of the form (4.1) can be compared to the one obtained for uncensored observations.

[Table 2]

As expected, the rmse displayed in table 1 and table 2 decrease when the sample size n and the percentage of censoring increase. The comparison of the error obtained with different distributions is not very useful as the results for a given distribution are greatly depending on the chosen distribution parameters. Nevertheless, for a given distribution, the results obtained with different censoring rates can be compared with the ones obtained in absence of censoring. For this purpose, we can use the ratio between the rmse under censoring and the rmse in absence of censoring. This ratio greatly depends on the percentage of observations with two uncensored components. Indeed, by construction, our estimator leads to better results when percentage of doubly censored observations is small. We can note that this ratio increases with the sample size n. As expected, the rmse in table 1 and table 2 for the normal distribution are similar since the generalized linear model for a normal distribution and identity link is equivalent to the gaussian linear model. Moreover, it appear that the rmse in table 1 and table 2 for the normal distribution are around ten times smaller than rmse for other distributions. We can also notice that the rmse obtained with 30% of censoring on T_1 and 15% on T_2 is always smaller than the rmse for 15% of censoring on T_1 and 30% on T_2 whereas the percentage of doubly censored observations is the same in this two case. It seems that it is worst to have censoring on the dependent variable than censoring on the covariate.

19

6 Conclusion

In this paper we proposed a new estimator of the joint distribution of bivariate censored lifetimes. Even if it not our main purpose (since we are focusing here on a regression framework), this technique can be used to estimate the joint distribution function. This estimator presents many advantages compared to previously existing approaches. First, it truely corresponds to a probability distribution. Second, its simplicity of computation make him a practical tool for bivariate censoring inference (or more generally multivariate censoring, since it can be extended to this case without difficulties). We derived general asymptotic representations of integrals with respect to this measure, that hold uniformly over some classes of functions. Moreover, these representation are valid on the whole distribution support (and not only on compact subsets strictly included in this support, as it is generally the case in the literature), which extends the result of Stute (1995). This key result allows us to provide asymptotic results for a large class of $M-$estimators like the one used in regression. These new $M-$estimation techniques for bivariate censoring present the advantage to be easily computable, since they only consists of adding appropriate weights into classical procedures used in absence of censoring.

7 Technicalities

Define $\mathcal{I}_\tau = \{(y^{(1)}, y^{(2)}) \ s.a. \ 0 \le y^{(1)} \le \tau^{(1)}, \ 0 \le y^{(2)} \le \tau^{(2)}\}$, where $\tau = (\tau^{(1)}, \tau^{(2)})$.

Lemma 7.1 *Let $\mathcal{F}$ be a class of functions. Let $P_n(t, \phi)$ be a process on $\mathcal{I}_{\tau_H} \times \mathcal{F}$. Define, for any $\tau \in \mathcal{I}_{\tau_H}$, $R_n(\tau, \phi) = P_n(\tau_H, \phi) - P_n(\tau, \phi)$. Assume that for all τ such that $\mathcal{I}_\tau$ is strictly included in the interior part of $\mathcal{I}_{\tau_H}$,*

$$(P_n(t, \phi))_{t \in \mathcal{I}_{\tau_H}, \phi \in \mathcal{F}} \Longrightarrow (W(V_\phi(t)))_{t \in \mathcal{I}_{\tau_H}, \phi \in \mathcal{F}},$$

where $W(V_\phi(t))$ is a Gaussian process with covariance function V_ϕ, and $\Longrightarrow$ denotes the weak convergence.

Assume that the following conditions hold,

1. *$\lim_{\tau \to \tau_H} V_\phi(\tau) = V_\phi(\tau_H)$, with $\sup_{\phi \in \mathcal{F}} |V_\phi(\tau_H)| < \infty$,*

2. *$|R_n(\tau', \phi)| \le M_n \times \Gamma_n(\tau)$, for all $\tau' \in \mathcal{I}_{\tau_H} - \mathcal{I}_\tau$,*

3. *$M_n = O_P(1)$,*

4. $\Gamma_n(\tau) \to \Gamma(\tau)$ in probability,

5. $\lim_{\tau \to \tau_H} \Gamma(\tau) = 0$.

Then $P_n(\tau_H, \phi) \Longrightarrow \mathcal{N}(0, V_\phi(\tau_H))$.

Proof. From Theorem 13.5 in Billingsley (1999) and from condition 1, it suffices to show that, for all $\varepsilon > 0$,

$$\lim_{\tau \to \tau_H} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{t \in \mathcal{I}_{\tau_H} - \mathcal{I}_\tau, \phi \in \mathcal{F}} |R_n(t, \phi)| > \varepsilon \right) = 0. \tag{7.1}$$

Using condition 2 in the Lemma, the probability in equation (7.1) is bounded, for all $M > 0$, by

$$\mathbb{P}(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)) + \mathbb{P}(M_n > M). \tag{7.2}$$

Moreover, from condition 4,

$$\limsup_{n \to \infty} \mathbb{P}(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)) = \mathbf{1}_{\varepsilon/M - \Gamma(\tau) \geq 0}.$$

Since $\Gamma(\tau) \to 0$ (condition 5), we can deduce that

$$\lim_{\tau \to \tau_H} \limsup_{n \to \infty} \mathbb{P}(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)) = 0.$$

Hence,

$$\lim_{\tau \to \tau_H} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{t > \tau, \phi \in \mathcal{F}} |R_n(t, \phi)| > \varepsilon \right) \leq \limsup_{n \to \infty} \mathbb{P}(M_n > M).$$

As a consequence,

$$\lim_{\tau \to \tau_H} \limsup_{n \to \infty} P\left(\sup_{t > \tau, \phi \in \mathcal{F}} |R_n(t, \phi)| > \varepsilon \right) \leq \lim_{M \to \infty} \limsup_{n \to \infty} P(M_n > M) = 0,$$

using the fact that $M_n = O_P(1)$ (condition 3). ∎

Lemma 7.2 *Let, for $i = 1, 2$,*

$$\mathcal{K}_i(y) = -\int_0^y \frac{dG_i(t)}{G_i(t-)^2 F_i(t)}.$$

Assume that Assumption 3 to 5 hold. Then, for all $\varepsilon > 0$,

$$\sup_{1 \leq j \leq n} \left| \frac{\delta_j^{(1)} \delta_j^{(2)}}{\mathcal{C}(\hat{G}_1(Y_j^{(1)}), \hat{G}_2(Y_j^{(2)}))} - \frac{\delta_j^{(1)} \delta_j^{(2)}}{\mathcal{C}(G_1(Y_j^{(1)}), G_2(Y_j^{(2)}))} \right|$$

$$\leq \mathcal{M}_n \left(\sum_{i=1}^2 \frac{\delta_j^{(1)} \delta_j^{(2)} G_i^{1-\alpha_i}(Y_j^{(i)}) \mathcal{K}_i^{1/2+\varepsilon}(Y_j^{(i)})}{G_{3-i}^{\alpha_{3-i}}(Y_j^{(3-i)}) \mathcal{C}(G_1(Y_j^{(1)}), G_2(Y_j^{(2)}))} \right),$$

where $|\mathcal{M}_n| = O_P(n^{-1/2})$.

21

Proof. Write

$$\frac{\delta_j^{(1)}\delta_j^{(2)}}{\mathcal{C}(\hat{G}_1(Y_j^{(1)}),\hat{G}_2(Y_j^{(2)}))} - \frac{\delta_j^{(1)}\delta_j^{(2)}}{\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)}))}$$

$$= \frac{\delta_j^{(1)}\delta_j^{(2)}[\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)})) - \mathcal{C}(\hat{G}_1(Y_j^{(1)}),\hat{G}_2(Y_j^{(2)}))]}{\mathcal{C}(\hat{G}_1(Y_j^{(1)}),\hat{G}_2(Y_j^{(2)}))\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)}))},$$

and use Assumptions 3 and 4 to bound the right-hand side by

$$\frac{M\delta_j^{(1)}\delta_j^{(2)}}{\mathcal{C}(G_1(Y_j^{(1)}),G_2(Y_j^{(2)}))}\left|\frac{|\hat{G}_1(Y_j^{(1)}) - G_1(Y_j^{(1)})|}{\hat{G}_1^{\alpha_1}(Y_j^{(1)})\hat{G}_2^{\alpha_2}(Y_j^{(2)})} + \frac{|\hat{G}_2(Y_j^{(2)}) - G_2(Y_j^{(2)})|}{\hat{G}_2^{\alpha_2}(Y_j^{(2)})\hat{G}_1^{\alpha_1}(Y_j^{(1)})}\right|,$$

for some positive constant M. The result follows from Gill (1983), using the fact that, for $i = 1, 2$,

$$\sup_{t \leq \tau_H^{(i)}} \frac{G_i(t)}{\hat{G}_i(t)} = O_P(1),$$

and that

$$\sup_{t \leq Y_{(n)}^{(i)}} \frac{|\hat{G}_i(t) - G_i(t)|}{\mathcal{K}_i^{1/2+\varepsilon}(t)G_i(t)} = O_P(n^{-1/2}).$$

This last point comes from Theorem 2.1 in Gill (1983), using that $\int \mathcal{K}_i^{-1-2\varepsilon}(t)d\mathcal{K}_i(t) < \infty$ (see condition (2.1) in Gill, 1983).

∎

Lemma 7.3 *For $y \in \mathcal{Y}_\tau$, there exists a Donsker class of functions $\mathcal{F}^*$ such as the function*

$$f^*(y) = \frac{1}{n}\sum_{j=1}^{n}\eta(Y_j, \Delta_j; y),$$

belongs to $\mathcal{F}^*$, where η is defined in equation (3.5).

Proof. For $i = 1, 2$, the functions

$$g_{1i}(y) = \frac{1}{n}\sum_{j=1}^{n}\frac{\Delta_j^{(i)}\mathbf{1}_{Y_j^{(i)}>y}}{F_i(Y_j^{(i)})},$$

$$g_{2i}(y) = \frac{1}{n}\sum_{j=1}^{n}\frac{(1-\Delta_j^{(i)})G_i(Y_j^{(i)}\vee y)}{H_i(Y_j^{(i)})},$$

$$g_{3i}(y) = \frac{1}{n}\sum_{j=1}^{n}\int\frac{\mathbf{1}_{Y_j^{(i)}\geq u}G_i(u\vee y)dF_i(u)}{H_i(u)F_i(u)},$$

22

are all monotone with respect to y, and are bounded on $\mathcal{Y}_\tau$. The class of monotone bounded functions is Donsker, see Theorem 2.7.5 in van der Vaart and Wellner (1996). The functions $y \rightarrow \partial_i \mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)})) \mathcal{C}(G_1(y^{(1)}), G_2(y^{(2)}))^{-2}$ are bounded continuously differentiable with bounded derivatives. From Corollary 2.7.2 in van der Vaart and Wellner (1996) and a permanence property of Donsker classes, the result of Lemma 7.3 follows. ∎

References

Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.*, 28(3):549–567.

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.

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

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Campbell, G. and Földes, A. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric statistical inference, Vol. I, II (Budapest, 1980)*, volume 32 of *Colloq. Math. Soc. János Bolyai*, pages 103–121. North-Holland, Amsterdam.

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

Delecroix, M., Lopez, O., and Patilea, V. (2008). Nonlinear censored regression using synthetic data. *Scand. J. Statist.*, 35(2):248–265.

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.

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1993). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. H. Poincare Probab. Statist.*, 31:547–597.

Hanley, J. A. and Parnes, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics*, 39(1):129–139.

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

Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, 9(6):1276–1288.

McCullagh, P. and Nelder, J. A. (1983). *Generalized linear models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79(3):495–512.

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

Pruitt, R. C. (1991b). Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis. Technical Report 543, University of Minnesota, Minneapolis.

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.

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.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.*, 21(3):1591–1607.

Tsai, W.-Y., Leurgans, S., and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.*, 14(4):1351–1365.

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 with applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York.

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

Table 1. Root mean squared error for the mean regression in different censoring schemes for normal, Weibull, Gamma and Pareto distribution for the explained variable.

| | | % of censoring on T_1 and T_2 | | | | |
		$0\% - 0\%$	$15\% - 15\%$	$15\% - 30\%$	$30\% - 15\%$	$30\% - 30\%$
Normal	$n = 50$	0.1041	0.1454	0.2003	0.1873	0.2377
	$n = 100$	0.0733	0.1075	0.1539	0.1324	0.1829
	$n = 300$	0.0418	0.0626	0.0994	0.0865	0.1291
	$n = 500$	0.0323	0.0479	0.0777	0.0686	0.1070
T_1 and T_2 uncensored		100%	75%	65%	65%	59%
Weibull	$n = 50$	1.1012	1.3820	1.7133	1.6008	1.8434
	$n = 100$	0.8086	1.1145	1.3178	1.2350	1.4812
	$n = 300$	0.5005	0.7664	1.0189	0.9560	1.1166
	$n = 500$	0.3782	0.6703	0.8926	0.8060	0.9932
T_1 and T_2 uncensored		100%	73%	61%	61%	52%
Pareto	$n = 50$	0.9806	1.2709	1.6042	1.5522	1.8571
	$n = 100$	0.6787	0.8753	1.0846	1.0707	1.3318
	$n = 300$	0.3790	0.5148	0.6232	0.6559	0.7718
	$n = 500$	0.2920	0.4209	0.4889	0.5395	0.5802
T_1 and T_2 uncensored		100%	70%	57%	57%	45%

Table 2. Root mean squared error in the generalized linear model with different censoring schemes for normal, exponential and Gamma distribution for the explained variable.

		% of censoring on T_1 and T_2				
		$0\% - 0\%$	$15\% - 15\%$	$15\% - 30\%$	$30\% - 15\%$	$30\% - 30\%$
Normal	$n = 50$	0.1041	0.1454	0.2003	0.1873	0.2377
	$n = 100$	0.0716	0.1035	0.1566	0.1382	0.1785
	$n = 300$	0.0423	0.0617	0.1006	0.0871	0.1306
	$n = 500$	0.0328	0.0477	0.0812	0.0674	0.1054
T_1 and T_2 uncensored		100%	74%	64%	64%	59%
Exponential	$n = 50$	1.1954	1.7062	2.3058	2.2722	2.7985
	$n = 100$	0.8716	1.2962	1.6910	1.8218	2.0815
	$n = 300$	0.4538	0.6883	0.9232	1.0633	1.3656
	$n = 500$	0.3513	0.5560	0.7024	0.8694	1.1135
T_1 and T_2 uncensored		100%	73%	61%	61%	52%
Gamma	$n = 50$	0.8205	1.1849	1.5013	1.4354	1.7978
	$n = 100$	0.5523	0.7993	1.0376	1.0221	1.2758
	$n = 300$	0.3181	0.4685	0.6023	0.6119	0.8181
	$n = 500$	0.2487	0.3562	0.4697	0.4931	0.6303
T_1 and T_2 uncensored		100%	73%	61%	61%	52%