



HAL
open science

A generalization of Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models

Olivier Lopez

► **To cite this version:**

Olivier Lopez. A generalization of Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models. 2012. hal-00663340

HAL Id: hal-00663340

<https://hal.science/hal-00663340>

Preprint submitted on 26 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A generalization of Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models

Olivier LOPEZ¹

January 26, 2012

Abstract

In this paper, we provide a new nonparametric estimator of the joint distribution of two lifetimes under random right censoring and left truncation, which can be seen as a bivariate extension of the Kaplan-Meier estimator. We derive asymptotic results for this estimator, including uniform $n^{1/2}$ -consistency, and develop a general methodology to study bivariate lifetime modelling, which is a critical issue in the study of pensions with a reversion condition. Application to goodness-of-fit for survival copula models is discussed. We show that the procedure that we use are consistent, and propose a bootstrap procedure based on our estimator to compute the critical values. All the new techniques that we propose are experimented on the Canadian data-set initially studied by Frees et al. (1996).

Key words: Kaplan-Meier estimator, bivariate censoring, left-truncation, survival analysis, lifetime analysis, nonparametric estimation, copula models, bootstrap.

Short title: Nonparametric estimation for two lifetimes.

¹ Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France, E-mail: olivier.lopez0@upmc.fr.

1 Introduction

Pension contracts with reversion clauses depend on two random lifetimes: the one of the policyholder, and the one of his/her wife/husband. While studying the lifetime of a single individual is a classical problem for insurers, bivariate mortality analysis is quite more delicate. The easiest solution would be to study each lifetime separately, assuming that these two random variables are independent. However, this independence assumption may be far to be true when we are considering two members of the same couple. Several approaches have been proposed to model such lifetimes jointly, see e.g. Frees et al. (1996), Carriere (2000), Luciano et al. (2008). These approaches are essentially based on parametric or semiparametric models (typically copula models), and therefore rely on rather strong assumptions. On the other hand, in this setting there is a serious lack of nonparametric estimators that may be used, for example, to perform model checking. The main reason for this absence stands in the difficulty to handle bivariate truncation and bivariate censoring, which compose the main specificities of lifetime datas. The main aim of this paper is to define a new nonparametric estimator of the joint probability law of a random vector of lifetimes, under a particular censoring and truncation scheme which corresponds to the typical situation encountered in insurance data-sets. As an example of application, we consider the Canadian data-set initially studied by Frees et al. (1996)¹. In the univariate case, Kaplan-Meier estimator (Kaplan and Meier (1958), generalized to the case of left-truncation by Tsai et al. (1987)) plays this role of a purely nonparametric estimator of the survival function which can be used as a benchmark for any parametric or semiparametric model. However, in the bivariate case, such an estimator of the joint survival function is more difficult to find. The most classical one can be found in Dabrowska (1988). The main drawback of this estimator stands in the fact that it does not define a true probability definition, making it hard to use when it comes to estimate any other quantity than the survival function itself. In fact, Dabrowska's estimator assigns negative mass to some observations (see Pruitt (1991a)) which is, at least, counterintuitive. Alternative estimators usually present similar drawbacks. Akritas and Van Keilegom (2003) and Lopez and Saint Pierre (2011), under bivariate censoring, proposed different kind of estimators that happen to define true distribution. However, the estimator of Akritas and Van Keilegom (2003) requires a careful choice of a smoothing parameter, while the one defined by Lopez and Saint Pierre (2011) require additional assumptions on the censoring

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

variables which may not be realistic for insurance data-sets. In the present paper, we restrain ourselves to a particular situation which corresponds to the situation observed for the Canadian data-set. The two main assumptions that we make is that censoring occurs at the same time for the two members of the couple, and that the only cause of censoring is the fact that one member (at least) of the couple is still alive at the end of the statistical study. Using these two assumptions allows to define a simple estimator of the joint distribution. We show that this estimator is uniformly $n^{1/2}$ -consistent for estimating the survival function and a large number of related quantities (such as Kendall's τ coefficient for instance). Applications to parametric estimation, model checking and bootstrap are considered. The rest of the paper is organized as follows. In section 2, we define the specific bivariate truncation and censoring scheme under which we work in the following. We use the assumptions of the model to define our nonparametric estimator of the joint distribution. In Section 3, we provide some theoretical properties of this estimator, giving some uniform convergence rate. We particularly focus on the estimation of quantities which may be used to perform copula modelling and fitting of the two lifetimes, and provide a bootstrap procedure that manage to take account for the presence of censoring and truncation. A detailed application of these techniques to the Canadian data-set is performed in section 4.

2 A nonparametric estimator of the joint distribution of two lifetimes

The aim of this section is to define a nonparametric estimator for the joint distribution of two lifetimes, under a right-censoring and left-truncation model which is presented in section 2.1. Our estimator is based on a partial derivative equation which is described in section 2.2, while the estimator itself is presented in section 2.3. Comparison with the construction of Kaplan-Meier estimator in the univariate case, and with alternative estimators, is done in section 2.4.

2.1 Bivariate right-censoring and left-truncated model

In the following, we consider two lifetimes (T, U) , and i.i.d. replications $(T_i, U_i)_{1 \leq i \leq n}$ of these random variables. We assume that these random variables are absolutely continuous with respect to Lebesgue's measure on the plane. A bivariate right-censoring model occurs

when the random variables $(T_i, U_i)_{1 \leq i \leq n}$ are not directly observed, that is when only the following data is available,

$$\begin{cases} Y_i &= \inf(T_i, C_i), \text{ \& } \delta_i = \mathbf{1}_{T_i \leq C_i} \\ Z_i &= \inf(U_i, D_i), \text{ \& } \gamma_i = \mathbf{1}_{U_i \leq D_i}, \end{cases}$$

where $(C_i, D_i)_{1 \leq i \leq n}$ consists of i.i.d. replications of a random bivariate censoring vector (C, D) , and $(\delta_i, \gamma_i)_{1 \leq i \leq n}$ are indicator functions allowing the distinction between censored and uncensored observations. In addition to right-censored observation, we consider bivariate left-truncation. Introducing two random variables (μ, ν) , the random vector (Y, Z, δ, γ) is observed only if $Y \geq \mu$ and $Z \geq \nu$. Therefore, observations are made of $(Y_i, Z_i, \delta_i, \gamma_i, \mu_i, \nu_i)_{1 \leq i \leq n}$ where $Y_i \geq \mu_i$ and $Z_i \geq \nu_i$. In order to be able to estimate the joint distribution of (T, U) , some identifiability assumptions are required. In the univariate case, classical assumptions consists of assuming the independence of T from (C, μ) , and of C from (T, μ) , and of assuming that the probability of observing non-truncated observations is non-zero (otherwise, the set of observations is almost surely empty). Throughout this paper, we will use a multivariate extension of such assumptions, which is summarized in Assumption 1 below.

Assumption 1 *Assume that*

1. (T, U) is independent from (C, D, μ, ν) ,
2. (C, D) is independent from (T, U, μ, ν) ,
3. $\mathbb{P}(Y \geq \mu, Z \geq \nu) > 0$.

It is important to mention that no additional assumption is made under the random variables (C, D) , or (μ, ν) , which may have a dependence structure that we do not need to specify.

If we consider the application of such a model in the study of the joint distribution of the lifetimes of two members of the same couple, T (resp. U) will denote the total lifetime of the husband (resp. his wife), C (resp. D) will denote the age at which the husband (resp. the wife) stops being under observation for any other cause than death. Causes for exiting the statistical study may be various. If one of the members of the couple is not dead at the end of the statistical study, his/her lifetime is censored, but other causes may exist, especially in cases where people can surrender their contract (and therefore stop being under observation). Regarding the truncation variables, μ (resp. ν) will denote the

age of the husband (resp. his wife) at the beginning of the statistical study (or at the time when he joins the study after its beginning). Of course, Y and Z , which correspond to the age at which the individuals leave the study for whatever cause, can only be observed if they are both larger than the ages at which the two members of the couple join the study.

Our aim is to provide a nonparametric estimation of the distribution of the random vector (T, U) . In absence of truncation or censoring, this task can be easily performed by using the empirical distribution, which is the distribution defined by the multivariate empirical distribution function, $F_{emp}(t, u) = n^{-1} \sum_{i=1}^n \mathbf{1}_{T_i \leq t, U_i \leq u}$, which estimates $F(t, u) = \mathbb{P}(T \leq t, U \leq u)$. Indeed, this standard procedure consists of assigning the mass n^{-1} at each observation, since, in a non-censored and non-truncated framework, each observation is expected to bring the same amount of information. In our model, F_{emp} is of course unavailable, since it would require complete observation of (T, U) . We propose to rely on an estimator of the following form,

$$\hat{F}(t, u) = \sum_{i=1}^n W_{i,n} \mathbf{1}_{Y_i \leq t, Z_i \leq u}. \quad (2.1)$$

An advantage of such a representation of our estimator \hat{F} is that there is a quick connection between such an estimator and the underlying distribution he defines. Indeed, $W_{i,n}$ can be interpreted as the mass assigned to observation i . The weight $W_{i,n}$ will be designed to compensate the lack of information caused by censoring and truncation in some parts of the distribution. Based on an estimator of the type (2.1), it becomes easy to estimate general quantities such as $E[\phi(T, U)] = \int \int \phi(t, u) dF(t, u)$ for some function ϕ , by computing

$$\int \int \phi(t, u) d\hat{F}(t, u) = \sum_{i=1}^n W_{i,n} \phi(Y_i, Z_i). \quad (2.2)$$

Estimation of the joint distribution or survival function is only a special case of this estimation problem. Our theoretical results will show that we can obtain consistency of integrals of the type (2.2) under appropriate conditions.

To properly define the weights $W_{i,n}$, we will proceed in two steps. Section 2.2 provides a key relationship that will only allow to define an ideal (but unfortunately uncomputable) estimator. Nevertheless, this relationship will allow us to define a true estimator of the type (2.1), which achieves some kind of asymptotic equivalence with the ideal one, as it is shown in section 3.1.

2.2 A key relationship for our bivariate lifetime model

In order to state the key result of our approach, let us first introduce some notation. Define

$$\begin{aligned} S_F(t, u) &= \mathbb{P}(T > t, U > u), \\ V(t, u) &= \mathbb{P}(Y \leq t, Z \leq u, \delta = \gamma = 1 | Y \geq \mu, Z \geq \nu), \\ H(t, u) &= \mathbb{P}(\mu \leq t \leq Y, \nu \leq u \leq Z | Y \geq \mu, Z \geq \nu). \end{aligned}$$

The following proposition establishes a relationship between these three functions and $dF(t, u)$.

Proposition 2.1 *Under Assumption 1,*

$$dF(t, u) = \frac{S_F(t, u)dV(t, u)}{H(t, u)} = \frac{\{\int_{t^+} \int_{u^+} dF(t', u')\} dV(t, u)}{H(t, u)}, \quad (2.3)$$

or, equivalently, for any $\phi \in L^1$,

$$E[\phi(T, U)] = \int \int \frac{\phi(y, z)S_F(y, z)dV(y, z)}{H(y, z)} = E[\delta\gamma w(Y, Z)\phi(Y, Z) | Y \geq \mu, Z \geq \nu],$$

with

$$w(y, z) = \frac{S_F(y, z)}{H(y, z)}. \quad (2.4)$$

Before proving Proposition 2.1, let us mention how it allows us to define an ideal estimator of F . Defining $W_{i,n}^* = n^{-1}\delta_i\gamma_i w(Y_i, Z_i)$, consider

$$F^*(t, u) = \sum_{i=1}^n W_{i,n}^* \mathbf{1}_{Y_i \leq t, Z_i \leq u}. \quad (2.5)$$

Based on Proposition 2.1, we see that

$$\int \int \phi(t, u)dF^*(t, u) \rightarrow \int \int \phi(y, z)w(y, z)dV(y, z) = E[\phi(T, U)], \text{ a.s.},$$

from the strong law of large numbers, when the sample size increases. Of course, F^* can not be computed, since it would require the knowledge of function H and of function $S_F(t, u) = \int_{t^+} \int_{u^+} dF(t', u')$. Nevertheless, as announced, F^* has to be understood as some benchmark estimator, which presents the advantage of depending only on $(Y_i, Z_i, \delta_i, \gamma_i)_{1 \leq i \leq n}$ and not on the unobservable variables $(T_i, U_i)_{1 \leq i \leq n}$.

Let us mention that F^* does not necessarily defines a true probability distribution, since there is no reason than $\sum_{i=1}^n W_{i,n}^* = 1$. We will denote as $W_{\infty, \infty}^* = 1 - \sum_{i=1}^n W_{i,n}^*$

which can be interpreted as the additional mass that has to be put at infinity (that is (∞, ∞)) in order to achieve a total mass equal to 1 (let us note that $W_{\infty, \infty}^*$ can even be negative, in the case where the total mass of F^* is larger than 1). Similar phenomenon occur when we look at Kaplan-Meier estimator in the univariate case. If the largest observation is uncensored, then the total mass of Kaplan-Meier estimator is one, otherwise, an additional mass is missing that can be assumed to be put at infinity.

We now state the proof of Proposition 2.1.

Proof of Proposition 2.1. Define

$$m^*(\phi) = \int \int \frac{\phi(y, z) S_F(y, z) dV(y, z)}{H(y, z)}.$$

This quantity is equal to

$$E[\delta\gamma w(Y, Z, \mu, \nu)\phi(Y, Z)|Y \geq \mu, Z \geq \nu] = E[\mathbf{1}_{T \leq C, U \leq D} w(T, U, \mu, \nu)\phi(T, U)|Y \geq \mu, Z \geq \nu].$$

Next, the right-hand side can be rewritten in the following way,

$$\begin{aligned} m^*(\phi) &= \frac{E[\mathbf{1}_{T \leq C, U \leq D} \mathbf{1}_{T \geq \mu, U \geq \nu} w(T, U, \mu, \nu)\phi(T, U)]}{\mathbb{P}(Y \geq \mu, Z \geq \nu)} \\ &= \frac{E[E[\mathbf{1}_{T \leq C, U \leq D} \mathbf{1}_{T \geq \mu, U \geq \nu} w(T, U, \mu, \nu)\phi(T, U)|T, U, \mu, \nu]]}{\mathbb{P}(Y \geq \mu, Z \geq \nu)} \\ &= \frac{E[E[\mathbf{1}_{T \leq C, U \leq D}|T, U, \mu, \nu] \mathbf{1}_{T \geq \mu, U \geq \nu} w(T, U, \mu, \nu)\phi(T, U)]}{\mathbb{P}(Y \geq \mu, Z \geq \nu)}. \end{aligned}$$

We now use the fact that (C, D) is independent from (T, U, μ, ν) , which leads to $E[\mathbf{1}_{T \leq C, U \leq D}|T, U, \mu, \nu] = S_G(T-, U-)$, where $S_G(t, u) = \mathbb{P}(C > t, D > u)$ is the joint survival function of (C, D) . Next, we get

$$m^*(\phi) = \frac{E[E[\mathbf{1}_{T \geq \mu, U \geq \nu}|T, U] S_G(T-, U-) w(T, U, \mu, \nu)\phi(T, U)]}{\mathbb{P}(Y \geq \mu, Z \geq \nu)},$$

and the independence of (μ, ν) from (T, U) leads to $E[\mathbf{1}_{T \geq \mu, U \geq \nu}|T, U] = L(T, U)$, where $L(t, u) = \mathbb{P}(\mu \leq t, \nu \leq u)$. The proof will be complete if we show that w defined by (2.4) is equal to $S_G(t-, u-)^{-1} L(t, u)^{-1} \mathbb{P}(Y \geq \mu, Z \geq \nu)$.

To this aim, observe that, by definition,

$$H(t, u) = \mathbb{P}(\mu \leq t \leq Y, \nu \leq u \leq Z|Y \geq \mu, Z \geq \nu) = \frac{\mathbb{P}(Y \geq t, Z \geq u) \mathbb{P}(\mu \leq t, \nu \leq u)}{\mathbb{P}(Y \geq \mu, Z \geq \nu)},$$

where we used the independence between (Y, Z) and (μ, ν) . Observing that $\mathbb{P}(Y \geq t, Z \geq u) = S_G(t-, u-) S_F(t-, u-)$ from the independence between (T, U) and (C, D) , we finally get $H(t, u) = S_G(t-, u-) S_F(t-, u-) L(t, u) \mathbb{P}(Y \geq \mu, Z \geq \mu)$, and the result follows, since $S_F(t-, u-) = S_F(t, u)$ for absolutely continuous variables. ■

2.3 A computable estimator

In the previous section, we showed that the distribution function F was solution of the partial derivative equation (2.3). We use the term of partial derivative equation, since, in the case where F corresponds to an absolutely continuous measure, $dF(t, u) = f(t, u)dtdu$, where $f(t, u) = \partial_t \partial_u F(t, u)$. This equation depends on two unknown quantities which are V and H . Nevertheless, V and H can be estimated empirically with $n^{1/2}$ -consistent estimators, that is

$$\begin{aligned}\hat{V}(t, u) &= \frac{1}{n} \sum_{i=1}^n \delta_i \gamma_i \mathbf{1}_{Y_i \leq t, Z_i \leq u}, \\ \hat{H}(t, u) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\mu_i \leq t \leq Y_i, \nu_i \leq u \leq Z_i}.\end{aligned}$$

Therefore, we propose to define our estimator \hat{F} as a solution of an empirical version of equation (2.3), that is \hat{F} solution of

$$d\hat{F}(t, u) = \frac{\left\{ \int_{t^+} \int_{u^+} d\hat{F}(t', u') \right\} d\hat{V}(t, u)}{\hat{H}(t, u)}. \quad (2.6)$$

with the condition that the total mass of \hat{F} is equal to one, in order to define a true probability distribution. Since \hat{V} is a sub-distribution function which puts mass only at observations such that $\delta_i = \gamma_i = 1$, \hat{F} is of the form (2.1), with $W_{i,n} = 0$ when $\delta_i = 0$ or $\gamma_i = 0$. This simple form will allow us to show existence and unicity of this function \hat{F} , and will also provide an easy way to compute it.

Indeed, in this case, solving (2.6) is equivalent to determine a vector $\mathbf{W} = (W_{1,n}, \dots, W_{n,n}, W_{\infty,\infty})$, where, again, $W_{\infty,\infty}$ represents the eventual additional mass that as to be put at infinity. Equation (2.6) can then be rewritten in a matricial form, as

$$(I_{n+1} - \hat{A})\mathbf{W} = 0, \quad (2.7)$$

where I_{n+1} denotes the identity matrix of \mathbb{R}^{n+1} , and $\hat{A} = (\hat{a}_{i,j})_{1 \leq i,j \leq n}$, with $\hat{a}_{n+1,n+1} = 1$, and, for $(i, j) \in \{1, \dots, n\}^2$,

$$\hat{a}_{i,j} = \frac{\delta_j \gamma_j \delta_i \gamma_i \mathbf{1}_{Y_j \geq Y_i, Z_j \geq Z_i}}{n \hat{H}(Y_i, Z_i)},$$

and $\hat{a}_{i,n+1} = \delta_i \gamma_i n^{-1} \hat{H}(Y_i, Z_i)^{-1}$, $\hat{a}_{j,n+1} = 0$. Let us observe that there is no problem of definition for coefficients $\hat{a}_{i,j}$, since $\hat{H}(Y_i, Z_i) > 0$, for all i .

Theorem 2.2 ensures existence of solutions of (2.7). Moreover, the proof furnishes a way to solve explicitly this equation.

Theorem 2.2 *Under Assumption 1, there exists a unique solution of (2.7) such that:*

1. $W_{\infty, \infty} > 0$,
2. $W_{i, n} = 0$ if $\delta_i = 0$ or $\gamma_i = 0$,
3. $\sum_{i=1}^{n+1} W_{i, n} = 1$.

Moreover, $W_{i, n} > 0$ for all i such that $\delta_i = \gamma_i = 1$.

Proof. Let m be the number of doubly uncensored observations. Since the weight of all other observations is assumed to be zero, solving (2.7) under the restriction that $W_{i, n} = 0$ if $\delta_i = 0$ or $\gamma_i = 0$ is equivalent to solving

$$(I_{m+1} - \tilde{A})\tilde{\mathbf{W}} = 0,$$

where \tilde{A} is a matrix of size $(m+1) \times (m+1)$, where we removed all the lines and columns corresponding to the censored observations, and $\tilde{\mathbf{W}}$ is the vector of all weights for doubly uncensored observations. Moreover, we can reorder the terms in $\tilde{\mathbf{W}}$ with respect to the values of Y (the first component of $\tilde{\mathbf{W}}$ corresponds to the weight that is affected to the doubly uncensored observation with the smallest value of Y , and so on until the m -th component which corresponds to the largest uncensored value of Y). We will denote by $Y_{(1)} < \dots < Y_{(m)}$ the ordered values of uncensored variables Y , and $Z_{(1)} < \dots < Z_{(n)}$ the corresponding values of uncensored variables Z (ordered with respect to Y). With a slight abuse in notations, we do not change the notation \tilde{A} after reordering. Next, one can observe that

$$I_{m+1} - \tilde{A} = I_{m+1} - \begin{pmatrix} 0 & \dots & \dots & \dots \\ 0 & \ddots & \dots & \dots \\ \vdots & \ddots & 0 & \dots \\ 0 & \dots & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \dots & \dots & \dots \\ 0 & \ddots & \dots & \dots \\ \vdots & \ddots & 1 & \dots \\ 0 & \dots & 0 & 0 \end{pmatrix}, \quad (2.8)$$

since the i -th component of $\tilde{A}\tilde{\mathbf{W}}$ corresponds to $n^{-1}\hat{H}(Y_{(i)}, Z_{(i)})^{-1} \sum_j \tilde{W}_j \mathbf{1}_{j \in R_i}$, where $R_i = \{j : Y_{(j)} > Y_{(i)}, Z_{(j)} > Z_{(i)}\} \cup \{\infty\}$. Therefore, we see that $R_i \subset \{i+1, \dots, n\} \cup \{\infty\}$, which leads to the triangular form of \tilde{A} with zeros on the diagonal, except for the last coefficient corresponding to the mass at infinity. It follows from (2.8) that the dimension of the kernel of $I_{m+1} - \tilde{A}$ is exactly one, showing the existence and unicity of a non-zero solution of (2.7) satisfying conditions 2 and 3 of the statement of Theorem 2.2.

We will now prove that the coefficients of \tilde{W} are strictly positive if we impose $W_{\infty, \infty} > 0$. To this aim, it suffices to consider a solution of (2.7) with $\tilde{W}'_{\infty} = 1$, the conclusion being deduced after normalization by the sum of the coefficients \tilde{W}'_j . This proof can be done recursively, and furnishes an explicit solution of (2.7). Indeed, $\tilde{W}'_m = n^{-1} \hat{H}(Y_{(m)}, Z_{(m)})^{-1} \tilde{W}'_{\infty}$ is strictly positive since W'_{∞} is. Since $R_{m-1} \subset \{m\} \cup \{(\infty, \infty)\}$, since \hat{H} is a positive function and since $\tilde{W}'_m > 0$ and $\tilde{W}'_{\infty} > 0$, we can conclude that $\tilde{W}'_{m-1} > 0$. The same procedure can be applied to show that $\tilde{W}'_{m-2} > 0, \dots, \tilde{W}'_1 > 0$. Moreover, we see that these weights can be computed recursively, beginning with the mass at infinity, since the mass at an observation only depends on the mass affected to the observations with a largest value of Y . The proof is complete by taking $W_{i,n} = \tilde{W}'_{i,n} (\sum \tilde{W}'_{j,n})^{-1}$, and $W_{\infty} = \tilde{W}'_{\infty} (\sum \tilde{W}'_{j,n})^{-1}$. ■

Remark 2.3 *Adding a mass at infinity is necessary if we wish to ensure existence of solutions of (2.7). However, as it will appear in the proof of our Theorem 3.1, we will see that this mass will be quite close to zero in practice (of order $n^{-3/2}$). Nevertheless, there exists a particular situation in which we can slightly modify our estimator in order to reallocate the small weight of (∞, ∞) at the "largest" observation. Consider the case where there exists some observation i_0 such that $Y_{i_0} = \max_{1 \leq i \leq n}(Y_i)$ and $Z_{i_0} = \max_{1 \leq i \leq n}(Z_i)$. In this case (Y_{i_0}, Z_{i_0}) can be considered as the largest observation in a bivariate sense, since all observations are smallest with respect to both Y and Z . If this observation is uncensored, that is $\delta_{i_0} = \gamma_{i_0} = 1$, we can slightly modify equation (2.7) by simply putting the residual mass at point (Y_{i_0}, Z_{i_0}) instead of putting it at (∞, ∞) . The new system of n equations then becomes,*

$$W_{i,n} = n^{-1} \hat{H}(Y_j, Z_j)^{-1} \sum_{i=1}^n W_{j,n} \mathbf{1}_{Y_j > Y_i, Z_j > Z_i}, \text{ for } i \neq i_0,$$

$$W_{i_0,n} = W_{i_0,n},$$

and the same proof as for Theorem 2.2 applies, by searching a solution of the system with $W_{i_0,n} = 1$ (instead of $W_{\infty, \infty} = 1$).

2.4 Comparison with other estimators

Comparison to Kaplan-Meier estimator.

In the univariate case, Kaplan-Meier estimator (Kaplan and Meier (1958)) is used to estimate the distribution of right-censored random variable, and Tsai et al. (1987)

proposed an adaptation which is used for dealing with left-truncation. This estimator satisfies a differential equation which has to be connected with our Proposition 2.1. Indeed, this estimator $\hat{F}(t)$ of $F(t) = \mathbb{P}(T \leq t)$ can be rewritten as

$$\hat{F}(t) = \sum_{i=1}^n \mathcal{W}_{in} \mathbf{1}_{Y_i \leq t}.$$

Looking at the expression of the weight \mathcal{W}_{in} in Sánchez Sellero et al. (2005), we see that this weight can be expressed as

$$\mathcal{W}_{in} = \frac{\delta_i(1 - \hat{F}(T_i))}{\hat{H}(T_i)},$$

where $H(t) = \mathbb{P}(\mu \leq t \leq T)$, which is the one-dimensional counterpart of (2.4).

Comparison to other existing estimators in the case of bivariate censoring.

Most of the estimators used for studying bivariate censoring only focus on the estimation of the joint survival function. A consequence is that many of them do not define regular distributions. The estimator proposed by Campbell and Földes (1982) is not a proper survival function due to nonmonotonicity. Similarly, Dabrowska's estimator (Dabrowska (1988)) assigns negative mass to some points in the plane, as it is shown in Pruitt (1991a). As a consequence, since Dabrowska's does not define a probability distribution, it can not be used for performing bootstrap. Moreover, assigning negative mass to some observations is, at most, counterintuitive. More generally, defining the mass that some estimator affects to each observation is generally difficult for previously existing estimators, making them difficult to use when it comes to estimate integrals with respect to the distribution of (T, U) . See also Prentice and Cai (1992), Gill et al. (1993) for modifications of these estimators that suffer from the same drawbacks.

Another class of estimators relies on nonparametric maximum likelihood techniques (NPMLE). The computation of the estimator is usually delicate, since it is defined implicitly (see Pruitt (1993), Pruitt (1991b) and van der Laan (1996)), while the convergence rate can be lower than $n^{1/2}$. This problem of slow convergence rates appears in Akritas and Van Keilegom (2003), which is one of the rare estimators being a true survival function. The estimator proposed by Lopez and Saint Pierre (2011), achieves this usual convergence rate and defines a true distribution, but relies on a restrictive assumption on the joint law of the censoring, that can be difficult to justify in the framework of insurance contracts.

As we will show in the following, the convergence rate of the estimator that we propose in the present paper is $n^{1/2}$, while its computation is quite simple (and can be done without

any approximation). Moreover, the fact that it defines a true distribution allows to use it to perform bootstrap procedure, as it is shown in section 3.3.

3 Theoretical behaviour of the new estimator and bootstrap procedure

The aim of this section is to show that our new estimator of the distribution is $n^{1/2}$ -consistent. In section 3.1, we provide a general result for integrals of the type $\int \int \phi(t, u) d\hat{F}(t, u)$. Some particular cases are considered in section 3.2. Finally, in section 3.3, we show how a bootstrap procedure can be implemented in order to evaluate estimation errors.

3.1 $n^{1/2}$ -consistency of the new estimator

To prove $n^{1/2}$ -consistency, we compare the estimator \hat{F} proposed in the previous section to the ideal estimator F^* defined in (2.5). Indeed, estimator F^* is very simple to study, since it is a sum of i.i.d. quantities. Therefore, the most delicate result that we need to prove is contained in Theorem 3.1 below, which aims to compare the mass $W_{i,n}$ to the ideal one $W_{i,n}^*$.

Theorem 3.1 *Under Assumption 1, assume that there exists a constant K such that*

$$\frac{1}{H(Y, Z)} < K, \text{ a.s.} \quad (3.1)$$

then

$$\sup_{1 \leq i \leq n} |W_{i,n} - W_{i,n}^*| + |W_{\infty, \infty} - W_{\infty, \infty}^*| = O_P(n^{-3/2}).$$

The proof of this result is postponed to the Appendix section. In this statement, let us discuss (3.1). This assumption has to be linked with the difficulty to estimate the tail of the distribution (in a bivariate sense, this means when y and z tend to infinity). A particular case in which this assumption holds is when the censoring has a support included in some compact set $[0, t_{max}] \times [0, u_{min}]$, strictly included in the support of (T, U) , with $\mathbb{P}((C, D) = (t_{max}, u_{min})) \neq 0$. However, as it will be shown in the simulations we conduct in section 3.2.2, the fact that this assumption does not hold does not seem to have a strong impact in practice. Some similar problems occur when studying the asymptotics of Kaplan-Meier estimator in the univariate case. Many existing results are only valid on compact subsets strictly included in the support of the distribution (see e.g. Van Keilegom

and Akritas (1999), Du and Akritas (2002), Lopez (2011)), because of similar problems of denominators close to zero that are delicate to handle from a theoretical point of view. With additional work, these conditions can be replaced by moments conditions (which, unfortunately, can not be checked in practice), such as the one provided by Stute (Stute (1995)). However, developing similar arguments in order to relax condition (3.1) is quite delicate in our framework (without significant practical impact), and is therefore left to future research.

With at hand Theorem 3.1, we can now state our main result on integrals of the form $\int \int \phi(t, u) d\hat{F}(t, u)$.

Theorem 3.2 *Let \mathcal{F} denote a class of functions such that, for all $\phi \in \mathcal{F}$, $\phi(\infty, \infty) = 0$, and $|\phi(y, z)| \leq \Phi(y, z)$ with $E[\Phi(Y, Z)] < \infty$. Then, under the assumptions of Theorem 3.1,*

$$\int \int \phi(t, u) d\hat{F}(t, u) = \int \int \phi(t, u) dF^*(t, u) + R_n(\phi), \quad (3.2)$$

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

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \gamma_i S_F(Y_i, Z_i) \phi(Y_i, Z_i)}{H(Y_i, Z_i)} - E[\phi(Y, Z)] \right| = O_P(n^{-1/2}), \quad (3.3)$$

we get

$$\sup_{\phi \in \mathcal{F}} \left| \int \int \phi(t, z) d\hat{F}(t, z) - \int \int \phi(t, z) dF(t, z) \right| = O_P(n^{-1/2}). \quad (3.4)$$

Condition (3.3) will hold as long as \mathcal{F} is a Donsker class of functions (see van der Vaart and Wellner (1996) for a precise definition), that is if \mathcal{F} satisfies an uniform central theorem property. In particular, the class of indicator functions $(T, U) \rightarrow \mathbf{1}_{T \leq t, U \leq u}$ where (t, u) belong to some set \mathcal{T} satisfies condition (3.3), allowing to obtain uniform consistency of \hat{F} as an estimator of F (see section 3.2.1 for more details).

Proof of Theorem 3.2. Write

$$\int \int \phi(t, u) d\hat{F}(t, u) = \int \int \phi(t, z) dF^*(t, z) + \sum_{i=1}^n (W_{i,n} - W_{i,n}^*) \phi(Y_i, Z_i).$$

The second term is bounded by

$$\left(n^{-1} \sum_{i=1}^n \Phi(Y_i, Z_i) \right) \times \sup_i n |W_{i,n} - W_{i,n}^*|.$$

Next, $n^{-1} \sum_{i=1}^n \Phi(Y_i, Z_i) = O_P(1)$, while $\sup_i n |W_{i,n} - W_{i,n}^*| = O_P(n^{-1/2})$, from Theorem 3.1, which shows that (3.2) holds. Finally, (3.4) can be obtained from Assumption (3.3).

■

3.2 Examples of applications of Theorems 3.1 and 3.2

We now present some applications of the general results of the previous section. The three examples that we give are motivated by survival copula models. Indeed, the bivariate survival function $S_F(t, u) = \mathbb{P}(T > t, U > u)$ of the random vector (U, V) admits, by Sklar's Theorem (Sklar (1959)), a copula representation, that is

$$S_F(t, u) = C(S_T(t), S_U(u)),$$

where $S_T(t) = \mathbb{P}(T > t)$ and $S_U(u) = \mathbb{P}(U > u)$, and where C is a copula function (see e.g. Nelsen (2006)). To understand the dependence between T and U , which is represented by the copula function C , it is natural to search for an estimator \hat{C} , usually based on a parametric or semiparametric model. Nevertheless, even in this framework, nonparametric inference is required, at least to validate the model.

In section 3.2.1, we show how our estimator can be used to compute a $n^{1/2}$ -consistent estimator of S_F . Section 3.2.2 illustrates its behaviour in view of estimating Kendall's τ coefficient. This nonparametric estimation of τ can be used, in a copula modelling approach, to obtain preliminary estimators that can be used in the maximum likelihood estimation procedure. Finally, section 3.2.3 provides an uniform consistency result for an extension of Kendall's process, that is classically used to assess the validity of Archimedean copula models in absence of right-censoring and left-truncation.

3.2.1 Estimation of the joint distribution function

As a particular example of the use of Theorem 3.2, we provide an uniform consistency result for the estimation of the joint distribution function (T, U) . The survival function S_F can be obtained in the same way, since both function are related. Indeed, we can apply Theorem 3.2 to the class of functions $\mathcal{F} = \{(T, U) \rightarrow \mathbf{1}_{T \leq t, U \leq u}\}$. This class is a Donsker class of functions (see Van der Vaart and Wellner, 1996, for a complete definition), and therefore satisfies condition (3.3). Moreover, $\hat{F}(t, u) = \int \int \mathbf{1}_{t' \leq t, u' \leq u} d\hat{F}(t', u')$, which leads to

$$\sup_{t, u} |\hat{F}(t, u) - F(t, u)| = O_P(n^{-1/2}). \quad (3.5)$$

The joint survival function can be studied in the same way by considering $\mathbf{1}_{T > t, U > u}$. We see that this convergence rate is of the same order as the one of Dabrowska's estimator (Dabrowska (1988)).

3.2.2 Estimation of Kendall's τ coefficient

Kendall's τ coefficient can be defined as

$$\tau = \mathbb{P}[(T_1 - T_2)(U_1 - U_2) > 0] - \mathbb{P}[(T_1 - T_2)(U_1 - U_2) < 0].$$

This coefficient is a popular measure of dependence between two random variables, see e.g. Nelsen (2006) for more details. After some standard calculations, Kendall's τ coefficient can be rewritten as

$$\tau = 4 \int \int F(t, u) dF(t, u) - 1.$$

Therefore, our estimator \hat{F} can be used to define a nonparametric estimator of τ in the following way,

$$\hat{\tau} = 4 \int \int \hat{F}(t, u) d\hat{F}(t, u) - 1. \quad (3.6)$$

Convergence of $\hat{\tau}$ towards τ can be shown combining (3.5) and Theorem 3.2. Indeed,

$$\begin{aligned} \hat{\tau} - \tau &= \int \int (\hat{F}(t, u) - F(t, u)) dF(t, u) + \int \int F(t, u) d\{\hat{F}(t, u) - F(t, u)\} \\ &\quad + \int \int (F(t, u) - \hat{F}(t, u)) d\{\hat{F}(t, u) - F(t, u)\}. \end{aligned}$$

The first and third term are $O_P(n^{-1/2})$ using (3.5), while the second is also $O_P(n^{-1/2})$ from Theorem 3.2.

In the following, we will use this estimator $\hat{\tau}$ to define preliminary estimators for the association parameter of survival copulas models under bivariate censoring and left truncation. Indeed, for some copula families, there exists a closed formula linking the parameter of association to Kendall's τ coefficient. Otherwise one can use numerical techniques to determine, for example through simulations, which value of the association parameter corresponds to $\hat{\tau}$. The preliminary estimators obtained by this method can then be used as starting point for maximum likelihood optimization. Indeed, these maximization techniques, especially in our incomplete data framework, rely on delicate numerical techniques that can be very sensitive to the initial point of the iterative maximization algorithm that is used. Since our estimator (3.6) is $n^{1/2}$ -consistent, it is expected to furnish a good initialization for such procedures.

Illustration.

To investigate the practical behaviour of our procedure, we consider a simulation study in order to assess the convergence of our estimator $\hat{\tau}$. We consider a couple of

lifetimes (T, U) with the same Weibull distribution for T and U . Weibull distribution is parametrized through a shape parameter α and a scale parameter β , and admits a density

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left(-\frac{t^\alpha}{\beta^\alpha}\right),$$

for $t \geq 0$. The two random variables are linked by a Clayton copula (see definition in Table 2) with parameter θ , so that Kendall's τ coefficient is known, since $\tau = \theta[\theta + 2]^{-1}$ for a Clayton copula. The censoring variables are assumed to be i.i.d. exponential with scale parameter equal to one. The truncation variables (μ, ν) are assumed to be i.i.d. with marginal distribution defined as the product $B \times W$ of a Bernoulli random variable B with mean value 0.7, and a Weibull random variable W of parameters $\alpha = 2$ and $\beta = 10$ which is independent from B .

We considered two different situations for the values of the parameters of the marginal laws of T and U . In the first situation, we consider $\alpha = 2$ and $\beta = 1.1$, which corresponds approximatively to a proportion of 20% of observations which are uncensored for both marginals. Additionally, we considered the case where $\alpha = 2$ and $\beta = 1.7$, corresponding to 10% of observations which are uncensored for both marginals. In each case, we took $\theta = 2$, corresponding to $\tau = 0.5$, and we considered two sample sizes $n = 1000$, and $n = 2000$. Through 1000 replications of each setting, we computed an estimation of the mean squared error $E[(\hat{\tau} - \tau)^2]$, and decomposed it into a bias term and a variance term. Results are shown in Table 1 below.

Model	Criterion	n=1000	n=2000
$\alpha = 2$ $\beta = 1.1$ (20% of uncensored)	MSE =	0.01502	0.00722
	Bias =	0.08983	0.05117
	Variance =	0.00695	0.00460
$\alpha = 2$ $\beta = 1.7$ (10% of uncensored)	MSE =	0.07433	0.04051
	Bias =	0.23792	0.16787
	Variance =	0.01772	0.01233

Table 1: Estimation of the mean-squared error and related quantities for the estimation of Kendall's τ coefficient.

3.2.3 Kendall's process

Kendall's process is a standard tool to perform model checking for copulas in absence of censoring or truncation, see Genest and Rivest (1993), or Genest et al. (2006). Applications will be shown in section 4.3. Defining $K(v) = \mathbb{P}(S_F(T, U) \leq v)$, empirical versions of K can be computed in absence of censoring using the empirical distribution function. In our framework, we can extend this definition, defining

$$\hat{K}(v) = \int \int \mathbf{1}_{S_{\hat{F}}(t,u) \leq v} d\hat{F}(t, u) = \sum_{i=1}^n W_{i,n} \mathbf{1}_{S_{\hat{F}}(Y_i, Z_i) \leq v}. \quad (3.7)$$

The proof of the next result is technical and is postponed to the appendix section.

Proposition 3.3 *Under the assumptions of Theorem 3.1, assume that the density of (T, U) is uniformly bounded by some constant $\mathcal{C} \geq 0$. Then,*

$$\sup_{v \in [0,1]} |\hat{K}(v) - K(v)| = O_P(n^{-1/2}).$$

This theoretical result is a key tool to assess the validity of testing procedures for copula models. More precisions on this are given in section 4.3. Wang and Wells (2000) proposed a similar estimator based on Dabrowska's estimator, without proving its convergence.

3.3 Bootstrap procedure

Theorem 2.2 only provides an uniform convergence rate, without giving an asymptotic law, which could be used to compute confidence intervals for estimations of integrals $\int \int \phi(t, u) dF(t, u)$. Nevertheless, the most appropriate way to estimate the asymptotic variance of such quantities is to rely on bootstrap approximations (even if we had determined the asymptotic law, bootstrap procedures would be more adapted for finite sample size).

Suppose that we wish to estimate the law of a quantity $\theta(\hat{F})$, the particular example we have in mind being $\text{Var} \left(\int \int \phi(t, u) d\hat{F}(t, u) \right)$, but more general quantities may be studied with the similar method. For censored variables in the univariate, there exists two principal strategies for defining resampling plans. Reid's (Reid (1981)) methodology consists of directly resampling the points $(Y_i, \delta_i)_{1 \leq i \leq n}$, while Efron's methodology (Efron (1981)) consists of simulating lifetimes and censoring variables to compute (Y_i, δ_i) . Although Reid's methodology is easier to implement, it can lead to unconsistant results,

which is not the case of Efron's methodology, see Akritas (1986). Therefore, we try to extend this second approach to the bivariate case (in the univariate case, Efron's resampling plan has already been extended to the case of left truncation by Wang (1991)).

In order to simulate censoring variables (C, D) , it is required to use an estimator of their joint distribution. Defining $G(c, d) = \mathbb{P}(C \leq c, D \leq d)$, we can, reversing the role of the censoring and lifetime variable, compute an estimator \hat{G} using the same method that we used for estimating the distribution of the lifetimes. We also need to simulate truncation variables (μ, ν) . For this, we use

$$\hat{L}(t, u) = \left(\sum_{i=1}^n \hat{S}_F(Y_i, Z_i)^{-1} \right)^{-1} \left(\sum_{i=1}^n \hat{S}_F(Y_i, Z_i)^{-1} \mathbf{1}_{\mu_i \leq t, \nu_i \leq u} \right),$$

as an estimator of their joint distribution.

We here provide a resampling plan, inspired by the methodology of Efron (1981) for censored observations in the univariate case.

Resampling plan.

Let B denote the number of bootstrap sample that we wish to generate. For $b = 1, \dots, B$,

1. Simulate (T_i^b, U_i^b) according to \hat{F} .
2. Simulate (C_i^b, D_i^b) according to \hat{G} .
3. Compute $(Y_i^b, Z_i^b, \delta_i^b, \gamma_i^b)$.
4. Simulate (τ_i^b, v_i^b) according to \hat{L} .
5. If $Y_i^b \geq \tau_i^b$ and $Z_i^b \geq v_i^b$, keep observation i . Otherwise, resimulate from Step 1 until this is the case.
6. Obtain an n -sample reiterating this procedure independently, and compute \hat{F}_b and $\theta(\hat{F}_b)$.

The vector $(\theta(\hat{F}_b))_{1 \leq b \leq B}$ can then be used to estimate different characteristics, such as the variance, of $\theta(\hat{F})$.

Remark 3.4 *Simulating random vector with respect to discrete distributions such as \hat{F} , \hat{G} and \hat{L} is easy. We here see one of the advantages of having used an estimator \hat{F} which is linked to a true distribution. If we use estimators such as the one provided by*

Dabrowska (1984), this resampling plan can not be adapted since Step 1 and 2 are not feasible. Nevertheless, there is a specificity of \hat{F} and \hat{G} that we must take into account: in some cases, these two distributions will present some mass at (∞, ∞) . In these cases, some observations which are resimulated should have the value (∞, ∞) . The probability occurrence of such events should be small (both distributions put a small mass at infinity). Such observations will only have the consequence to add some weight at (∞, ∞) for \hat{F}_b , without leading to failures. However, an alternative would be to remove these observations from the simulated sample, which would not strongly modify the results.

4 Application to the Canadian data set

In this section, we apply our technique to a portfolio from a Canadian insurer, which has been studied in several papers. In this paper, the authors provided a very detailed study of the joint mortality using Archimedean copula models. The choice of the copula family is made using a technique based on a version of Kendall's process proposed by Wang and Wells (2000) and based on Dabrowska's estimator. We here compare our results with the new estimator that we propose in order to assess its validity. Moreover, in section 4.3, we provide a test procedure to perform model-checking, and explain how to compute the critical values.

4.1 Description of the data-set

The dataset that we study has been initially introduced by Frees et al. (1996), and was studied by Carriere (2000), Youn and Shemyakin (1999), Youn and Shemyakin (2001) and Luciano et al. (2008). We refer to Frees et al. (1996) for a more detailed description of this dataset. The dataset concerns 14947 contracts from a large Canadian insurer, observed between Decembre 29th, 1988 and Decembre 31th, 1993. After elimination of same-sex contracts and of couples with more than one policies (for which we only keep one policy), 11454 contracts remain. To model the data, we consider three copula models proposed by Luciano et al. (2008) to study a particular generation (here, we apply our method to the whole population), and show how our technique allows to perform estimation and goodness-of-fit.

4.2 Copula models and estimation of the association parameter

In this section, we consider a semiparametric copula model for describing the distribution of (T, U) . In a semiparametric copula model, the marginal distribution of each variable is estimated nonparametrically (using Kaplan-Meier estimator since we are in presence of censoring and truncation), while the dependence between the two variables is described by a parametric copula family $\{C_\theta : \theta \in \Theta\}$ for $\Theta \subset \mathbb{R}^k$, see e.g. Shih and Louis (1995). We consider the three families of copulas that have been used by Luciano et al. (2008) for studying the joint mortality of a specific generation in the Canadian data-set, that is Clayton, Frank and Nelsen 4.2.20. These three families are of Archimedean type, which means that the copula function is uniquely defined by a decreasing convex function ϕ , that is $C_\phi(t, u) = \phi^{-1}(\phi(u) + \phi(v))$. The generators corresponding to the three families are presented in Table 2 below, and we recall the link between τ and the association parameter in these families.

Using our estimator (3.6), we obtain $\hat{\tau} = 0.4332$. Using the relationship between the association parameter θ and τ , we obtain the following preliminary estimators of θ , presented in Table 2.

Model	$\phi_\theta(t)$	τ	Preliminary
Clayton	$\theta^{-1}(t^{-\theta} - 1)$	$\tau = \frac{\theta}{\theta+2}$	1.5283
Frank	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$1 - \frac{4}{\theta} \left(\int_0^\theta \frac{t dt}{\theta(e^t-1)} - 1 \right)$	4.6201
4.2.20	$\exp(t^{-\theta}) - e$	$1 - \frac{4}{\theta} \left(\frac{1}{\theta+2} - \int_0^\theta t^{\theta+1} [\exp(1-t^{-\theta})] dt \right)$	1.0413

Table 2: Expression of the different copula families considered. The column τ presents the expression of Kendall's τ if available, the column preliminary represents the preliminary estimator obtained by calibration with τ coefficient.

These preliminary estimators are then used to perform maximum likelihood estimation, using a procedure which has been studied by Shih and Louis (1995), Wang and Wells (2000), and which is also used in Luciano et al. (2008). In this bivariate censoring framework, the log-likelihood can be decomposed into four terms, that is

$$\begin{aligned}
 L(\theta) = & \sum_{i=1}^n \delta_i \gamma_i \log \left(c_\theta(\hat{S}_T(Y_i), \hat{S}_U(Z_i)) \right) + (1 - \delta_i) \gamma_i \log \left(\partial_z C_\theta(\hat{S}_T(Y_i), \hat{S}_U(Z_i)) \right) \\
 & + \delta_i (1 - \gamma_i) \log \left(\partial_y C_\theta(\hat{S}_T(Y_i), \hat{S}_U(Z_i)) \right) + (1 - \delta_i) (1 - \gamma_i) C_\theta(\hat{S}_T(Y_i), \hat{S}_U(Z_i)),
 \end{aligned}$$

Model	Final estimator	Corresponding value of Kendall's τ
Clayton	1.3116	0.3960
Frank	2.91290	0.3012
Nelsen 4.2.20	0.5098	0.4020

Table 3: Final maximum likelihood estimators of the association parameter in the three copula models, with corresponding value of Kendall's τ coefficient.

where \hat{S}_T (resp. \hat{S}_U) denotes the Kaplan-Meier estimator of $S_T(t) = \mathbb{P}(T > t)$ (resp. $S_U(u) = \mathbb{P}(U > u)$), or any other consistent estimator of the survival function. We perform maximization of this quantity using a Newton-Raphson algorithm, with initial value taken as the preliminary estimator of Table 2. The final estimators are presented in Table 3, along with the corresponding value of Kendall's τ coefficient.

4.3 A class of model-checking procedures using Kendall's process

4.3.1 Test-statistics

Archimedean families of copulas present the advantage of being uniquely defined by their generator ϕ , which can be linked to function K through the relation,

$$K(v) = v - \frac{\phi(v)}{\phi'(v)}.$$

Genest and Rivest (1993) and Genest et al. (2006) proposed test procedures (in absence or censoring) that can be used for model-checking based on the comparison of $K_{\hat{\theta}}(v) = v - \phi_{\hat{\theta}}(v)/\phi'_{\hat{\theta}}(v)$, where $\hat{\theta}$ denotes the estimator of the association parameter in a parametric copula model, and a nonparametric version \hat{K} .

Let C denote the true copula function linking the marginal distributions of T and U . If one wishes to test $H_0 : C \in \mathcal{C}_{\Theta}$ against $H_1 : C \notin \mathcal{C}_{\Theta}$, where \mathcal{C}_{Θ} is a parametric family of copulas, the procedure can be summarized as follows:

1. Estimate θ by $\hat{\theta}$ assuming that model \mathcal{C}_{Θ} is true.
2. Compute \hat{K} , $K_{\hat{\theta}}$, and $d(\hat{K}, K_{\hat{\theta}})$ where d denotes some distance.
3. If $d(\hat{K}, K_{\hat{\theta}})$ is larger than some critical value s_{α} , reject H_0 .

Luciano, Vigna and Spreuw (2008) use a L^2 -distance, that is

$$d(K, K') = \int_0^1 \{K(v) - K'(v)\}^2 dv, \quad (4.1)$$

and we will use the same distance for easier comparison. Next, we explain in section 4.3.2 how to compute the critical values.

Based on Proposition 3.3, one can see that this testing approach is consistent. Indeed, if H_0 is not true, our estimator \hat{K} converges towards some limit K such that $\inf_{\theta \in \Theta} d(K, K_\theta) > 0$. Hence, since $d(\hat{K}, K_{\hat{\theta}}) > |d(K, K_{\hat{\theta}}) - d(\hat{K}, K)|$, we get

$$d(\hat{K}, K_{\hat{\theta}}) > \inf_{\theta \in \Theta} d(K, K_\theta) + O_P(n^{-1/2}).$$

As a consequence, for any α , $\mathbb{P}\left(d(\hat{K}, K_{\hat{\theta}}) > s_\alpha | H_1\right) \rightarrow_{n \rightarrow \infty} 1$.

4.3.2 Computation of the critical values

We propose to compute the critical values using a procedure which is adapted from Genest et al. (2006) in the uncensored case. The critical values are computed using a bootstrap resampling plan. Let us mention that Wang and Wells (2000), in our framework, proposed a very simple resampling plan. However, Genest et al. (2006) showed that this resampling plan was inconsistent and therefore could not be used. Therefore, we propose to adapt our bootstrap procedure defined in section 3.3 in order to extend the procedure of Genest et al. (2006). We apply this technique for $\theta(\hat{F}) = d(\hat{K}, K_{\hat{\theta}})$, where d is the distance defined in (4.1). However, since we are interesting in knowing the distribution of $d(\hat{K}, K_{\hat{\theta}})$ under the null hypothesis, we simply change Step 1 in our bootstrap procedure of section 3.3 by

- 1'. Simulate (T, U) according to the distribution $C_{\hat{\theta}}(\hat{F}_T(t), \hat{G}_U(u))$.

The bootstrap vector $(d(\hat{K}^b, K_{\hat{\theta}^b}))$ that we obtain using this procedure can then be used to determine approximated p -values of our test procedure.

Another modification of the bootstrap procedure is required to reduce computation time. Indeed, this simulation scheme suffers from the difficulty to estimate the distribution in the left-tail (caused by truncation). Even if condition (3.1) holds, the constant K may be large and impact on the performance for finite sample-size (this is actually the case in the univariate situation). In practice, we see that the nonparametric estimator of the distribution of (C, D) is very sensitive to this phenomenon. As an illustration, the smallest censored observations (for both components) has mass 0.5. Moreover, the set of observations such that one at least is less than 20000 days is affected of a mass greater than 0.8, although it represents only 2% of the observations. This is not a problem from a theoretical point of view: if we draw some point (C, D) in this part of the distribution, the truncation phenomenon will frequently reject this observation, and therefore it will not

appear in the final simulated sample. However, in practice, according to our procedure, each time a point is rejected due to truncation, another simulation is performed. If these rejection happen very often, this dramatically increases the computation time. Therefore, we chose to use modify the distribution. We affect to the set of censored points such that $C < 20000$ or $D < 20000$ a mass 0.2, and renormalize the rest of the mass in order to obtain a total mass equal to 1.

4.3.3 Goodness-of-fit on the Canadian data-set

In Figure 1, we graphically present the function \hat{K} computed from the Canadian data-set, and compare it with the function \hat{K}_θ obtained for the three copula models.

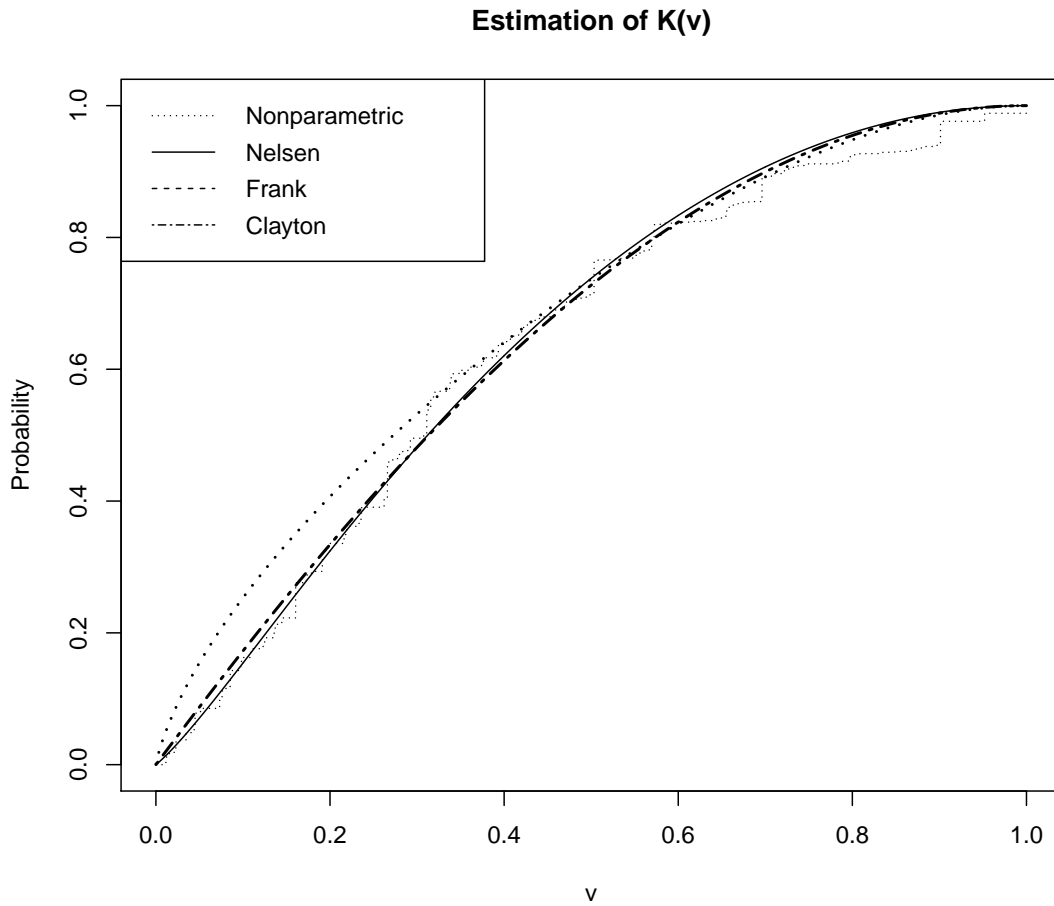


Figure 1: Comparison of the estimators \hat{K}_θ to the nonparametric estimator \hat{K} .

One can observe that the estimated version using Clayton's model or Nelsen's 4.2.20

model seem to be closer to the nonparametric curve. Nevertheless, all three models seem to be less accurate for high values of v , which may be caused by the presence of censoring. The application of the test procedure described in the previous section is presented in Table 4. In this table, we compare the value of the distance obtained by our method for the different models that we considered, to the one obtained by Luciano et al. (2008) using the version of \hat{K} proposed by Wang and Wells (2000) instead of our estimator \hat{K} defined in (3.7). We present the p -value corresponding to each case.

Model	Test statistic	95% quantile	97.5 % quantile	99 % quantile	p-value
Clayton	0.006360	0.02301	0.02708	0.03383	0.3840
Frank	0.02439	0.02893	0.02980	0.03975	0.1520
Nelsen 4.2.20	0.006253	0.02866	0.02972	0.03571	0.3630

Table 4: Goodness-of-fit procedure for a the three survival copula models considered (Clayton, Frank, Nelsen 4.2.20).

According to these results, Frank’s model corresponds to the smallest p-value, and seems less adapted than the two other models, which corresponds to what was observed on Figure 1. At a classical 5% level, one can see that, for all three models, the null hypothesis is not rejected.

5 Conclusion

In this paper, we proposed a new nonparametric estimator of the joint distribution of a random vector of lifetimes (T, U) . This estimator is designed to compensate the classical problem that impact lifetime data, such as left truncation and right-censoring. Unlike previously existing estimators in this framework, our estimator defines a true distribution, and does not only focus on the particular case of estimating the joint survival function. An important advantage of this approach, is that it allows to perform simulations quite easily, as it has been proposed in the bootstrap procedures that we discussed in section 3.3 and 4.3.2. These simulations procedure may be quite valuable in the computation of critical values of goodness-of-fit tests, as it is illustrated on the Canadian data-set of Frees et al. (1996).

6 Appendix

6.1 Proof of Theorem 3.1

In this proof, we introduce several notations in order to make some simplifications. We will denote $H_i = H(Y_i, Z_i)$, $\Delta_i = \delta_i \gamma_i$, $S_i = S_F(Y_i, Z_i)$, and $S_i^* = 1 - F^*(Y_i, Z_i) = n^{-1} \sum_{j=1}^n \Delta_j H_j^{-1} \mathbf{1}_{Y_j > Y_i, Z_j > Z_i}$. The vector W^* will represent the weights affected by F^* to each observation. Only the doubly uncensored observations (which are the one where some mass is put) are present in this vector of size $m+1$ (as in the proof of Theorem 2.2, m will denote the number of doubly uncensored observations) This vector is sorted accordingly to Y (the first component corresponds to the mass assigned at the doubly uncensored observation with the smallest Y , and so on). W_i^* will denote the i -th component of W^* .

We recall some facts that can be easily proven using empirical process theory, see e.g. Van der Vaart and Wellner (1996) for more details:

$$\sup_{t,z} |\hat{H}(t, z) - H(t, z)| = O_P(n^{-1/2}), \quad (6.1)$$

$$\sup_{t,z} |F^*(t, z) - F(t, z)| = O_P(n^{-1/2}). \quad (6.2)$$

Now, let us recall that

$$W_{\infty, \infty}^* = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i S_i}{H_i},$$

that is the residual mass that the ideal estimator F^* puts at infinity. $W_{\infty, \infty}^*$ is a sum of i.i.d. quantities, with $E[W_{\infty, \infty}^*] = 0$ (from Proposition 2.1), and with finite variance. Then, the CLT applies, leading to $W_{\infty, \infty}^* = O_P(n^{-1/2})$.

Next, as in the proof of Theorem 2.2, we remove the components of W which are supposed to be zero, considering the equation $(I_{m+1} - \tilde{A})\tilde{W} = 0$, where m denotes the number of doubly uncensored responses. We will also assume that the components of \tilde{W} are sorted with respect to the variable Y . Define W' as a solution of $(I_{m+1} - \tilde{A})W' = 0$ with $W'_\infty = W_\infty^*$. Note that the vector W' does not necessarily define a probability measure, since the some of its components is not equal to one. However, W' is proportional to W , and defines a positive measure. In this proof, we first show that W' is close to W^* , and then show that W is close to W' .

We denote $\mathcal{S}_j = \sum_{i \geq j} |W'_i - W_i^*| + |W'_{\infty, \infty} - W_{\infty, \infty}^*|$, with $\mathcal{S}_{m+1} = |W'_{\infty, \infty} - W_{\infty, \infty}^*| = 0$. Let R_i be defined as in the proof of Theorem 2.2. We have, for $j \leq m$, $0 \leq \mathcal{S}_j \leq$

$\mathcal{S}_{j+1} + |W'_j - W_j^*|$. Next, observe that

$$W'_j = \frac{1}{n} \frac{\sum_{i>j} W'_i \mathbf{1}_{i \in R_j}}{\hat{H}_j} = \frac{1}{n} \frac{\sum_{i>j} W'_i \mathbf{1}_{i \in R_j}}{H_j} + \frac{1}{n} \frac{\sum_{i>j} W'_i \mathbf{1}_{i \in R_j} [H_j - \hat{H}_j]}{\hat{H}_j H_j}, \quad (6.3)$$

$$W_j^* = \frac{1}{n} \frac{S_j}{H_j} = \frac{1}{n} \frac{\sum_{i>j} W_i^* \mathbf{1}_{i \in R_j}}{H_j} + \frac{1}{n} \frac{S_j - S_j^*}{H_j}. \quad (6.4)$$

Therefore,

$$\mathcal{S}_j \leq \left(1 + \frac{K}{n}\right) \mathcal{S}_{j+1} + \frac{K^2}{n} \sup_j \frac{H_j |\hat{H}_j - H_j|}{\hat{H}_j} + \frac{K}{n} \sup_j |S_j^* - S_j|.$$

Since $\mathcal{S}_{m+1} = 0$, one gets

$$\mathcal{S}_j \leq \varepsilon_n \frac{(1 + Kn^{-1})^{m-j} - 1}{Kn^{-1}} \leq C(K) \varepsilon_n,$$

where

$$\varepsilon_n = \frac{K^2}{n} \sup_j \frac{H_j |\hat{H}_j - H_j|}{\hat{H}_j} + \frac{K}{n} \sup_j |S_j^* - S_j|,$$

and where $C(K)$ is a positive constant which does not depend from n or j . Moreover, on A_n we have

$$\sup_j \frac{H_j |\hat{H}_j - H_j|}{\hat{H}_j} \leq \sup_{t,u} |\hat{H}(t,u) - H(t,u)| \times \sup_{t,u} \left| \frac{H(t,u)}{\hat{H}(t,u)} \right| = O_P(n^{-1/2}),$$

from (6.1). Indeed, the first supremum in the right-hand side is exactly the supremum considered in (6.1). To see that $\sup_{t,u} \left| \frac{H(t,u)}{\hat{H}(t,u)} \right| = O_P(1)$, we can use the uniform consistency of \hat{H} coming from (6.1) to see that, for n large enough, on a set with probability tending to one, $|\hat{H}(t,u) - H(t,u)| \leq 1/2K$, and therefore $H(t,u)/\hat{H}(t,u) \leq 1/[1 - 1/2]$. Next, recall that

$$\sup_j |S_j^* - S_j| \leq \sup_{t,u} |S^*(t,u) - S(t,u)| = O_P(n^{-1/2}),$$

from (6.2). This leads to $\sup_j \mathcal{S}_j = O_P(n^{-1/2})$.

Going back to (6.3) and (6.4), we see that

$$|W'_j - W_j^*| \leq \frac{1}{n} \sup_j \mathcal{S}_j + \frac{C(K)}{n} \varepsilon_n := \frac{\eta_n}{n},$$

with $\eta_n = O_P(n^{-1/2})$. From this, we deduce that $\sum_{j=1}^n W'_j = \sum_{j=1}^n W_j^* + r_n = 1 + r_n$, with $r_n = O_P(n^{-1/2})$. Next, we use the fact that $W_j = W'_j [\sum_{j=1}^n W'_j]^{-1}$ to deduce that

$$|W_j - W_j^*| \leq \frac{|W'_j - W_j^*|}{1 + r_n} + W_j^* r_n = \frac{\theta_n}{n},$$

with $\theta_n = O_P(n^{-1/2})$, since $W_j^* \leq Kn^{-1}$.

6.2 Proof of Proposition 3.3

Consider the event $A_{n,\varepsilon} = \{\sup_{t,u} n^{1/2} |S_{\hat{F}}(t, u) - S_F(t, u)| \leq \varepsilon\}$, for some $\varepsilon > 0$. By (3.5) and the definition of the $O_P(n^{-1/2})$ convergence rate, $\lim_{x \rightarrow \infty} \bar{\lim}_n \mathbb{P}(A_{n,\varepsilon}^c) = 0$. On this set, $\Delta_i := |\mathbf{1}_{S_{\hat{F}}(Y_i, Z_i) \leq v} - \mathbf{1}_{S_F(Y_i, Z_i) \leq v}|$ is equal to zero, except if $|S_F(Y_i, Z_i) - v| \leq \varepsilon n^{-1/2}$. Hence, it is bounded by $\mathbf{1}_{|S_F(Y_i, Z_i) - v| \leq \varepsilon n^{-1/2}}$. Define

$$\tilde{K}(v) = \int \int \mathbf{1}_{S_F(t,u) \leq v} d\hat{F}(t, u) = \sum_{i=1}^n W_{i,n} \mathbf{1}_{S_F(Y_i, Z_i) \leq v}.$$

We see that, on $A_{n,\varepsilon n^{-1/2}}$,

$$|\hat{K}(v) - \tilde{K}(v)| \leq \int \int \mathbf{1}_{|S_F(t,u) - v| \leq \varepsilon n^{-1/2}} d\hat{F}(t, u).$$

Now, obviously the class $\mathcal{F} = \{(T, U) \rightarrow \mathbf{1}_{|S_F(T,U) - v| \leq \eta}, v \in [0, 1], \eta > 0\}$ satisfies condition (3.3) (it is a Donsker class of bounded functions, see van der Vaart and Wellner (1996)). Hence, Theorem 3.2 applies. Therefore, on $A_{n,\varepsilon}$

$$|\hat{K}(v) - \tilde{K}(v)| \leq \mathbb{P}(|S_F(T, U) - v| \leq \varepsilon n^{-1/2}) + R_n(\varepsilon, v), \quad (6.5)$$

with $\sup_{\varepsilon, v} |R_n(\varepsilon, v)| = O_P(n^{-1/2})$. Since we assumed that the density of (T, U) is bounded by some constant \mathcal{C} , the probability in the right-hand side of (6.5) is less than $\mathcal{C}\varepsilon n^{-1/2}$.

We can deduce that

$$\mathbb{P}\left(\sup_v n^{1/2} |\hat{K}(v) - \tilde{K}(v)| > x\right) \leq \mathbb{P}\left(\sup_v n^{1/2} |\hat{K}(v) - \tilde{K}(v)| > x, A_{n,\varepsilon}\right) + \mathbb{P}(A_{n,\varepsilon}^c). \quad (6.6)$$

Using (6.5), we see that, for x large enough, the first probability on the right-hand side of (6.6) is less than $\mathbb{P}(\sup_{\varepsilon, v} n^{1/2} |R_n(\varepsilon, v)| > x)$. By definition of the $O_P(n^{-1/2})$ rate, $\lim_{x \rightarrow \infty} \bar{\lim}_n \mathbb{P}(\sup_{\varepsilon, v} n^{1/2} |R_n(\varepsilon, v)| > x) = 0$, and consequently $\sup_v |\hat{K}(v) - \tilde{K}(v)| = O_P(n^{-1/2})$. Finally, Theorem 3.2 applies to \tilde{K} , leading to $\sup_v |\tilde{K}(v) - K(v)| = O_P(n^{-1/2})$, completing the proof.

References

- Akritis, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.*, 81(396):1032–1038.
- Akritis, 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.

- 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.
- Carriere, J. F. (2000). Bivariate survival models for coupled lives. *Scand. Actuar. J.*, (1):17–32.
- Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, 16(4):1475–1489.
- Du, Y. and Akritas, M. G. (2002). Uniform strong representation of the conditional Kaplan-Meier process. *Math. Methods Statist.*, 11(2):152–182.
- Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, 76(374):312–319.
- Frees, E. W., Carriere, J., and Emiliano, V. (1996). Annuity valuation with dependent mortality. *Journal of Risk and Insurance*, 63(2):229–261.
- Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.*, 33(2):337–366.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.*, 88(423):1034–1043.
- 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. Poincaré Probab. Statist.*, 31:547–597.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 53:457–481.
- Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Communications in Statistics: Theory and Methods*, 40(15):2639–2660.
- Lopez, O. and Saint Pierre, P. (2011). Bivariate censored regression relying on a new estimator of the joint distribution function. *Preprint*.

- Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance Math. Econom.*, 43(2):234–244.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.
- 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.
- Pruitt, R. C. (1993). Identifiability of bivariate survival curves from censored data. *J. Amer. Statist. Assoc.*, 88(422):573–579.
- Reid, N. (1981). Estimating the median survival time. *Biometrika*, 68(3):601–608.
- Sánchez Sellero, C., González Manteiga, W., and Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist.*, 32(4):563–581.
- Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.*, 23(2):422–439.
- Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4):883–886.
- 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.
- Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.*, 27(5):1745–1784.
- Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86(413):pp. 130–143.
- Wang, W. and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.*, 95(449):62–76. With a comment by Edsel A. Peña and a rejoinder by the authors.
- Youn, H. and Shemyakin, A. (1999). Statistical aspects of joint life insurance pricing. *Proceedings of the business and statistics section of the American Statistical Association*, pages 34–38.
- Youn, H. and Shemyakin, A. (2001). Pricing practices for joint last survivor insurance. *Actuarial Research Clearing House*.