Universidade de Vigo

A Semiparametric Estimator of the Bivariate Distribution Function for Censored Gap Times

Jacobo de Uña Álvarez and Ana Paula Amorim

Report 09/03

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://eioweb.uvigo.es/
E-mail: depc05@uvigo.es

Universidade de Vigo

A Semiparametric Estimator of the Bivariate Distribution Function for Censored Gap Times

Jacobo de Uña Álvarez and Ana Paula Amorim

Report 09/03

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita: Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

A SEMIPARAMETRIC ESTIMATOR OF THE BIVARIATE DISTRIBUTION FUNCTION FOR CENSORED GAP TIMES

Jacobo de Uña-Álvarez[1,*] and Ana Paula Amorim[2]

November 2009

[1]Department of Statistics and OR, University of Vigo

[2]Department of Mathematics, University of Minho

[*]Corresponding author. *Full postal address:* Departamento de Estadística e Investigación Operativa, Facultad de CC. Económicas y Empresariales, Universidad de Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain. *Phone:* (+34) 986812492. *Fax:* (+34) 986812401. *E-mail:* jacobo@uvigo.es

Abstract

Let (T_1, T_2) be gap times corresponding to two consecutive events, which are observed subject to random right-censoring. In this paper a semiparametric estimator of the bivariate distribution function of (T_1, T_2) and, more generally, of a functional $E[\varphi(T_1, T_2)]$ is proposed. We assume that the probability of censoring for T_2 given the (possibly censored) gap times belongs to a parametric family of binary regression curves. We investigate the conditions under which the introduced estimator is consistent. We explore the finite sample behavior of the estimator through simulations. The main conclusion of this paper is that the semiparametric estimator may be much more efficient than purely nonparametric methods. Real data illustration is included.

Key Words and Phrases: bivariate censoring, Kaplan-Meier, presmoothing, recurrent events, semiparametric censorship model

1 Introduction

The statistical analysis of consecutive gap times is an issue of much importance in a number of fields, including engineering, economy, epidemiology, and survival analysis. Most of the times, one will be interested in describing not only the marginal distribution of the gap times but also the correlation structure among them. This happens, for example, when analyzing recurrent event data, which arise when each individual may go through a well-defined event several times along his history. Then, the inter-event times are referred to as the gap times, and they are of course determined by the times at which the recurrences take place (i.e. the recurrence times). See Cook and Lawless (2007) for an up-to-date revision of statistical methods for recurrent event data. In this paper, the interest is focused on a given couple of (successive) gap times. In our real data example in Section 4, these will be the time up to first recurrence and the time from first to second recurrence for bladder cancer patients. In order to formalize the discussion, we now introduce our notation.

Let (T_1, T_2) be a pair of gap times of successive events, which are observed subject to random right-censoring. Let C be the right-censoring variable, assumed to be independent of (T_1, T_2), and let $Y = T_1 + T_2$ be the total time. Due to censoring, rather than (T_1, T_2) we observe $\left(\widetilde{T}_1, \widetilde{T}_2, \Delta_1, \Delta_2\right)$, where $\widetilde{T}_1 = T_1 \wedge C$, $\Delta_1 = I(T_1 \leq C)$ and $\widetilde{T}_2 = T_2 \wedge C_2$, $\Delta_2 = I(T_2 \leq C_2)$, where $C_2 = (C - T_1) \, I(T_1 \leq C)$ is the censoring variable for the second gap time. Note that $\Delta_2 = 1$ implies $\Delta_1 = 1$. Hence, $\Delta_2 = \Delta_1 \Delta_2 = I(Y \leq C)$ is the censoring indicator pertaining to the total time. We put $\widetilde{Y} = Y \wedge C$. Let $\left(\widetilde{T}_{1i}, \widetilde{T}_{2i}, \Delta_{1i}, \Delta_{2i}\right)$, $1 \leq i \leq n$, be iid data with the same distribution as $\left(\widetilde{T}_1, \widetilde{T}_2, \Delta_1, \Delta_2\right)$. Since the censoring time is assumed to be independent of the process, the marginal distribution of the first gap time T_1 may be consistently estimated by the Kaplan-Meier estimator based on the $\left(\widetilde{T}_{1i}, \Delta_{1i}\right)$'s. Similarly, the distribution of the total time may be consistently estimated by the Kaplan-Meier estimator based on the $\left(\widetilde{T}_{1i} + \widetilde{T}_{2i}, \Delta_{2i}\right)$'s. However, T_2 and C_2 will be in general dependent (because the expected correlation between the gap times), and hence the estimation of the marginal distribution of the second gap time is not such a simple issue. Also, it is not clear in principle how the bivariate distribution function $F_{12}(x, y) = P(T_1 \leq x, T_2 \leq y)$ can be efficiently estimated. This issue was investigated, among others, by Wang and Wells (1998), Lin et al. (1999), Wang and Chang (1999), Peña *et al.* (2001), van der Laan *et al.* (2002), Schaubel and Cai (2004), van Keilegom (2004), or de Uña-Álvarez and Meira-Machado (2008).

In this paper we propose a semiparametric estimator for the bivariate distribution function of the gap times, $F_{12}(x, y)$. For this, we assume that the probability of censoring for T_2 given the (possibly censored) gap times belongs to a parametric family of binary regression curves. That is, letting $m(x, y) = P(\Delta_2 = 1 | \widetilde{T}_1 = x, \widetilde{Y} = y)$, it is assumed that $m(x, y)$ follows some parametric model. In Section 2 we will see that, in essence, this implies assuming a parametric (smooth) model for $m_1(x, y) = P(\Delta_2 = 1 | \widetilde{T}_1 = x, \widetilde{Y} = y, \Delta_1 = 1)$. Note that, since $\widetilde{T}_1$, $\widetilde{Y}$, Δ_1, and Δ_2 are observed, this assumption is testable in practice. On the basis of this parametric assumption, we are able to introduce a new estimator. Basically, the new method uses a presmoothed version of the Kaplan-Meier estimator (see e.g. Dikta, 1998) pertaining to the distribution of the total time (the Y) to weight the bivariate data. In the limit case of no presmoothing, the estimator we propose reduces to that in de Uña-Álvarez and Meira-Machado (2008), which was shown to have nice properties. However, the introduction of parametric presmoothing may greatly reduce the variance in the estimation, particularly at the right tail of the (bivariate) distribution or for heavy censoring on T_2. This will become clear below.

The idea of presmoothing the Kaplan-Meier estimator through a parametric model goes back to Dikta (1998), who termed this method as 'semiparametric censorship modeling'. See also Dikta (2000, 2001) and Dikta et al. (2005). Parametric presmoothing with covariates was considered by de Uña-Álvarez and Rodríguez-Campos (2004), Yuan (2005), or Iglesias-Pérez and de Uña-Álvarez (2008). All these references conclude that the presmoothed (semiparametric) estimators have improved variance when compared to purely nonparametric estimators. In this paper we adapt the main ideas in de Uña-Álvarez and Rodríguez-Campos (2004) to the case in which the 'covariate' is the first gap time, which can be eventually censored.

The paper is organized as follows. In Section 2 we introduce the new semiparametric estimator for the joint distribution function of (T_1, T_2) and we establish its consistency. More generally, we prove consistency for an estimator of a functional of the form $S(\varphi) = E[\varphi(T_1, T_2)]$, where φ is a given transformation of the vector of gap times. Note that in the case in which φ is the indicator of the event $\{T_1 \leq x, T_2 \leq y\}$, the expectation $S(\varphi)$ reduces to $F_{12}(x, y)$. In Section 3 we investigate the finite sample performance of the semiparametric estimator of $F_{12}(x, y)$ in a simulated scenario, while Section 4 is devoted to the analysis of real medical data. Main conclusions and some final remarks are reported in Section 5. The technical proofs are collected in the Appendix.

2 The estimator: consistency

Let $\widetilde{Y}_i = \widetilde{T}_{1i} + \widetilde{T}_{2i}$ be the $i-$th recorded total time, and let W_i be the Kaplan-Meier weight attached to $\widetilde{Y}_i$ when estimating the marginal distribution of Y from the $\left(\widetilde{Y}_i, \Delta_{2i}\right)$'s. That is,

$$W_i = \frac{\Delta_{2i}}{n - R_i + 1} \prod_{R_j=1}^{i-1} \left[1 - \frac{\Delta_{2j}}{n - R_j + 1}\right] \qquad \text{where } R_i = Rank(\widetilde{Y}_i),$$

and where, by convention, the ranks of the censored $\widetilde{Y}_i$'s are higher than those for uncensored values in the case of ties. In the uncensored case we have $W_i = n^{-1}$ for each i. In de Uña-Álvarez and Meira-Machado (2008) the following estimator was proposed:

$$\widehat{F}_{12}(x, y) = \sum_{i=1}^{n} W_i I(\widetilde{T}_{1i} \leq x, \widetilde{T}_{2i} \leq y). \tag{1}$$

These authors showed that this estimator is consistent whenever $x + y$ is smaller than the upper bound of the support of the censoring time. In general, one only has (as usual)

$$\lim_{n \to \infty} \widehat{F}_{12}(x, y) = P(T_1 \leq x, T_2 \leq y, T_1 + T_2 \leq \tau_H) \equiv F^0_{12}(x, y),$$

where τ_H is the upper bound of the support of the distribution function H of $\widetilde{Y}$, assumed to be continuous throughout the paper. The estimator (1) was proved to be more efficient than previous estimators, while being more natural at the same time. Indeed, unlike other available estimators, it is an empirical distribution assigning nonnegative mass to each pair of gap times. Note that this estimator only assigns positive mass to those pairs of gap times with both components uncensored. Now we will modify this estimator in order to incorporate the semiparametric information.

Put $m(x,y) = P(\Delta_2 = 1 | \widetilde{T}_1 = x, \widetilde{Y} = y)$, that is, the probability of uncensoring for the total time Y given the observable information on both gap times. Note that this function is only defined for $x \leq y$; indeed, assuming $P(T_2 = 0) = 0$ (which of course holds under continuity), we have $m(x,x) = 0$, since the event $\left\{ \widetilde{T}_1 = \widetilde{Y} \right\}$ corresponds exactly to $\Delta_1 = 0$, and since $\Delta_1 = 0$ implies $\Delta_2 = 0$. This shows the discontinuous nature of the function m, and consequently prevents us from using any smooth fit to this unknown curve. On the other hand, for $x < y$, we obtain $m(x,y) = P(\Delta_2 = 1 | \widetilde{T}_1 = x, \widetilde{Y} = y, \Delta_1 = 1) \equiv m_1(x,y)$, since the event $\Delta_1 = 1$ is superfluous in the presence of $\widetilde{T}_1 < \widetilde{Y}$. Introduce the presmoothed Kaplan-Meier weights through

$$W_i(m) = \frac{m(\widetilde{T}_{1i}, \widetilde{Y}_i)}{n - R_i + 1} \prod_{R_j = 1}^{i-1} \left[1 - \frac{m(\widetilde{T}_{1j}, \widetilde{Y}_j)}{n - R_j + 1} \right],$$

that is, each censoring indicator $\Delta_{2j}, j = 1, ..., i$, in W_i is replaced by the conditional probability $m(\widetilde{T}_{1j}, \widetilde{Y}_j)$. We assume that $m(x,y) = m(x,y;\beta)$ where β is a vector of parameters and

$$m(x,y;\beta) = \begin{cases} 0 & \text{if } x = y \\ m_1(x,y;\beta) & \text{if } x < y \end{cases},$$

and $m_1(.,.;\beta)$ stands for a (smooth) parametric binary regression model (e.g. logistic) for m_1. In practice, β is replaced by some consistent estimator β_n, which typically will be computed by maximizing the conditional likelihood of the Δ_2's given $\left(\widetilde{T}_1, \widetilde{T}_2 \right)$, for those individuals with $\Delta_1 = 1$ (see e.g. Dikta, 1998, 2000). Thus, we introduce the parametrically presmoothed Kaplan-Meier weights as

$$W_i(\beta_n) = \frac{m(\widetilde{T}_{1i}, \widetilde{Y}_i; \beta_n)}{n - R_i + 1} \prod_{R_j = 1}^{i-1} \left[1 - \frac{m(\widetilde{T}_{1j}, \widetilde{Y}_j; \beta_n)}{n - R_j + 1} \right],$$

where $m(x,y;\beta_n) = I(x < y) m_1(x,y;\beta_n)$. Note that this definition of $m(x,y;\beta_n)$ mimics the discontinuous behavior of the true m. On the basis of these weights, we introduce the new semiparametric estimator of $F_{12}(x,y)$ as

$$\widehat{F}_{12}^{sp}(x,y) = \sum_{i=1}^{n} W_i(\beta_n) I(\widetilde{T}_{1i} \leq x, \widetilde{T}_{2i} \leq y). \tag{2}$$

Unlike for (1), the estimator F_{12}^{sp} may attach positive mass to pairs of gap times with a censored T_2, while the weight attached to pairs with first gap time censored remains to be zero. As a consequence, the differences between (2) and (1) will be more evident when increasing the proportion of censoring on T_2 for the subpopulation $\Delta_1 = 1$.

More generally, we are concerned with the estimation of $S(\varphi) = E[\varphi(T_1, T_2)]$ for a given transformation φ. Specific transformations give the joint and the marginal distributions of the gap times, the moments of these variables, or the correlation coefficient. By noting $S(\varphi) = \int \varphi dF_{12}$, we introduce the following estimator of this expectation:

$$S_n(\varphi) = \int \varphi d\widehat{F}_{12}^{sp} = \sum_{i=1}^{n} W_i(\beta_n)\varphi(\widetilde{T}_{1i}, \widetilde{T}_{2i}).$$

Note that this is just $\widehat{F}_{12}^{sp}(x,y)$ when we take $\varphi(u,v) = I(u \leq x, v \leq y)$. Next result establishes the strong consistency of $S_n(\varphi)$ under an integrability condition. We will also refer to the following assumption:

$$U: \sup_{x,y} |m_1(x,y;\beta_n) - m_1(x,y)| \to 0 \qquad \text{w. p. 1},$$

which says that the function m_1 can be accurately approximated (in a uniform way) by some member of the parametric family $m_1(.,.;\beta)$, see Dikta (1998, 2000) for further discussion on this.

Theorem 1 *Assume $P(T_2 = 0) = 0$. Assume that H is continuous, that U hold, and that*

$$\int \frac{|\varphi(u,v)| F_{12}^{0}(du,dv)}{m_1(u,u+v)(1-H(u+v))^{\rho}} < \infty$$

is satisfied for some $\rho > 0$. Then, with probability 1

$$\int \varphi d\widehat{F}_{12}^{sp} \to \int \varphi dF_{12}^{0}.$$

Theorem 1 can be regarded as an adaptation of the Strong Law in Dikta (2000) to the context of censored gap times. Moreover, the result remains valid when using any presmoothing function $m_{1n}(x,y)$ satisfying assumption U, so it is not restricted to parametric presmoothing. We also indicate here that the integrability assumption in Theorem 1 is a consequence of our unknowledge on the binary regression $m_1(x,y)$; indeed, under the stronger assumption

$$U': \sup_{x,y} \left| \frac{m_1(x,y;\beta_n)}{m_1(x,y)} - 1 \right| \to 0 \qquad \text{w. p. 1},$$

it is easily seen from the proofs in the Appendix that one can state Theorem 1 merely under

$$\int \frac{|\varphi(u,v)|\, F_{12}^0(du,dv)}{(1-H(u+v))^\rho} < \infty,$$

which basically imposes the existence of the limit $\int \varphi dF_{12}^0$.

Now, an application of Theorem 1 to $\varphi(u,v) = I(u \le x, v \le y)$ leads to the pointwise convergence of $\widehat{F}_{12}^{sp}(x,y)$ to $F_{12}^0(x,y)$. Then, a standard uniformity argument gives the uniform consistency of the semiparametric estimator. This is stated as a Corollary.

Corollary 2 *Under the conditions of Theorem 1, with probability 1*

$$\sup_{x,y} \left| \widehat{F}_{12}^{sp}(x,y) - F_{12}^0(x,y) \right| \to 0.$$

From (2) we can obtain an estimator for the marginal distribution of the second gap time, $F_2(y) = P(T_2 \le y)$, namely

$$\widehat{F}_2^{sp}(y) = \widehat{F}_{12}^{sp}(\infty, y) = \sum_{i=1}^n W_i(\beta_n)\, I(\widetilde{T}_{2i} \le y). \tag{3}$$

Note that $\widehat{F}_2^{sp}(y)$ is not Dikta (1998)'s presmoothed Kaplan-Meier estimator based on the $\left(\widetilde{T}_{2i}, \Delta_{2i}\right)$'s. This is because the weights $W_i(\beta_n)$ are based on the $\widetilde{Y}_i$-ranks rather than on the $\widetilde{T}_{2i}$-ranks. Indeed, since T_2 and C_2 are expected to be dependent, the ordinary Kaplan-Meier estimator of F_2 will be in general inconsistent. As for (2), in general we have (assuming continuity for H)

$$\lim_{n\to\infty} \widehat{F}_2(y) = P(T_2 \le y, T_1 + T_2 \le \tau_H) \equiv F_2^0(y),$$

and again the restriction $T_1 + T_2 \le \tau_H$ plays a role. Hence, it is interesting to discuss the conditions under which both estimators $\widehat{F}_{12}^{sp}(x,y)$ and $\widehat{F}_2^{sp}(y)$ converge to their respective targets.

Let F and G denote the distribution functions of Y and C, respectively. Let τ_F be the upper bound of the support of F, and similarly define τ_G. Assume again that H is continuous (see de Uña-Álvarez and Meira-Machado, 2008, for a more general discussion). In essence, two different situations are possible. (A) If $\tau_F \le \tau_G$, then we get that $\widehat{F}_{12}^{sp}(x,y)$ is consistent for any (x,y). (B) If $\tau_G < \tau_F$, then $\tau_H < \tau_F$ and consistency is only ensured for $x + y \le \tau_H$. This is not surprising, since in this case relevant information on F is missing on the whole interval $(\tau_G, \tau_F]$. The bivariate estimators proposed in Wang and Wells

(1998), Lin *et al.* (1999) and de Uña-Álvarez and Meira-Machado (2008) suffer from the same problem. Similar comments hold for (3). However, note that in this latter case, to get consistency of $\widehat{F}_2^{sp}(y)$ in situation (B) one should require $P(T_1 \leq \tau_H - y) = 1$, a condition that will typically fail for y at the right tail of F_2. Specifically, if τ_1 stands for the upper bound of the support of T_1, we have $\widehat{F}_2^{sp}(y) \to F_2(y)$ w.p. 1 for $y \leq \tau_H - \tau_1$. The practical consequences of these issues will be more clearly seen when analyzing the real medical data in Section 4.

3 Simulation study

In this Section we investigate the performance of the proposed estimator $\widehat{F}_{12}^{sp}(x,y)$ through simulations. The simulated scenario is the same as that described in Lin et al. (1999) and de Uña-Álvarez and Meira-Machado (2008). To be precise, the gap times (T_1, T_2) were generated according to the bivariate distribution

$$F_{12}(x,y) = F_1(x)F_2(y)\left[1 + \theta\left\{1 - F_1(x)\right\}\left\{1 - F_2(y)\right\}\right]$$

where the marginal distribution functions F_1 and F_2 are exponential with rate parameter 1. This corresponds to the so-called Farlie-Gumbel-Morgenstern copula, where the single parameter θ controls for the amount of dependency between the gap times. The parameter θ was set to 0 for simulating independent gap times, and also to 1, corresponding to 0.25 correlation between T_1 and T_2. An independent uniform censoring time C was generated, according to models $U[0,4]$ and $U[0,3]$. The first model resulted in 24% of censoring on the first gap time, and in 47% of censoring on the second gap time. The second model increased these censoring levels to 32% and about 57%, respectively. Sample sizes 50, 100, 250 and 500 were considered. In each simulation, 1,000 samples were generated.

We considered as (x, y) pairs four different points, corresponding to the four different combinations of the percentiles 20% and 80% of the marginal distributions of the gap times. In this manner, we were able to explore the relative behavior of the estimator at the different corners of the joint distribution. As a measure of efficiency, we took the Mean Squared Error (MSE) of $\widehat{F}_{12}^{sp}(x,y)$ along the 1,000 trials. In the simulations, the MSE's were mainly determined by the variances, while the bias terms (squared) were of a smaller order of magnitude. In Tables 1 and 2 we report the MSE's attained by the proposed estimator when based on several presmoothing functions. The row labeled with m corresponds to presmoothing with the true function $m(x,y) = P(\Delta_2 = 1|\widetilde{T}_1 = x, \widetilde{Y} = y)$. This is unrealistic in practice, because this function will be typically unknown, but the figures are relevant because they represent the optimal situation in which the presmoothing function is 'perfectly estimated' (so the attained MSE's are expected to be lower bounds for the error of any realistic estimator). In the simulated models the function m is given by (for $x < y$)

$$m(x,y) = \frac{1}{1+\eta(x,y)}, \qquad \text{where } \eta(x,y) = \frac{\lambda_G(y)}{\lambda_{2|1}(y-x|x)},$$

and where $\lambda_G(.)$ and $\lambda_{2|1}(.|x)$ stand for the hazard rate functions of C and T_2 given $T_1 = x$, respectively. Note that $\lambda_G(y) = 1/(\tau_G - y)$ when $C \sim U\,[0, \tau_G]$ and that $\lambda_{2|1}(.|x)$ is given by

$$\lambda_{2|1}(y-x|x) = \frac{2+4\exp(-y)-2\exp(-x)-2\exp(-y+x)}{2+2\exp(-y)-2\exp(-x)-\exp(-y+x)} \qquad \text{if } \theta = 1,$$

being 1 when $\theta = 0$.

	$C \sim$	$U\,[0,4]$			$C \sim$	$U\,[0,3]$		
n	50	100	250	500	50	100	250	500
$m(.;\beta)$	0.7024	0.3247	0.1244	0.0708	0.7347	0.3293	0.1309	0.0663
$m(.;\gamma)$	0.7250	0.3411	0.1352	0.0786	0.7582	0.3444	0.1380	0.0725
m	0.6749	0.3095	0.1246	0.0690	0.6495	0.2900	0.1186	0.0591
KM	0.8298	0.3987	0.1604	0.0865	0.8408	0.4094	0.1579	0.0839
$m(.;\beta)$	2.9085	1.4435	0.5471	0.2989	3.0520	1.4900	0.5476	0.2821
$m(.;\gamma)$	2.9595	1.4500	0.5526	0.3080	3.0670	1.4964	0.5507	0.2842
m	2.6497	1.2990	0.5148	0.2759	2.5405	1.2782	0.4842	0.2549
KM	3.4877	1.7482	0.6752	0.3537	3.7107	1.9175	0.7235	0.3641
$m(.;\beta)$	2.9347	1.3820	0.5378	0.2664	3.2162	1.4967	0.5657	0.2922
$m(.;\gamma)$	2.9575	1.3994	0.5486	0.2737	3.2462	1.5109	0.5742	0.2970
m	2.7510	1.2487	0.5123	0.2499	2.7115	1.2622	0.5006	0.2511
KM	3.5112	1.6836	0.6582	0.3406	3.9618	1.8774	0.7539	0.3862
$m(.;\beta)$	6.8489	3.4705	1.4054	0.6695	10.211	4.7643	2.0116	0.9723
$m(.;\gamma)$	6.9993	3.5538	1.4405	0.7007	10.447	4.9738	2.1642	1.0673
m	5.4665	2.8684	1.1615	0.5388	6.5614	3.0111	1.2640	0.5878
KM	8.3579	4.3358	1.7308	0.8117	13.083	7.1644	2.8870	1.3184

Table 1. $10^3 \times MSE$ of $\widehat{F}_{12}^{sp}(x,y)$ for several presmoothing functions (see text) along 1,000 simulated samples, case $\theta = 0$. From top to bottom: $(x,y) = \left(F_1^{-1}(0.2), F_2^{-1}(0.2)\right)$, $\left(F_1^{-1}(0.8), F_2^{-1}(0.2)\right)$, $\left(F_1^{-1}(0.2), F_2^{-1}(0.8)\right)$, and $\left(F_1^{-1}(0.8), F_2^{-1}(0.8)\right)$.

Secondly, the row labeled with $m(.;\beta)$ corresponds to a presmoothing based on a certain parametric family which contains the true m. Specifically, we consider a logistic model with a preliminary transformation of the variables $\widetilde{T}_1 = x$ and $\widetilde{Y} = y$, as follows. When $\theta = 0$, for $x < y$ we took

$$m(x,y;\beta) = \frac{1}{1+\exp(\beta_0 + \beta_1\psi(x) + \beta_2\psi(y))}$$

where $\psi(s) = \ln \lambda_G(s)$. Hence, the true m corresponds to $\beta_0 = \beta_1 = 0$, $\beta_2 = 1$ in this case. When $\theta = 1$, we just took $(x < y)$

$$m(x, y; \beta) = \frac{1}{1 + \exp(\beta_0 + \beta_1 \ln(\eta(x, y)))},$$

so again the true presmoothing function is included in the parametric family, specifically it corresponds to $\beta_0 = 0$ and $\beta_1 = 1$. In order to investigate the robustness of the proposed estimator with respect to miss-specifications of the binary regression family, we considered also presmoothing via a standard logistic model, without any preliminar transformation of the gap times. This is labeled with $m(.;\gamma)$ in Tables 1 and 2. Note that the true m does not belong to this parametric family. Finally, we also report in Tables 1 and 2 the errors pertaining to the estimator in de Uña-Álvarez and Meira-Machado (2008), which corresponds to the situation with no presmoothing at all. This is labeled in the Tables as KM.

	$C \sim$	$U\,[0,4]$			$C \sim$	$U\,[0,3]$		
n	50	100	250	500	50	100	250	500
$m(.;\beta)$	1.2979	0.5735	0.2168	0.1173	1.0919	0.5928	0.2437	0.1153
$m(.;\gamma)$	1.2958	0.5708	0.2162	0.1174	1.1091	0.6047	0.2486	0.1180
m	1.2600	0.5572	0.2158	0.1141	1.0253	0.5841	0.2335	0.1120
KM	1.4068	0.6267	0.2408	0.1345	1.2210	0.6647	0.2776	0.1313
$m(.;\beta)$	3.0332	1.4798	0.5670	0.3137	3.0125	1.4339	0.6353	0.3242
$m(.;\gamma)$	3.2090	1.5668	0.6083	0.3311	3.2587	1.5223	0.6781	0.3655
m	2.9112	1.3844	0.5405	0.2969	2.7051	1.3348	0.5770	0.2857
KM	3.6242	1.8101	0.6789	0.3747	3.8507	1.8790	0.8021	0.4182
$m(.;\beta)$	3.0088	1.4905	0.6743	0.3225	3.3173	1.5772	0.6647	0.3621
$m(.;\gamma)$	3.0129	1.4956	0.6723	0.3233	3.3363	1.5768	0.6683	0.3621
m	2.8146	1.4273	0.6459	0.3079	3.0422	1.5135	0.6214	0.3390
KM	3.3812	1.6898	0.7565	0.3565	3.8003	1.8664	0.7748	0.4177
$m(.;\beta)$	6.6111	3.3523	1.4540	0.7006	9.2472	4.3998	1.8009	0.9804
$m(.;\gamma)$	6.8618	3.4152	1.4742	0.7402	10.233	4.8078	2.0860	1.2115
m	5.1991	2.7842	1.1823	0.5716	5.6046	2.6988	1.1484	0.6081
KM	8.0523	3.9276	1.6765	0.7967	13.055	6.8854	2.7521	1.6888

Table 2. $10^3 \times MSE$ of $\widehat{F}_{12}^{sp}(x, y)$ for several presmoothing functions (see text) along 1,000 simulated samples, case $\theta = 1$. From top to bottom: $(x, y) = \left(F_1^{-1}(0.2), F_2^{-1}(0.2)\right)$, $\left(F_1^{-1}(0.8), F_2^{-1}(0.2)\right)$, $\left(F_1^{-1}(0.2), F_2^{-1}(0.8)\right)$, and $\left(F_1^{-1}(0.8), F_2^{-1}(0.8)\right)$.

Some expected features are clearly seen in the Tables. For example, we see that the MSE goes down with an increasing sample size, while it increases at the right corners of the joint distribution, where the censoring effects are stronger.

Besides, results for $C \sim U[0,3]$ are in general worse than those for $C \sim U[0,4]$, although this is not true for all the situations; a possible explanation is that the presmoothing induces a kind of 'informative censoring' model, a discussion that goes back at least to Cheng and Lin (1987). On the other hand, the MSE tends to be a bit larger when introducing some correlation between the gap times (case $\theta = 1$), although some exceptions are found at the right corner of the joint distribution.

More interestingly, from Tables 1 and 2 we see that the minimum MSE is attained by the estimator which makes use of the true m. Compared to the estimator without any presmoothing (KM), it is seen that the relative efficiency of this one is about 67%-75% when taking the average along the four considered (x, y) points for each simulated scenario. However, a more careful inspection of the results reveals that, in special cases, this relative efficiency is as small as 42%. As expected, these cases correspond to the right corner of the joint distribution ($(x,y) = \left(F_1^{-1}(0.8), F_2^{-1}(0.8)\right)$) and the heavily censored case. As discussed above, in practice one has to estimate the function m. In Tables 1 and 2, the best performance among the realistic versions of $\widehat{F}_{12}^{sp}(x,y)$ corresponds to the estimator based on the right parametric family of binary regression curves. The relative efficiency of KM with respect to this estimator is about 82%-85% on average, but again in some extreme situations (right corner, heavy censoring) it goes down to only 67%. Finally, we see that the presmoothed estimator based on the wrong parametric model $m(.;\gamma)$ is still (much) better than KM; the practical message is that it is worthwhile doing some parametric presmoothing even when we are not completely sure about the parametric family.

An interesting point to discuss is that of the relative benefits of presmoothing when increasing the sample size. The figures in Tables 1 and 2 suggest that there exist a first order improvement related to presmoothing. That is, if the MSE of the KM estimator in de Uña-Álvarez and Meira-Machado (2008) is $MSE(KM) \sim c_{KM}/n$, and if the MSE pertaining to the semiparametric estimator is $MSE(SP) \sim c_{SP}/n$, then we would have $c_{SP}/c_{KM} < 1$. This is an interesting feature, since it is known that presmoothing ideas only lead to second-order improvements of the error in a number of applications (see e.g. Cao et al., 2005).

4 Real data illustration

In this Section we consider data from a cancer bladder study (Byar, 1980) conducted by the Veterans Administration Cooperative Urological Research Group. In this study, patients had superficial bladder tumors that were remove transurethrally. Many patients had multiple recurrences of tumors during the study, and new tumors were removed at each visit. Here we analyzed the n =85 individuals in the placebo and thiotepa treatment groups; these data are listed in Wei et al. (1989). Only the first two recurrence times T_1 and Y

(or the corresponding gap times T_1 and $T_2 = Y - T_1$) are considered. Among the 85 patients, 47 relapsed at least once (45% of censoring on T_1) and, among these, 29 had another recurrence (38% of extra censoring). The presence of a reasonable amount of censored Y's among the uncensored T_1's suggests that presmoothing could lead to an important reduction of variance in estimation. We will quantify this below.

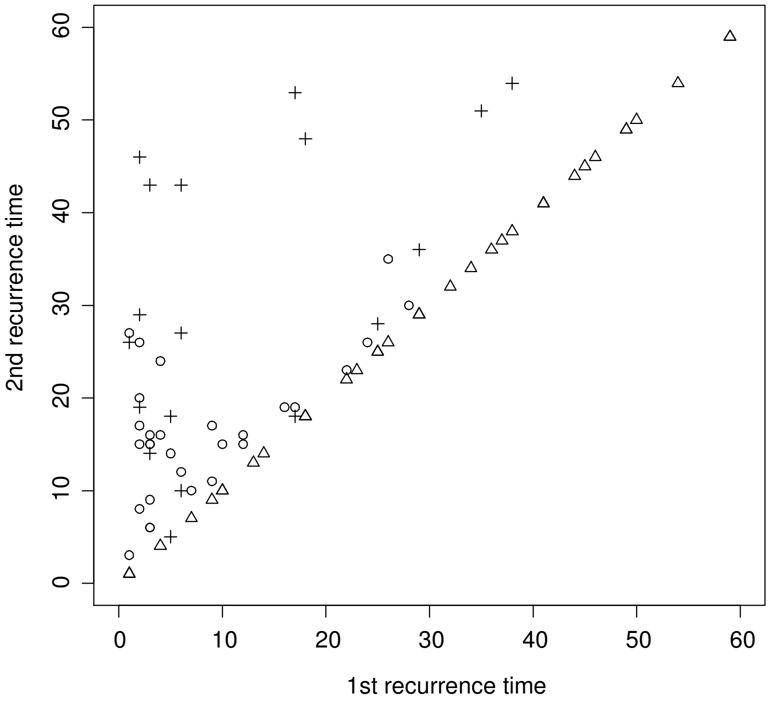

Figure 1. Time to first recurrence vs. time to second recurrence for the 85 cases of bladder cancer. Triangles indicate censoring in both times, while crosses indicate censoring on the second gap time.

In Figure 1 we represent the 85 observed values for the recurrence times $\left(\widetilde{T}_1, \widetilde{Y}\right)$ (months). Cases with both times censored are located on the line $y = x$. On the other hand, 18 points among those out of this line (labelled with a cross) correspond to observations with second gap time censored. From this Figure it is not clear in principle which type of correlation (if any) exists between both gap times T_1 and T_2. Figure 2 depicts the survival curves corresponding to T_1 (solid line) and Y (dashed line). It is clearly seen that the first recurrence is almost restricted to the first 3 years after randomization, while a large proportion of patients (about 60%) do not relapse in 5 years.

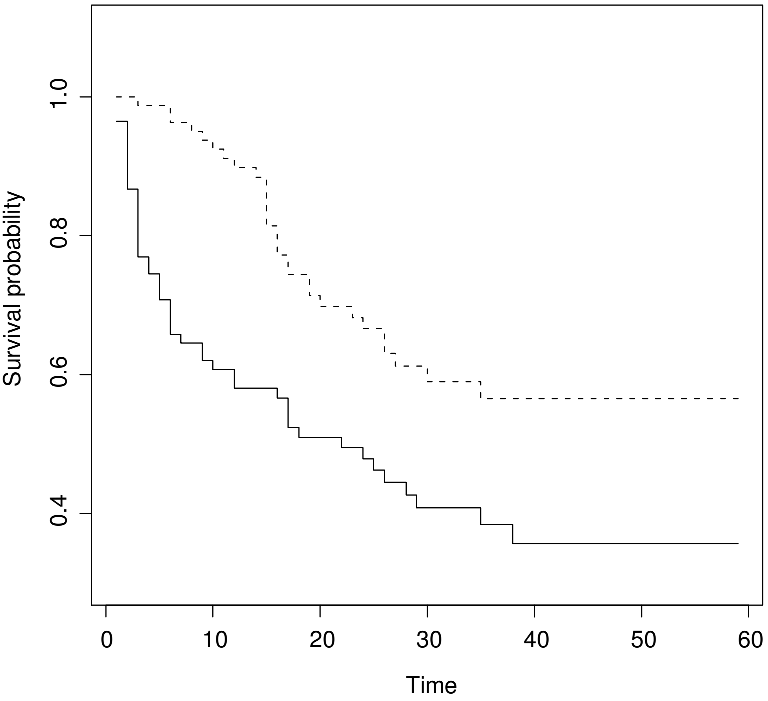

Figure 2. Kaplan-Meier curves for the bladder cancer data: time to first recurrence (solid line) and time to second recurrence.

In order to compute the semiparametric estimator (2), we have fitted a logistic model to the binary regression $m_1(x,y) = P(\Delta_2 = 1|\widetilde{T}_1 = x, \widetilde{Y} = y, \Delta_1 = 1)$. The results indicate that $\widetilde{Y}$ is highly significative (p=0.002590) while $\widetilde{T}_1$ does not reach significance (p=0.339851). The coefficient of $\widetilde{Y}$ in the model is negative, thus censoring probability increases with the observed time up to second recurrence. With this parametric presmoothing we computed the estimator $\widehat{F}_{12}^{sp}(x,y)$ for $x = 5, 10, 15, 20, 30$ months and $y = 5, 10, 20$ months. Results are displayed in Table 3, top. For comparison, we also report in this Table 3 (bottom) the values of the estimator corresponding to no presmoothing, $\widehat{F}_{12}(x,y)$. From this Table we see that both methods provide similar point estimates. We estimate the standard errors for both estimators through the simple bootstrap, which resamples (with replacement) each $\left(\widetilde{T}_{1i}, \widetilde{T}_{2i}, \Delta_{1i}, \Delta_{2i}\right)$ with probability $1/n$. The results in Table 3 (based on 5000 bootstrap resamples) reveal that: (a) the errors increase at the right corner of the joint distribution of the gap times, where the censoring effects are stronger; and (b) the semiparametric estimator has smaller standard errors, with a minimum relative efficiency of $\widehat{F}_{12}(x,y)$ of about 86% (91% when averaging the 15 cases of (x,y)).

$\widehat{F}_{12}^{sp}(x,y)$	$y=5$	$y=10$	$y=20$
$x=5$	.0454 (.0216)	.0783(.0283)	.1896 (.0433)
$x=10$	.0906 (.0294)	.1455 (.0377)	.2568 (.0488)
$x=15$	.1133 (.0335)	.1683 (.0412)	.2796 (.0514)
$x=20$	.1482 (.0374)	.2031 (.0440)	.3144 (.0528)
$x=30$	.1965 (.0462)	.2715 (.0554)	.3828 (.0604)

$\widehat{F}_{12}(x,y)$	$y=5$	$y=10$	$y=20$
$x=5$	.0372 (.0210)	.0761 (.0298)	.1921 (.0462)
$x=10$	.0775 (.0303)	.1439 (.0401)	.2598 (.0513)
$x=15$	.1056 (.0354)	.1719 (.0436)	.2879 (.0534)
$x=20$	.1359 (.0402)	.2023 (.0469)	.3183 (.0551)
$x=30$	.1920 (.0488)	.2829 (.0574)	.3989 (.0624)

Table 3. Top: Semiparametric estimator of the joint distribution function of the gap times $F_{12}(x,y)$ for the colon cancer data (standard errors between brackets). Bottom: Same information for the estimator without presmoothing.

In Figure 3 we report the semiparametric estimator of the distribution function of T_2 for the individuals with a recurrence during the first $x=30$ months of follow-up. Note that this conditional distribution is

$$F_{2|1}(y|x) = P\left(T_2 \leq y | T_1 \leq x\right) = \frac{F_{12}(x,y)}{F_1(x)},$$

where $F_1(x) = P(T_1 \leq x)$, which can be estimated by plugging-in $\widehat{F}_{12}^{sp}(x,y)$ in the numerator and the (ordinary) Kaplan-Meier for the first gap time in the denominator. We also report in this Figure 3 the estimator constructed with $\widehat{F}_{12}(x,y)$. The main difference between both curves is that the semiparametric estimator has more jump points, explicitly the censored values of T_2 for which condition $T_1 \leq 30, \Delta_1 = 1$ is satisfied. This implies that the mass is more distributed, being the reason behind the variance reduction which is achieved by presmoothing. The vertical line at $y=29$ in Figure 3 indicates that, according to our remarks to Theorem 1, both estimators should only be interpreted as empirical versions of $F_{2|1}^{\tau_H}(y|x) = P\left(T_2 \leq y, Y \leq \tau_H | T_1 \leq x\right)$ from that point on. Note that $\tau_H = 59$ in our application and hence $Y \leq \tau_H$ is not superfluous when $x=30$ and $y>29$.

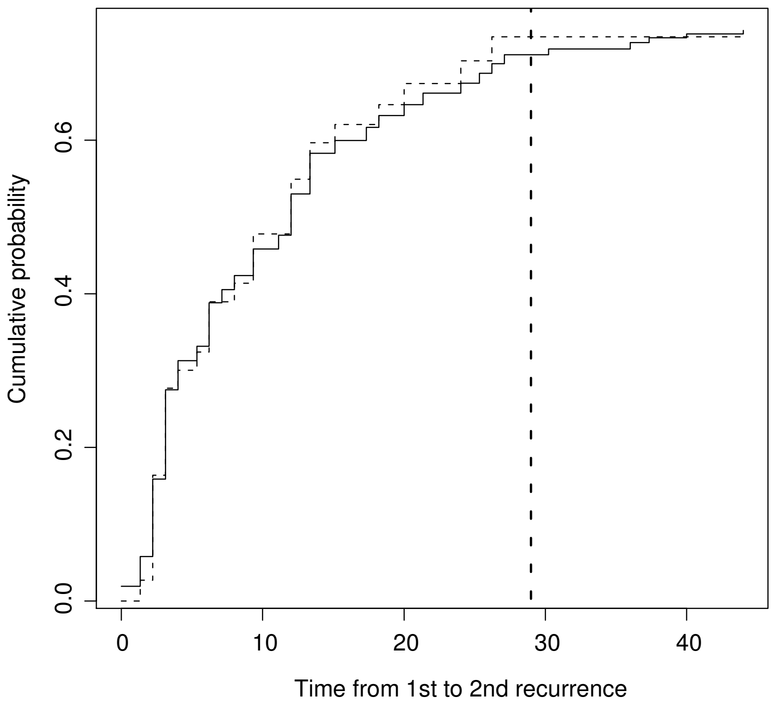

Figure 3. Semiparametric estimator (solid line) and no-presmoothed estimator (dashed line) of the distribution of time from first to second recurrence, for the subgroup with a recurrence in the first 30 months after randomization.

Finally, we give in Figures 4 and 5 two other plots which depict the joint behavior of both gap times. In Figure 4, two estimated distribution functions of T_2 based on the semiparametric estimator are plotted. The solid line corresponds to the subgroup $T_1 \leq 10$ months, while the dashed line refers to the subpopulation $10 < T_1 \leq 30$. This Figure suggests a negative correlation between both gap times. Figure 5 depicts the surface $\widehat{F}_{12}^{sp}(.,.)$, and again suggests that large times to first recurrence are connected with relatively small values of T_2.

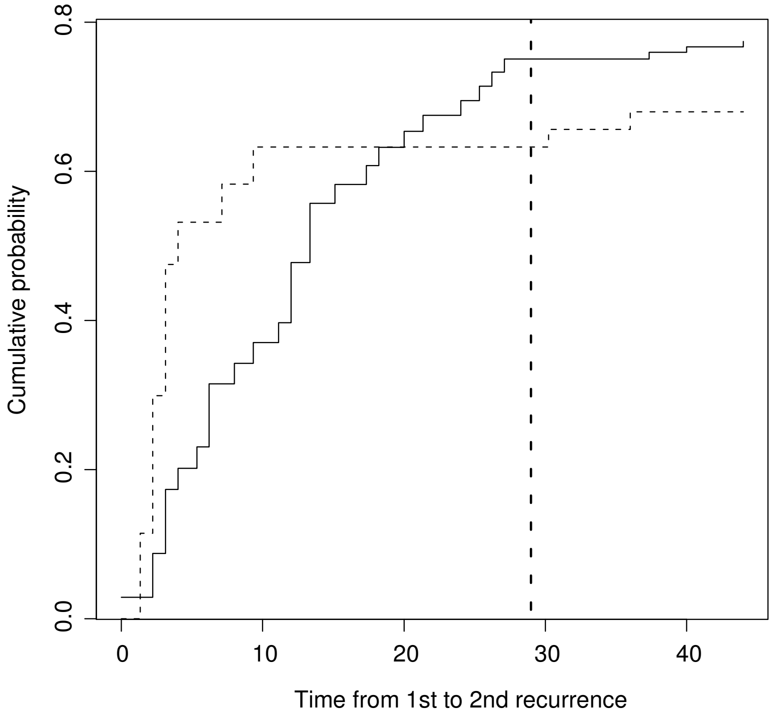

Figure 4. Semiparametric estimator of the distribution of time from first to second recurrence: relapse in the first 10 months (solid line) and relapse between month 10 and 30 (dashed line). Negative correlation between both gap times suggested.

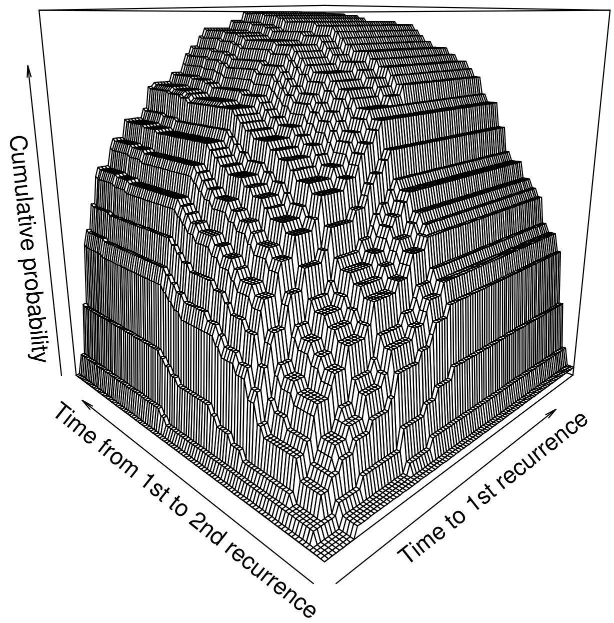

Figure 5. Cumulative joint distribution of the two gap times, based on the semiparametric estimator.

5 Conclusions and final remarks

In this paper we have introduced a new semiparametric estimator $\widehat{F}_{12}^{sp}(x,y)$ of the bivariate distribution of gap times which are observed under censoring. The semiparametric estimator is based on a parametric specification of the conditional probability of censoring for the second gap time, given the available information. This specification can be tested in practice. We have derived the consistency of the proposed estimator and, more generally, of an empirical functional based on it. We have verified through simulations that the semiparametric estimator may be much more efficient than other available estimators. This will be particularly true at points in which there is a large proportion of censored T_2 among those with $\Delta_1 = 1$. Besides, we have seen that the method is robust against miss-specifications of the parametric model. A real data illustration has been provided.

An issue of much practical interest is that of the construction of confidence limits from $\widehat{F}_{12}^{sp}(x,y)$. This could be done in a number of ways. For example, a central limit theorems and mathematical derivation of the asymptotic variance seems to be possible following the ideas in Dikta et al. (2005), see also Stute (1996a) and references therein. Besides, variance estimation could be performed on the basis of the jackknife as in Stute (1996b), or through other resampling methods such as the bootstrap. In Section 4 we have used the simple bootstrap to approximate the standard errors of the estimators, and a general result of the validity of the bootstrap in our setup would be of interest. These problems are currently under investigation.

The proposed estimator falls in the scope of the so-called presmoothing methods, which are based on the idea of replacing the censoring indicators by some smooth fit. This very idea could be applied to more complex multi-state models, as the $k-$th state progressive model or the illness-death model (see e.g. Meira-Machado et al., 2009). In the first case, the censoring indicator for the total survival time $Y = T_1 + ... + T_k$ should be replaced by a smooth (parametric) fit to the probability of censoring given the observed (possibly censored) gap times, and given that the $k-1$ first gap times are uncensored. We are exploring the benefits of this method in real data applications, and we will provide the corresponding results when more evidence on it is reached. There is some hope that presmoothing ideas can be applied for the illness-death model too, when the sojourn time in the first ('healthy') state plays the same role as the first gap time here. But this issue demands for more investigation.

Interestingly, nonparametric presmoothing is also possible for the proposed methods, as the main consistency result remains valid. This avoids the problem of choosing a proper parametric family for the binary regression. However, the gains in efficiency when using a nonparametric binary regression curve should be explored in detail. Typically, this nonparametric presmoothing will involve the selection of several smoothing parameters, which may be a critical point in the final performance of the estimator. In any case, this seems to be another promising field of research.

Acknowledgement. Work supported by the Grants MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovación and PGIDIT07PXIB300191PR of the Xunta de Galicia.

6 References

Byar, D. P. (1980). The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine and Topical Thiotepa. In: Bladder Tumors and Other Topics in Urological Oncology, eds. M. Pavone-Macaluso, P.H. Smith, and F. Edsmyn, New York: Plenum, pp. 363-370.

Cao, R., López de Ullibarri, I., Janssen, P. and Veraverbeke, N. (2005). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. Journal of Nonparametric Statistics 17, 31-56.

Cheng, P.E. and Lin, G.D. (1987). Maximum lilkelihood estimation of a survival function under the Koziol-Green proportional hazards model. Statistics & Probability Letters 5, 75-80.

Cook, R. J. and Lawless, J. F. (2007). The Analysis of Recurrent Event Data. Springer, New York.

Dikta, G. (1998). On semiparametric random censorship models. Journal of Statistical Planning and Inference 66, 253-279.

Dikta, G. (2000). The strong law under semiparametric random censorship models. Journal of Statistical Planning and Inference 83, 1-10.

Dikta, G. (2001). Weak representation of the cumulative hazard function under semiparametric random censorship models. Statsitics 35, 395-409.

Dikta, G., Ghorai, J. and Schmidt, C. (2005). The central limit theorem under semiparametric random censorship models. Journal of Statistical Planning and Inference 127, 23-51.

Iglesias-Pérez, M.C. and de Uña-Álvarez, J. (2008). Nonparametric estimation of the conditional distribution function in a semiparametric censorship model. Journal of Statistical Planning and Inference 138, 3044-3058.

Lin, D. Y., Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. Biometrika 86, 59-70.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C. and Andersen, P.K. (2009) Multi-state models for the analysis of time to event data. Statistical Methods in Medical Research 18, 195-222.

Neveu, J. (1975). *Discrete-parameter Martingales.* North-Holland, Amsterdam/Oxford.

Peña, E.A., Strawderman, R.L. and Hollander, M. (2001). Nonparametric estimation with recurrent event data. Journal of the American Statistical Association 96, 1299-1315.

Schaubel, D. E. and Cai, J. (2004). Non-parametric estimation of gap-time survival functions for ordered multivariate failure time data. Statistics in Medicine 23, 1885-1900.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. Journal of Multivariate Analysis 45, 89-103.

Stute, W. (1996a). Distributional convergence under random censroship when covariables are present. Scand. J. Statist. 23, 461-471.

Stute, W. (1996b). The jackknife estimate of variance of a Kaplan-Meier integral. Annals of Statistics 24, 2679-2704.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. Annals of Statistics 21, 1591-1607.

de Uña-Álvarez, J. and Meira-Machado, L. (2008). A simple estimator of the bivariate distribution function for censored gap times. Statistics & Probability Letters 78, 2440-2445.

de Uña-Álvarez, J. and Rodríguez-Campos, C. (2004). Strong consistency of presmoothed Kaplan-Meier integrals when covariables are present. Statistics 38, 483-496.

van der Laan, M.J., Hubbard, A.E. and Robins, J.M. (2002). Locally efficient estimation of a multivariate survival function in longitudinal studies. Journal of the American Statistical Association 97, 494-507.

Van Keilegom, I. (2004). A note on the nonparametric estimation of the bivariate distribution under dependent censoring. J. Nonpar. Statist., 16, 659-670.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. Journal of the American Statistical Association 94, 146-153.

Wang, W. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. Biometrika 85, 561-572.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84, 1065-1073.

Yuan, M. (2005). Semiparametric censorship model with covariates. Test 14, 489-514.

7 Appendix: Technical proofs

In this Section we give the technical proof to our main result (Theorem 1). We will see that this proof is similar to that of Theorem 2.1 in de Uña-Álvarez and Rodríguez-Campos (2004); here, the role of their covariate vector is played by the first gap time, while the total time Y is taken as the 'response'. Note that, since C is assumed to be independent of (T_1, T_2), the identifiability conditions H1 and H2 in de Uña-Álvarez and Rodríguez-Campos (2004) automatically hold. In our setup, these conditions read

H1. Y and C are independent
H2. $P(Y \leq C|T_1, Y) = P(Y \leq C|Y)$

which clearly follow from the independence between the censoring time and the gap times.

In order to formalize things, introduce

$$S_n(m) = \sum_{i=1}^{n} W_i(m)\varphi(\widetilde{T}_i, \widetilde{T}_{2i}) = \sum_{i=1}^{n} W_i(m)\xi^{\varphi}(\widetilde{T}_i, \widetilde{Y}_i),$$

where $W_i(m)$ are the presmoothed weights introduced in Section 2 and where $\xi^{\varphi}(u, v) = \varphi(u, v - u)$. Note that this $S_n(m)$ is an 'estimator' of $S(\varphi) = E[\varphi(T_1, T_2)] = E[\xi^{\varphi}(T_1, Y)]$ based on the true m which in practice will be unknown. Recall that the proposed semiparametric estimator of $S(\varphi)$ is

$$S_n(\varphi) = \int \varphi d\widehat{F}_{12}^{sp} = \sum_{i=1}^{n} W_i(\beta_n)\varphi(\widetilde{T}_{1i}, \widetilde{T}_{2i}) = \sum_{i=1}^{n} W_i(\beta_n)\xi^{\varphi}(\widetilde{T}_{1i}, \widetilde{Y}_i),$$

where $W_i(\beta_n) = W_i(m_n)$ with $m_n(x, y) = m(x, y; \beta_n)$ the presmoother based on the parametric model. As in de Uña-Álvarez and Rodríguez-Campos (2004), we proceed in two steps. First, we show the convergence of $S_n(m)$ to $\int \varphi dF_{12}^0 = E[\xi^{\varphi}(T_1, Y)I(Y \leq \tau_H)]$, and then we prove that the difference $S_n(\varphi) - S_n(m)$ goes to zero under appropriate conditions.

For proving the consistency of $S_n(m)$ we need three Lemmas. The first one states the supermartingale structure of $S_n(m)$, which enables us to apply powerful convergence results. The other two Lemmas allow for the identification of the limit.

Let $\widetilde{Y}_{i:n}$ be the $i-$th ordered statistic among the $\widetilde{Y}_j$'s, and let $\widetilde{T}_{[1i:n]}$ be the first gap time corresponding to $\widetilde{Y}_{i:n}$, i.e., the $i-$th concomitant. Introduce the sequence $(\mathcal{F}_n)_{n\geq 1}$, where

$$\mathcal{F}_n = \sigma\left(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}, 1 \leq i \leq n, \widetilde{T}_{1,n+1}, \widetilde{Y}_{n+1,\ldots}\right).$$

Note that $S_n(m)$ is adapted to $\mathcal{F}_n$. Note also that $\mathcal{F}_n \downarrow$ and set $\mathcal{F}_\infty = \cap_{n\geq 1}\mathcal{F}_n$ for the limit of $\mathcal{F}_n$.

Lemma 3 *Assume that H is continuous. Then,*

$$E\left[S_n(m) \mid \mathcal{F}_{n+1}\right] = S_{n+1}(m) - \frac{\xi^{\varphi}(\widetilde{T}_{[1,n+1:n+1]}, \widetilde{Y}_{n+1:n+1})}{n+1} \times$$

$$\times m(\widetilde{T}_{[1,n+1:n+1]}, \widetilde{Y}_{n+1:n+1})(1 - m(\widetilde{T}_{[1n:n+1]}, \widetilde{Y}_{n:n+1}) \prod_{j=1}^{n-1} \left[1 - \frac{m(\widetilde{T}_{[1j:n+1]}, \widetilde{Y}_{j:n+1})}{n-j+1}\right].$$

In particular, for $\varphi \geq 0$, $\left(S_n(m), \mathcal{F}_n\right)_{n \geq 1}$ is a reverse-time supermartingale.

Proof. The proof follows exactly the same steps as in the proof to Lemma 4.1 in de Uña-Álvarez and Rodríguez-Campos (2004), which in its turn is a consequence of Lemma 2.1 in Stute (1993), Lemma 2.2 in Stute and Wang (1993), and Lemma 2.1 in Dikta (2000).□

Lemma 3 allows for the application of the convergence result in Neveu (1975), Proposition V-3-11. Indeed, the Hewitt-Savage 0-1 law ensures that the limit S of $S_n(m)$ is constant with probability 1. In order to determine $S = \lim_{n\to\infty} E\left[S_n(m)\right]$, we will need the following lemma. This is a proper adaptation to our context of Lemma 2.3 in Stute (1993). Introduce the notation

$$\varphi_n(t) = \prod_{i=1}^{n} \left[1 + \frac{1 - \widetilde{m}(\widetilde{Y}_{i:n})}{n-i+1}\right]^{I\left(\widetilde{Y}_{i:n} < t\right)}, \qquad \text{where } \widetilde{m}(z) = E(\Delta_2 \mid \widetilde{Y} = z),$$

and

$$g_n(t) = E\left[\varphi_n(t)\right]; \qquad g_0(t) \equiv 1.$$

Finally,

$$\widetilde{\xi}(z) = E\left[\xi^{\varphi}(\widetilde{T}_1, \widetilde{Y})\Delta_2 \mid \widetilde{Y} = z\right].$$

Lemma 4 *Under the assumptions of Lemma 3 , we have*

$$E\left[S_n(m)\right] = E\left[\widetilde{\xi}(\widetilde{Y}) g_{n-1}(\widetilde{Y})\right].$$

Proof. Similar to that in Stute (1993), Lemma 2.3, after noting that

$$E\left[m(\widetilde{T}_1, \widetilde{Y}) \mid \widetilde{Y} = z\right] = \widetilde{m}(z), \qquad E\left[\xi^{\varphi}(\widetilde{T}_1, \widetilde{Y}) m(\widetilde{T}_1, \widetilde{Y}) \mid \widetilde{Y} = z\right] = \widetilde{\xi}(z).$$

Note that the fact that the 'covariate' $\widetilde{T}_1$ is a censored version of the 'true covariate' T_1 is not an issue here, since the outer expectation integrate this variable out.□

Now, by Stute and Wang (1993), we have

$$g_n(t) \uparrow \frac{1}{1-G(t)} \qquad \text{for each } t \text{ such that } H(t) < 1.$$

This fact together with Lemma 4 will allow for the identification of S.

Lemma 5 *Under the assumptions of Lemma 3, we have with probability 1*

$$S_n(m) \to S = \lim_{n\to\infty} E\left[S_n(m)\right] = \int \varphi dF_{12}^0.$$

Proof. Assume $\varphi \geq 0$ w.l.o.g. The general case is obtained by decomposing φ into its positive and negative part. Lemma 5 and the monotone convergence theorem give

$$S = E\left[\widetilde{\xi}(\widetilde{Y})\frac{1}{1-G(\widetilde{Y})}\right] = E\left[\frac{\xi^{\varphi}(\widetilde{T}_1,\widetilde{Y})\Delta_2}{1-G(\widetilde{Y})}\right] = E\left[\frac{\xi^{\varphi}(T_1,\widetilde{Y})\Delta_2}{1-G(\widetilde{Y})}\right] = \int \varphi dF_{12}^0,$$

where for the last equality we have used the independence between C and (T_1, Y).□

For proving that the difference $S_n(\varphi) - S_n(m)$ goes to zero, we need the following result, which is a proper adaptation of Lemma 2.2 in Dikta (2000) to our setup. Introduce for any pair of functions $p(x,z)$ and $q(x,z)$ with $0 \leq q \leq 1$ the quantity

$$\overline{S}_n(p,q) = \sum_{i=1}^{n} \overline{W}_{i,n}(p,q)\varphi(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n})$$

where

$$\overline{W}_{i,n}(p,q) = \frac{p(\widetilde{T}_{[1i:n]},\widetilde{Y}_{i:n})}{n-i+1}\prod_{j=1}^{i-1}\left[1-\frac{q(\widetilde{T}_{[1j:n]},\widetilde{Y}_{j:n})}{n-j+1}\right].$$

The proof, which we omit, is based on martingale properties (as those described in Lemma 3) of both $\overline{S}_n(p,q)$ and

$$\varphi_{q,n}(t) = \prod_{i=1}^{n}\left[1+\frac{1-\widetilde{q}(\widetilde{Y}_{i:n})}{n-i+1}\right]^{I\left(\widetilde{Y}_{i:n}<t\right)}, \qquad \text{where } \widetilde{q}(z) = E(q(\widetilde{T}_1,\widetilde{Y}) \mid \widetilde{Y} = z).$$

Lemma 6 *Under assumptions of Lemma 3, we have with probability 1*

$$\overline{S}_n(p,q) \to \overline{S}(p,q) \equiv E\left[\varphi(\widetilde{T}_1,\widetilde{Y})p(\widetilde{T}_1,\widetilde{Y})\exp\left\{\int_0^{\widetilde{Y}}\frac{1-\widetilde{q}}{1-H}dH\right\}\right].$$

Assume now that condition U holds. Then, since both $m(x,y)$ and $m(x,y;\beta_n)$ are zero for $x=y$, we have

$$\sup_{x,y} |m(x,y;\beta_n) - m(x,y)| \to 0 \qquad \text{w. p. 1.}$$

We have, for a given $\varepsilon > 0$,

$$0 \le m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}; \beta_n) \le \left| m_n(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}; \beta_n) - m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}) \right| + m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}) \le$$

$$\le \varepsilon + m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n})$$

eventually. Similarly, since $a+b \ge |a| - |b|$ whenever $a+b \ge 0$, we eventually have

$$m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}; \beta_n) \ge m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}) - \left| m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}; \beta_n) - m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}) \right| \ge$$

$$\ge m(\widetilde{T}_{[1i:n]}, \widetilde{Y}_{i:n}) - \varepsilon.$$

Introduce the functions

$$M_{1,\varepsilon}(x,z) = \max(0, m(x,z) - \varepsilon), \qquad M_{2,\varepsilon}(x,z) = \min(1, m(x,z) + \varepsilon).$$

Assume $\varphi \ge 0$ w.l.o.g. Since $M_{2,\varepsilon}(x,z) \le M_{1,\varepsilon}(x,z) + 2\varepsilon$, we get (with $m_n = m(.,.;\beta_n)$)

$$S_n(m_n) \le \overline{S}_n(M_{2,\varepsilon}, M_{1,\varepsilon}) \le S_n(M_{1,\varepsilon}) + 2\varepsilon \overline{S}_n(1, M_{1,\varepsilon})$$

where we use the obvious notation $S_n(q) = \overline{S}_n(q,q)$. We also have

$$S_n(m_n) \ge \overline{S}_n(M_{1,\varepsilon}, M_{2,\varepsilon}) \ge S_n(M_{2,\varepsilon}) - 2\varepsilon \overline{S}_n(1, M_{2,\varepsilon}).$$

Use Lemma 6 to obtain

$$\begin{aligned} S(M_{2,\varepsilon}) - 2\varepsilon\overline{S}(1, M_{2,\varepsilon}) &\le \liminf_{n\to\infty} S_n(m_n) \le \limsup_{n\to\infty} S_n(m_n) \le \\ &\le S(M_{1,\varepsilon}) + 2\varepsilon \overline{S}(1, M_{1,\varepsilon}) \end{aligned}$$

where we put $S(q) = \overline{S}(q,q)$. Bounds for $S(M_{2,\varepsilon}) - 2\varepsilon\overline{S}(1, M_{2,\varepsilon})$ and $S(M_{1,\varepsilon}) + 2\varepsilon\overline{S}(1, M_{1,\varepsilon})$ can be easily found as in Dikta (2000):

$$\begin{aligned} S(M_{1,\varepsilon}) + 2\varepsilon\overline{S}(1, M_{1,\varepsilon}) &\le \int\int \frac{\varphi(x,y)}{(1-H(x+y))^{\varepsilon}} F^0_{12}(dx,dy) + \\ &+2\varepsilon \int\int \frac{\varphi(x,y)}{m(x,x+y)(1-H(x+y))^{\varepsilon}} F^0_{12}(dx,dy), \end{aligned}$$

$$
\begin{aligned}
S(M_{2,\varepsilon}) - 2\varepsilon\overline{S}(1, M_{2,\varepsilon}) \quad \geq \quad & \int\int \varphi(x,y)(1-H(x+y))^{\varepsilon} F_{12}^{0}(dx,dy) - \\
& -2\varepsilon \int\int \frac{\varphi(x,y)}{m(x,x+y)} F_{12}^{0}(dx,dy).
\end{aligned}
$$

Note that $m(x, x+y) = m_1(x, x+y)$ unless T_2 has positive mass at zero, a situation excluded by assumption $P(T_2 = 0) = 0$. Let $\varepsilon \downarrow 0$ and apply the monotone convergence theorem to end with the proof of Theorem 1.□