# Presmoothing the transition probabilities in the illness-death model

Ana Paula Amorim, Jacobo de Uña-Álvarez, Luís Meira-Machado

# Presmoothing the transition probabilities in the illness-death model

Ana Paula Amorim[a], Jacobo de Uña-Álvarez[b], Luís Meira-Machado[c]

[a] *Department of Mathematics and Applications*
*University of Minho*
*Campus de Gualtar, 4710-057 Braga, Portugal*
*Telephone: (+351) 253604340*
*Fax: (+351) 253604369*
*E-mail: apamorim@math.uminho.pt*
[b] *Department of Statistics and O.R.*
*University of Vigo, Spain.*
[c] *Department of Mathematics and Applications*
*University of Minho, Portugal.*

**Abstract**

One major goal in clinical applications of multi-state models is the estimation of transition probabilities. In a recent paper, Meira-Machado, de Uña-Álvarez and Cadarso-Suárez (2006) introduce a substitute for the Aalen-Johansen estimator in the case of a non-Markov illness-death model. The idea behind their estimator is to weight the data by the Kaplan-Meier weights pertaining to the distribution of the total survival time of the process. In this paper we propose a modification of Meira-Machado et al (2006) estimator based on presmoothing. Consistency is established. We investigate the finite sample performance of the new estimator through simulations. Data from a study on colon cancer are used for illustration purposes.

*Keywords:* Kaplan-Meier, Markov condition, Multi-state models, semiparametric censorship

## 1. Introduction

Multi-state models (Andersen et al. 1993; Meira-Machado et al. 2009) are the most common models used for the description of longitudinal survival data. A multi-state model is a model for a stochastic process, which is characterized by a set of states and the possible transitions among them.

The states represent different situations of the individual (healthy, diseased, etc) along a follow-up. Special multi-state models that have been widely used in biomedical applications are the three-state progressive model, the illness-death model, or the bivariate model (Hougaard, 2000).

Let $X(t)$ represent the state occupied by the process at time $t \geq 0$. For two states $i$,$j$ and $s < t$, introduce the transition probability

$$p_{ij}(s,t) = P\left(X(t) = j | X(s) = i\right).$$

There has been much interest in the estimation of $p_{ij}(s,t)$ since it allows for long-term predictions of the process. Aalen and Johansen (1978) introduced a nonparametric estimator of $p_{ij}(s,t)$ for Markov models. The Markov assumption states that the future evolution of the process is independent of the previously visited states and the times of transition amongst them given the present state of the process. This simplifying assumption allows for the construction of simple estimators, since individuals with different past histories become comparable. However, it has been quoted that the Markov assumption is violated in some applications (e.g. Andersen et al., 2000). This is a relevant remark, since Aalen-Johansen estimator may be inconsistent if the process is non-Markov. Estimators of $p_{ij}(s,t)$ which are consistent in non-Markov situations are hardly found in literature.

Meira-Machado et al (2006) introduced a substitute for the Aalen-Johansen estimator in the case of a non-Markov illness-death model. They showed that when the Markov assumption does not hold, the new estimator may behave much better than the Aalen-Johansen which may be systematically biased. However, by removing the Markov condition, the proposed substitute for the Aalen-Johansen estimator provides undesirable large standard errors. This problem becomes worse when there is a large proportion of censored data. In order to overcome this issue, we propose here a modification of Meira-Machado et al (2006)'s estimator based on presmoothing, which allows for a variance reduction in the presence of censoring.

The idea of presmoothing goes back at least to Dikta (1998), see also Dikta (2000, 2001) and Dikta et al. (2005). By 'presmoothing' it is meant that each censoring indicator is replaced by a smooth fit of a binary regression of the indicator on observables. This replacement results in estimators with improved variance. Presmoothing has been successfully applied in different problems, including nonparametric curve estimation (Cao and Jácome, 2004; Cao et al., 2005) and regression analysis (de Uña-Álvarez and

Rodríguez-Campos, 2004; Yuan, 2005; Iglesias-Pérez and de Uña-Álvarez, 2008). Recently, an application of presmoothing to the estimation of the bivariate distribution of censored gap times has been provided too (de Uña-Álvarez and Amorim, 2011). In this paper we will propose presmoothed estimators of the transition probabilities $p_{ij}(s,t)$ in the scope of the illness-death model.

In order to illustrate our estimators using real data, we consider data from one of the first successful trials of adjuvant chemotherapy for colon cancer, which is freely available from the R survival package. In this study, 929 patients affected by colon cancer underwent a potential curative surgery. Unfortunately, some of these patients had residual cancer, which lead to the recurrence of disease and death (in some cases). Therefore, we may consider the recurrence as an associated state of risk, and use the so-called illness-death model with states "alive and disease-free", "alive with recurrence" (local-regional or metastases) and "dead". See Section 2 for a more formal definition of the model.

The organization of the paper is as follows. Section 2 introduces the notation, the new estimators and the main results. The finite sample performance of the proposed estimator is investigated via simulations in Section 3. In Section 4 we analyze the colon cancer data with the proposed methods. Main conclusions are reported in Section 5. Further illustration, complete simulation results and technical proofs are deferred to the web-only Appendix.

## 2. The estimator: main result

In this paper we consider the illness-death model depicted in Figure 1. In this model, all the subjects are in State 1 ('healthy') at time $t = 0$. At some future time, they will arrive at State 3 ('dead'), which is absorbing. In the meanwhile they may visit State 2 ('diseased') at some time point; or not, passing directly to State 3 without visiting State 2. Note that this multi-state model is progressive (Hougaard, 2000), in the sense that past states can not be revisited. For this model the set of states is $\mathcal{S} = \{1, 2, 3\}$, and the transitions allowed are $1 \to 2$, $1 \to 3$, and $2 \to 3$. Given two time points $s < t$, there are in essence three different transition probabilities to estimate: $p_{11}(s,t)$, $p_{13}(s,t)$, and $p_{23}(s,t)$. The two other transition probabilities ($p_{12}(s,t)$ and $p_{22}(s,t)$) are easily obtained from $p_{12}(s,t) = 1 - p_{11}(s,t) - p_{13}(s,t)$ and $p_{22}(s,t) = 1 - p_{23}(s,t)$.

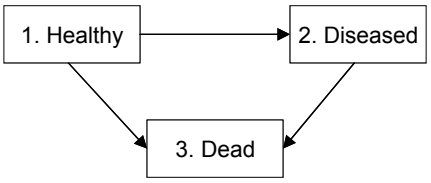


Figure 1: Illness-death model: the three states (boxes) and the possible transition among them (arrows).

Let $T_{ij}$ be the potential transition time from state $i$ to state $j$. This means that a subject not visiting state 2 will reach the 'dead' state at time $T_{13}$, while this time will be $T_{12}+T_{23}$ if he/she passes through state 2 before. We denote by $\rho = I(T_{12} \leq T_{13})$ the indicator of visiting state 2 at some time. Let $Z = T_{12} \wedge T_{13}$ be the sojourn time in state 1, and let $T = Z + \rho T_{23}$ be the total survival time of the process (up to reaching the absorbing state). We denote the censoring variable by $C$ which is assumed to be independent of the process; finally, we put $\widetilde{Z} = Z \wedge C$ and $\widetilde{T} = T \wedge C$ for the censored versions of $Z$ and $T$, and $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$ for the respective censoring indicators. With this notation, the transition probabilities are written as

$$\begin{aligned} p_{11}(s,t) &= \frac{P(Z>t)}{P(Z>s)}, \quad p_{13}(s,t) = \frac{P(s<Z, T \leq t)}{P(Z>s)}, \\ p_{23}(s,t) &= \frac{P(Z \leq s, s<T \leq t)}{P(Z \leq s<T)}. \end{aligned}$$

All these quantities involve expectations of particular transformations of the pair $(Z,T)$, $S(\varphi) = E[\varphi(Z,T)]$ say. Thus we now discuss how these expectations can be empirically approximated from the data

$$\left\{ \left( \widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i, \Delta_{1i}\rho_i \right), 1 \leq i \leq n \right\},$$

which are assumed to form a random sample of the vector $\left( \widetilde{Z}, \widetilde{T}, \Delta_1, \Delta, \Delta_1 \rho \right)$. Note that $p_{11}(s,t)$ and the denominator of $p_{13}(s,t)$ only involve the $Z$ variable, and that they can be estimated by the ordinary Kaplan-Meier estimator of the sojourn time distribution in state 1. However, the remaining quantities cannot be estimated so simply.

Let $\widetilde{T}_{1:n} \leq \cdots \leq \widetilde{T}_{n:n}$ denote the ordered $\widetilde{T}_i$'s, and let $W_i$ be the Kaplan-Meier weight attached to $\widetilde{T}_{i:n}$ when estimating the marginal distribution of

$T$ from $(\widetilde{T}_i, \Delta_i)$'s. That is,

$$W_i = \frac{\Delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left[ 1 - \frac{\Delta_{[j:n]}}{n-j+1} \right]$$

where $\Delta_{[i:n]}$ is the $i$th concomitant of $\widetilde{T}_{i:n}$. Here, ties within the censored or within the uncensored times are ordered arbitrarily, and ties among the uncensored and censored times are treated as if the former precede the later.

In the uncensored case we have $W_i = n^{-1}$ for each $i$. In Meira-Machado et al (2006) the following estimator of $S(\varphi)$ was proposed:

$$S_n(\varphi) = \sum_{i=1}^{n} W_i \varphi(\widetilde{Z}_{[i:n]}, \widetilde{T}_{i:n}).$$

where $\widetilde{Z}_{[i:n]}$ is the concomitant of $\widetilde{T}_{i:n}$. Consider now the presmoothed version of $S_n(\varphi)$ given by

$$S_n(\varphi; m_n) = \sum_{i=1}^{n} W_i(m_n) \varphi(\widetilde{Z}_{[i:n]}, \widetilde{T}_{i:n})$$

where

$$W_i(m_n) = \frac{m_n\left(\widetilde{Z}_{[i:n]}, \widetilde{T}_{i:n}\right)}{n-i+1} \prod_{j=1}^{i-1} \left[ 1 - \frac{m_n\left(\widetilde{Z}_{[j:n]}, \widetilde{T}_{j:n}\right)}{n-j+1} \right]$$

and where $m_n(z,t)$ stands for an estimator of the binary regression function

$$m(z,t) = P\left(\Delta = 1 | \widetilde{Z} = z, \widetilde{T} = t\right).$$

Since $\left(\widetilde{Z}, \widetilde{T}, \Delta\right)$ are observable, the function $m(z,t)$ can be estimated by standard methods. However, the naive construction of a smooth estimator for $m(z,t)$ will generally fail. This is because the function $m(z,t)$ will typically be discontinuous along the line $t = z$, that is, for those covariate values $\left(\widetilde{Z}, \widetilde{T}\right)$ corresponding to individuals who are censored while being in state 1 or who suffer a direct transition $1 \to 3$ to the absorbing state.

In order to see this, note that for $z < t$ we have

$$m(z,t) = P\left(\Delta = 1 | \widetilde{Z} = z, \widetilde{T} = t, \Delta_1 \rho = 1\right) \equiv m_1(z,t),$$

that is, $m_1\left(\widetilde{Z},\widetilde{T}\right)$ is the conditional probability of censoring on $T$ given $\left(\widetilde{Z},\widetilde{T}\right)$ and given that transition $1 \to 2$ is observed ($\Delta_1\rho = 1$). However, when $z = t$ we get

$$m(t,t) = P\left(\Delta_1 = 1 | \widetilde{Z} = t, \Delta_1\rho = 0\right) \equiv m_2(t),$$

which is the conditional probability of observing $1 \to 3$ given $\widetilde{Z} = t$ (or $\widetilde{T} = t$) and given that transition $1 \to 2$ is never observed. We implicitly assume that the events $\left\{\widetilde{Z} = \widetilde{T}\right\}$ and $\Delta_1\rho = 0$ are the same. This is reasonable unless there is a significative proportion of individuals with zero sojourn time in state 2. These formulae show that the functions $m_1$ and $m_2$ represent different binary regression problems and that they are based on disjoint subpopulations (according to the value of $\Delta_1\rho$). Furthermore, the limit of $m_1(z,t)$ as $z$ approaches to $t$ does not coincide with $m_2(t)$ in reality. Figure 2 displays these functions for the colon cancer data, when estimated separately by two logistic models. The noise around $m_1(z,t)$ comes from the fact that the variable $z$ is omitted from the plot while it is present in the model (although without reaching statistical significance, p-value=0.285). Both functions are clearly separated.

In summary, in order to construct $m_n(z,t)$ we propose that one estimate the functions $m_1(z,t)$ and $m_2(t)$ independently by fitting some smooth models, $m_{1n}(z,t)$ and $m_{2n}(t)$ say, so we finally have

$$m_n(z,t) = m_{1n}(z,t)\, I(z < t) + m_{2n}(t) I(z = t),$$

or

$$\begin{aligned} m_n\left(\widetilde{Z}_i,\widetilde{T}_i\right) &= m_{1n}\left(\widetilde{Z}_i,\widetilde{T}_i\right) I(\widetilde{Z}_i < \widetilde{T}_i) + m_{2n}(\widetilde{Z}_i) I(\widetilde{Z}_i = \widetilde{T}_i) \\ &= m_{1n}\left(\widetilde{Z}_i,\widetilde{T}_i\right) \Delta_{1i}\rho_i + m_{2n}(\widetilde{Z}_i)(1 - \Delta_{1i}\rho_i). \end{aligned}$$

The estimator $m_{1n}(z,t)$ is based on the subsample $\{i : \Delta_{1i}\rho_i = 1\}$, while $m_{2n}(t)$ is computed from $\{i : \Delta_{1i}\rho_i = 0\}$. The only condition we assume on these two functions is that they should approximate well their targets in a uniform sense; more specifically, set

$$U_1 : \sup_{z,t} |m_{1n}(z,t) - m_1(z,t)| \to 0 \qquad \text{w. p. 1},$$

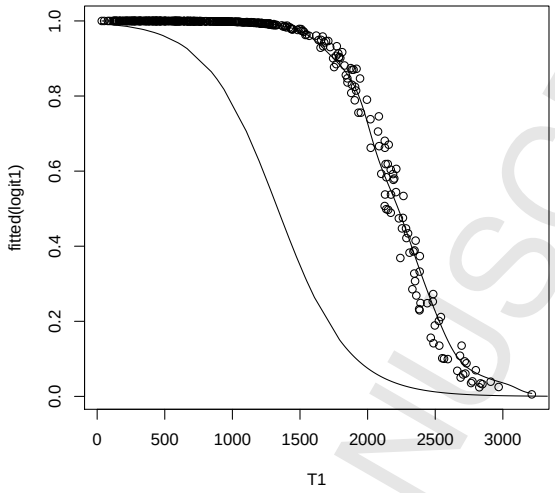


Figure 2: Presmoothing functions $m_1$ and $m_2$ estimated by logistic models vs. $\widetilde{T}$ (variable $\widetilde{Z}$ not shown). Colon cancer data.

and

$$U_2: \ \sup_t |m_{2n}(t) - m_2(t)| \to 0 \qquad \text{w. p. 1.}$$

Since $m(z,t) = m_1(z,t)\, I(z < t) + m_2(t) I(z = t)$, under $U_1$ and $U_2$ we have

$$U: \ \sup_{z,t} |m_n(z,t) - m(z,t)| \to 0 \qquad \text{w. p. 1.}$$

and hence Theorem 2.1 in de Uña-Álvarez and Rodríguez-Campos (2004) can be applied with some adaptation to the present context (see the Appendix). Conditions under which $U_1$ and $U_2$ hold are investigated in a number of papers, including Devroye (1978a, 1978b), Mack and Silverman (1982), and Härdle and Luckhaus (1984). Now we state our main result and the corresponding corollaries. Let $H$ be the distribution function of $\widetilde{T}$ and let $\tau_H = \inf\{t : H(t) = 1\}$.

**Theorem 1.** Assume that $H$ is continuous, that $U_1$ and $U_2$ hold, and that

$$E\left[\frac{|\varphi(Z,T)|\, I(T \leq \tau_H)}{m(Z,T)(1 - H(T))^{\rho}}\right] < \infty$$

is satisfied for some $\rho > 0$. Then, $S_n(\varphi; m_n) \to S^\tau(\varphi)$ with probability 1, where $S^\tau(\varphi) = E[\varphi(Z,T)I(T \leq \tau_H)]$.

Theorem 1 is a proper adaptation of the Strong Law in Dikta (2000) to our scenario. We note that the result is not restricted to parametric presmoothing; the only thing one should have in mind is that condition $U$ must be verified by the chosen estimator $m_n(z,t)$. Note also that, in general, one can not ensure that $S^\tau(\varphi)$ and $S(\varphi)$ will coincide; indeed, as always with censored data, one should not expect consistency beyond the upper bound of the censoring distribution, because there is no sampling information regarding the lifetime there. As a particular case, we have $S^\tau(\varphi) = S(\varphi)$ if the support of $T$ is contained in that of $C$.

The proof to Theorem 1 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 sojourn time in State 1 ($Z$), while the total time $T$ up to reaching the absorbing State 3 is taken as the 'response'. However, since $Z$ is also subject to right-censoring, the results in the referred paper (see also Stute, 1993, and Stute and Wang, 1993) do not directly apply here. Note that, since $C$ is assumed to be independent of $(Z,T)$, the identifiability conditions H1 and H2 in de Uña-Álvarez and Rodríguez-Campos (2004) automatically hold. In our setup, these conditions read

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

which clearly follow from the independence between the censoring time and the process. See the web-appendix for details.

Now, we come back to our initial goal of estimating the transition probabilities $p_{ij}(s,t)$. Recall that $p_{11}(s,t)$ can be estimated by the ordinary Kaplan-Meier based on the $\left(\widetilde{Z}_i, \Delta_{1i}\right)$'s. In order to introduce some presmoothing, we recommend to replace the $\Delta_{1i}$'s by some smooth fit to the binary regression function $P\left(\Delta_1 = 1|\widetilde{Z} = z\right)$ (see e.g. Dikta, 1998). Now, we focus on the estimation of $p_{13}(s,t)$ and $p_{23}(s,t)$. Write

$$p_{13}(s,t) = \frac{P(s < Z, T \leq t)}{P(s < Z)} = \frac{E[\varphi_{s,t}(Z,T)]}{P(s < Z)},$$

where $\varphi_{s,t}(u,v) = I(u > s, v \leq t)$. Introduce the presmoothed estimator

$$\widehat{p}_{13}(s,t) = \frac{S_n(\varphi_{s,t}; m_n)}{\widehat{P}(s < Z)}$$

where $\widehat{P}(s < Z)$ stands for a consistent estimator (e.g. Kaplan-Meier) of $P(s < Z)$.

Similarly, we have

$$p_{23}(s,t) = \frac{P(Z \leq s, s < T \leq t)}{P(Z \leq s < T)} = \frac{E[\widetilde{\varphi}_{s,t}(Z,T)]}{E[\overline{\varphi}_s(Z,T)]}$$

where $\widetilde{\varphi}_{s,t}(u,v) = I(u \leq s, s < v \leq t)$ and $\overline{\varphi}_s(u,v) = I(u \leq s < v)$. Therefore, in this case, we estimate the transition probability through

$$\widehat{p}_{23}(s,t) = \frac{S_n(\widetilde{\varphi}_{s,t}; m_n)}{S_n(\overline{\varphi}_s; m_n)}.$$

We have the following Corollary.

**Corollary 1.** Assume that the conditions in Theorem 1 hold for the special $\varphi$-functions $\varphi_{s,t}$, $\widetilde{\varphi}_{s,t}$ and $\overline{\varphi}_s$. Then, for any consistent estimator $\widehat{P}(Z > s)$ of $P(Z > s)$ we have with probability 1 $\widehat{p}_{13}(s,t) \to p^{\tau}_{13}(s,t)$ and $\widehat{p}_{23}(s,t) \to p^{\tau}_{23}(s,t)$, where $p^{\tau}_{13}(s,t) = P(T \leq t, T \leq \tau_H \ / \ Z > s)$ and $p^{\tau}_{23}(s,t) = P(T \leq t \ / \ Z \leq s < T, T \leq \tau_H)$. □

Corollary 1 is an immediate consequence of Theorem 1. As for the Theorem, consistency can not be ensured in general. However, when the support of $C$ contains that of $T$ we have $p^{\tau}_{13}(s,t) = p_{13}(s,t)$ and $p^{\tau}_{23}(s,t) = p_{23}(s,t)$. In particular this may happens whenever $\tau_H = \infty$.

**Remark** In practice, when $n$ is small, it may happen $\widehat{p}_{13}(s,t) > 1$ and/or $\widehat{p}_{11}(s,t) + \widehat{p}_{13}(s,t) > 1$. When any of these inequalities occurs, we propose the modification $\widehat{p}_{13}(s,t) = 1 - \widehat{p}_{11}(s,t)$, which ensures $\widehat{p}_{13}(s,t) \leq 1$ and $\widehat{p}_{11}(s,t) + \widehat{p}_{13}(s,t) \leq 1$. With this remark in mind, we always have $\widehat{p}_{12}(s,t) = 1 - \widehat{p}_{11}(s,t) - \widehat{p}_{13}(s,t) \geq 0$. For moderate or large sample sizes this problem disappears.

## 3. Simulation study

In this Section we investigate the performance of the proposed estimators $\widehat{p}_{ij}(s,t)$ through simulations. More specifically, the estimators $\widehat{p}_{11}(s,t)$, $\widehat{p}_{13}(s,t)$ and $\widehat{p}_{23}(s,t)$ introduced in Section 2 are considered.

To simulate the data in the illness-death model, we separately consider the subjects passing through State 2 at some time (that is, those cases with $\rho = 1$), and those who directly go to the absorbing State 3 ($\rho = 0$). For the first subgroup of individuals ($\rho = 1$), the successive gap times $(Z, T - Z)$ are simulated 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 $\theta$ controls for the amount of dependency between the gap times. The parameter $\theta$ was set to 0 for simulating independent gap times, and also to 1, corresponding to 0.25 correlation between $Z$ and $T - Z$. This simulated scenario is the same as that described in Lin et al. (1999) and de Uña-Álvarez and Meira-Machado (2008). For the second subgroup of individuals ($\rho = 0$), the value of $Z$ is simulated according to an exponential with rate parameter 1. In summary, the simulation procedure is as follows:

Step 1. Draw $\rho \sim Ber(p)$ where $p$ is the proportion of subjects passing through State 2.

Step 2. If $\rho = 1$ then:

(2.1) $V_1 \sim U(0,1), V_2 \sim U(0,1)$ are independently generated;

(2.2) $U_1 = V_1, A = \theta(2U_1 - 1) - 1, B = (1 - \theta(2U_1 - 1))^2 + 4\theta V_2(2U_1 - 1)$

(2.3) $U_2 = 2V_2 / \left(\sqrt{B} - A\right)$

(2.4) $Z = \ln(1/(1 - U_1)), T = \ln(1/(1 - U_2)) + Z$

If $\rho = 0$ then $Z = \ln(1/(1 - U(0,1)))$.

Situations with $p = 1$ corresponds to the three-state progressive model, in which a direct transition $1 \to 3$ is not allowed. In our simulation we consider $p = 0.7$. An independent uniform censoring time $C$ is generated, according to models $U[0,4]$ and $U[0,3]$. The first model results in 24% of censoring on the first gap time $Z$, and in 47% of censoring on the second gap time $T - Z$,

for those individuals with $\rho = 1$. The second model increases these censoring levels to 32% and about 57%, respectively.

After some algebra, it is seen that the function

$$m_1(z,t) = P\left(\Delta = 1 | \widetilde{Z} = z, \widetilde{T} = t, \Delta_1 \rho = 1\right)$$

is written as

$$m_1(z,t) = \frac{1}{1+\eta_1(z,t)}, \qquad \text{where} \quad \eta_1(z,t) = \frac{\lambda_G(t)}{\lambda^1_{T|Z=z}(t|z)}$$

and where $\lambda_G(.)$ and $\lambda^1_{T|Z=z}(.|z)$ stand respectively for the hazard rate of the censoring variable and the hazard rate of $T$ given $Z = z$ under restriction $\rho = 1$. Note that $\lambda_G(t) = 1/(\tau_G - t)$ when $C \sim U[0, \tau_G]$ and that $\lambda^1_{T|Z=z}(t|z)$ is given by

$$\lambda^1_{T|Z=z}(t|z) = \frac{2 + 4\exp(-t) - 2\exp(-z) - 2\exp(-t+z)}{2 + 2\exp(-t) - 2\exp(-z) - \exp(-t+z)} \qquad \text{if } \theta = 1,$$

being 1 when $\theta = 0$. The function $m_1(z,t)$ belongs to the logistic family with some preliminary transformation of the conditioning variables. To be more specific we have (for $\beta_0 = 0$ and $\beta_1 = 1$)

$$m_1(z,t;\beta) = \frac{1}{1+\exp(\beta_0 + \beta_1 \ln(\eta_1(z,t)))}.$$

This is the parametric model we fit to $m_1(z,t)$ in the simulations. The $\beta$ parameter in model $m_1(.;\beta)$ is estimated via maximization of the conditional likelihood of the $\Delta_i$'s given the $(\widetilde{Z}_i, \widetilde{T}_i)$'s, for those subjects with $\Delta_1\rho = 1$ (see e.g. Dikta, 1998, 2000). The same estimation criterium is used for the other presmoothing functions ($m_0$ and $m_2$) in this section. For $m_2(t) = P\left(\Delta_1 = 1 | \widetilde{Z} = t, \Delta_1\rho = 0\right)$, we have

$$m_2(t) = \frac{1}{1+\eta_2(t)}, \qquad \text{where} \quad \eta_2(t) = \frac{\lambda_G(t)}{\lambda^0_Z(t)}$$

and where $\lambda^0_Z(t)$ stands for the sub-hazard function of $Z$ restricted to $\rho = 0$, namely

$$\lambda^0_Z(t) = P(Z = t, \rho = 0 | Z \geq t) = 1 - p.$$

Similarly as above, we fit the logistic model

$$m_2(t;\gamma) = \frac{1}{1+\exp(\gamma_0 + \gamma_1 \ln(\eta_2(t)))}$$

to estimate the function $m_2(t)$ in the simulations. As before, this logistic model has the true presmoothing function $m_2$ as a special case ($\gamma_0 = 0$,$\gamma_1 = 1$).

The aim of this simulation study is to compare the estimator by Meira-Machado et al (2006) and the new estimator based on presmoothing ideas. In order to measure the estimates' relative performance, we computed the integrated absolute bias, integrated variance and the integrated Mean Square Error (MSE) of the estimates. For each simulated setting we derived the analytic expression of $p_{11}(s,t)$, $p_{13}(s,t)$ and $p_{23}(s,t)$ so that the bias and the MSE of the estimator could be examined. Sample sizes 50, 100 and 200 were considered. In each simulation, $K = 1000$ samples were generated.

Let $\hat{p}_{ij}^k(s,t)$ denote the estimated transition probability based on the $k$th generated data set. For each fixed $(s,t)$ we obtained the mean for all generated data sets, $\overline{\hat{p}_{ij}(s,t)} = \frac{1}{K}\sum_{k=1}^{K} \hat{p}_{ij}^k(s,t)$. We then computed the pointwise estimates of the bias, variance and MSE as:

$$\widehat{bias}(s,t) = p_{ij}(s,t) - \overline{\hat{p}_{ij}(s,t)}$$

$$\widehat{var}(\hat{p}_{ij}(s,t)) = \frac{1}{K-1}\sum_{k=1}^{K}[\hat{p}_{ij}^k(s,t) - \overline{\hat{p}_{ij}(s,t)}]^2$$

$$\widehat{MSE}(\hat{p}_{ij}(s,t)) = \frac{1}{K}\sum_{k=1}^{K}[\hat{p}_{ij}^k(s,t) - p_{ij}(s,t)]^2$$

To summarize the results we also calculated the integrated absolute bias, integrated variance and the integrated MSE, defined in Table 1. We fixed the values of $s$ using the quantiles 0.25, 0.5 and 0.75 of the exponential distribution with rate 1. The results obtained in Table 2 and 3 were obtained by numerical integration on the interval $[s,t_1]$ with $t_1 = 3$, taking a grid of step $\delta = 0.05$.

In Tables 2 and 3 we report the results for the integrated absolute bias, integrated variance and the integrated MSE attained by the proposed estimators for $p_{23}(s,t)$ when based on several presmoothing functions, in the

| Statistic | Definition | Estimator |
|---|---|---|
| Integrated absolute bias | $\int_s^{t_1} \lvert bias(s,t)\rvert dt$ | $\sum_{t=s}^{t_1} \lvert\widehat{bias}(s,t)\rvert\delta$ |
| Integrated variance | $\int_s^{t_1} var(\hat{p}_{ij}(s,t))dt$ | $\sum_{t=s}^{t_1} \widehat{var}(\hat{p}_{ij}(s,t))\delta$ |
| Integrated MSE | $\int_s^{t_1} MSE(\hat{p}_{ij}(s,t))dt$ | $\sum_{t=s}^{t_1} \widehat{MSE}(\hat{p}_{ij}(s,t))\delta$ |

Table 1: Summary statistics measuring bias, variance and mean square error.

scenario with $\theta = 1$ (dependent transition times). The row labeled $m$ corresponds to presmoothing with the true function $m(z,t) = P(\Delta = 1|\tilde{Z} = z, \tilde{T} = t)$, which is unrealistic because this function will generally be unknown. However, this row represents a 'gold standard' the other methods can be compared to. The row labeled with $m(.;\beta,\gamma)$ corresponds to a semiparametric estimator which is obtained using a presmoothing based on a parametric family which contains the true $m$. Specifically, we consider a logistic model with the preliminary transformation of the conditioning variables $\tilde{Z} = z, \tilde{T} = t$ shown before. Results for $\widehat{p}_{11}(s,t)$ and $\widehat{p}_{13}(s,t)$ are shown in the web-appendix. Similarly, for $p_{11}(s,t)$ and for the denominator of $p_{13}(s,t)$ we also perform logistic presmoothing for the function $m_0(z) = P\left(\Delta_1 = 1|\widetilde{Z} = z\right)$, with the variable $\widetilde{Z}$ transformed by $-\ln(\tau_G - \widetilde{Z})$ (so the parametric family contains the true $m_0(z)$).

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 standard logistic models, without any preliminary transformation of the transition times. This is labeled with $m(.,\xi)$ in Tables 2 and 3. Note that the true $m$ and the true $m_0$ do not belong to this parametric family. Finally, we also report the results pertaining to the estimator in Meira-Machado et al (2006), which corresponds to the situation with no presmoothing at all. This is labeled in the Tables as KM.

Some expected features are clearly seen in Tables 2 and 3. For example, we see that the (integrated) MSE, bias and variance of $\hat{p}_{23}(s,t)$ decrease with an increasing sample size, while they increase with the censoring degree. The best performance is attained by the estimator which makes use of the true $m$, which was expected. However, in practice one must estimate the function $m$. The lowest errors among the realistic versions of the estimators correspond to the estimator based on the correctly specified parametric family, $m(.;\beta,\gamma)$. Finally, we see that the presmoothed estimator based on the wrong parametric model $m(.;\xi)$ is still (much) better than KM; the practical

message is that it is worthwhile doing some presmoothing even when we are not completely sure about the parametric family.

Compared to the estimator without any presmoothing (KM), it is proven that the relative efficiency of the estimators based on presmoothing is always above 1. In special cases, the relative deficiency of the Kaplan-Meier estimator is below 50%; this occurs for larger values of $s$, where the censoring effects are stronger. This supports the belief that the relative benefits of presmoothing will be seen more clearly in the presence of large censoring degrees.

Although we restrict the integrated bias, variance and mean square error (MSE) to the interval $[s, 3]$, we verified that, in both settings, the enlargement of this interval favors the estimator based on presmoothing (results not shown). This happens because higher levels of censoring are expected in the right tail of the distribution. We also run simulations for mutually independent transition times $T_{ij}$. The results (shown in the web-appendix) also revealed advantages on the use of presmoothing.

## 4. Colon cancer study

For illustration, we apply the proposed methods of Section 2 to data from a large clinical trial on Duke's stage III patients, affected by colon cancer, that underwent a curative surgery for colo-rectal cancer (Moertel et al. 1990). In this study, from the total of 929 patients, 468 developed recurrence and among these 414 died. 38 patients died without recurrence. The rest of the patients (423) remained alive and disease-free up to the end of the follow-up. As mentioned in the Introduction recurrence can be expressed as an intermediate event which can be modeled using an illness-death model.

Using the Cox proportional hazards model, we verified that the transition rate from state 2 to state 3 is affected by the time spent in the previous state. This allowed us to conclude that the Markov assumption may be unsatisfactory for the colon cancer data set. In this section we will present estimated transition probabilities calculated using the new approach, based on presmoothed Kaplan-Meier weights and the estimators of Meira-Machado et al (2006). Neither one of the approaches assume the process as being Markovian.

In Figure 3 we illustrate differences between the estimated transition probabilities $p_{ij}(s,t)$, $1 \leq i \leq j \leq 3$ based on presmoothing the Kaplan-Meier weights and the estimator corresponding to no presmoothing (KM;

Meira-Machado et al 2006). The presmoothed estimator was obtained by standard logistic regression for both $m_1$ and $m_2$. The value $s$ was chosen to be the 75th percentile of the sojourn time in state 1 ($s = 1549$ days). From this figure we conclude that the new estimator have more jump points (corresponding to patients with censored values of the total time) but with smaller steps. The number of jump points and the size of the steps are related to the censoring degree and to the sample size. We can also verify that both methods provide similar point estimates for small time values. Departures between both estimated curves can be more appreciated for larger time values where the censoring effects are stronger. In summary, the new approach provides more reliable curves with less variability, specially at the right tail of the lifetime distribution. Other values of $s$ reported similar results.

## 5. Conclusions and final remarks

In this paper we have introduced new estimators for the transition probabilities of a censored illness-death model. The new estimators are based on a preliminary estimation (presmoothing) of the probability of censoring for the total time, given the available information. This idea has been used before in univariate survival analysis, but its application to multi-state survival data is complicated since new problems arise. More explicitly, the illness-death model involves presmoothing functions which are discontinuous, so a naive estimation approach fails.

We have derived the consistency of the proposed estimators. The consistency result is not restricted to parametric presmoothing, but it also includes the possibility of using some nonparametric estimators to this end. The finite sample performance of the introduced estimators was investigated through simulations. The main conclusion is that presmoothing leads to improved estimators, even when there is some miss-specication in the parametric family assumed for the presmoothing function. The relative benefits of presmoothing are more clearly seen in the heavily censored case. The new method has been illustrated using data from a colon cancer study.

The new estimators for the transition probabilities are consistent regardless the Markov condition. This is interesting because real problems are often far from markovianity and therefore the consistency of the time-honored Aalen-Johansen estimator can not be ensured. To this regard, one may think about the methods introduced here as a remarkable improvement (in the

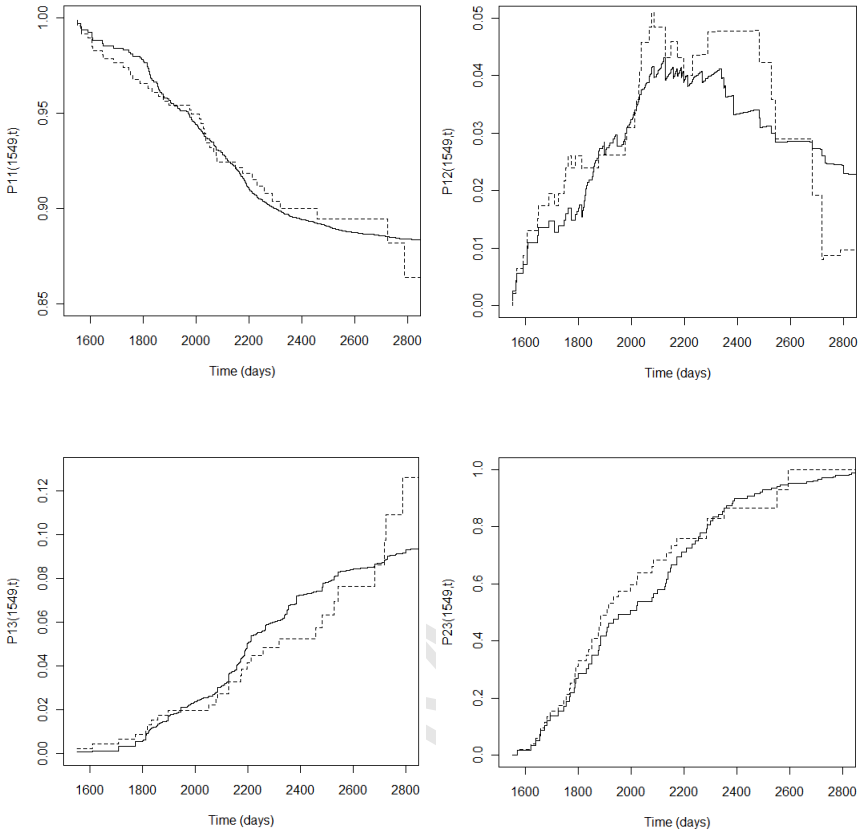


Figure 3: Estimated transition probabilities for $p_{ij}(s,t)$ with $s = 1549$ based on the Kaplan-Meier weights (dashed line) and based on presmoothed Kaplan-Meier weights (solid line). Colon cancer data.

sense of having less variance) of previous non-Markovian estimators (Meira-Machado et al. 2006).

**Acknowledgement.** The authors tank the AE and a referee for suggestions which have improved the presentation of the paper. Work supported by the Grants MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovación and 10PXIB300068PR of the Xunta de Galicia. Support from the INBIOMED project of the Xunta de Galicia (DXPCTSUG, Ref. 2009/063) is also acknowledged. Luis F. Meira-Machado acknowledges financial support by Grant PTDC/MAT/104879/2008 (FEDER support included) of the

Portuguese Ministry of Science, Technology and Higher Education. Ana P. Amorim and Luis Meira-Machado acknowledge financial support provided by the research Centre of Mathematics of the University of Minho through the FCT Pluriannual Funding Program.

**References**

Aalen, O. Johansen, S., 1978. An emperical transition matrix for nonhomogeneous Markov and chains based on censored observations. Scand J Stat 5, 141-150.

Andersen, P.K., Borgan, O, Gill, R.D, Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, New York.

Andersen, P.K., Esbjerj S., Sorensen, T.I.A., 2000. Multistate models for bleeding episodes and mortality in liver cirrhosis. Statistics in Medicine 19, 587-599

Cao, R. and Jácome, M.A., 2004. Presmoothed kernel density estimator for censored data, Journal of Nonparametric Statistics 16, 289309.

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.

Devroye, L. P., 1978a. The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory* 24, 142-151.

Devroye, L. P., 1978b. The uniform convergence of the Nadaraya-Watson regression function estimate. *Canadian Journal of Statistics* 6, 179-191.

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.

Härdle, W. and Luckhaus, S., 1984. Uniform consistency of a class of regression function estimators. Annals of Statistics 12, 612-623.

Hougaard, P., 2000. Analysis of multivariate survival data. Springer, New York.

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.

Mack, Y. P. and Silverman, B. W., 1982. Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 61, 405-415.

Meira-Machado, L., de Uña-Álvarez, J. and Cadarso-Suárez, C. 2006. Nonparametric estimation of transition probabilities in non-Markov illness-death model. Lifetime Data Anal 12, 325-344.

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.

Moertel, C.G., Fleming, T.R., McDonald, J.S. et al., 1990. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. New Engl. J. Med. 322, 352-358.

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

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

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

de Uña-Álvarez, J. and Amorim A.P., 2011. A semiparametric estimator of the bivariate distribution function for censored gap times. Biometrical Journal 53, 113-127.

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.

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

Table 2: Integrated absolute bias, integrated variance and the integrated MSE of $\hat{p}_{23}(s,.)$ along 1,000 trials, case $\theta = 1$ and $C \backsim U[0,4]$.

| | $n$ | 50 | | | 100 | | | 200 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_{23}(s,t)$ | $Method$ | $MSE$ | $BIAS$ | $VAR$ | $MSE$ | $BIAS$ | $VAR$ | $MSE$ | $BIAS$ | $VAR$ |
| $P_{23}(0.2877,t)$ | $m(.;\beta,\gamma)$ | 0.11278 | 0.33928 | 0.06426 | 0.08010 | 0.33575 | 0.03299 | 0.06054 | 0.32289 | 0.01665 |
| | $m(.;\xi)$ | 0.11800 | 0.34077 | 0.06895 | 0.08311 | 0.33317 | 0.03632 | 0.06179 | 0.31516 | 0.01950 |
| | $KM$ | 0.13334 | 0.35142 | 0.08173 | 0.09356 | 0.34754 | 0.04299 | 0.06818 | 0.32733 | 0.02322 |
| | $m$ | 0.11267 | 0.33787 | 0.06466 | 0.07973 | 0.33510 | 0.03271 | 0.06017 | 0.32254 | 0.01636 |
| $P_{23}(0.6931,t)$ | $m(.;\beta,\gamma)$ | 0.07035 | 0.21315 | 0.04952 | 0.04205 | 0.19120 | 0.02526 | 0.02784 | 0.17532 | 0.01370 |
| | $m(.;\xi)$ | 0.07708 | 0.20687 | 0.05734 | 0.04515 | 0.17437 | 0.03111 | 0.02845 | 0.14607 | 0.01846 |
| | $KM$ | 0.09932 | 0.23282 | 0.07454 | 0.05878 | 0.19997 | 0.04046 | 0.03876 | 0.18576 | 0.02297 |
| | $m$ | 0.06874 | 0.20093 | 0.05025 | 0.04073 | 0.18794 | 0.02452 | 0.02698 | 0.17190 | 0.01337 |
| $P_{23}(1.3863,t)$ | $m(.;\beta,\gamma)$ | 0.06433 | 0.11788 | 0.05405 | 0.03299 | 0.09703 | 0.02598 | 0.01696 | 0.07198 | 0.01304 |
| | $m(.;\xi)$ | 0.07135 | 0.10612 | 0.06325 | 0.03598 | 0.07266 | 0.03221 | 0.01724 | 0.03337 | 0.01645 |
| | $KM$ | 0.10830 | 0.15046 | 0.09176 | 0.05788 | 0.11084 | 0.04867 | 0.03064 | 0.07963 | 0.02579 |
| | $m$ | 0.06038 | 0.09580 | 0.05336 | 0.03057 | 0.08890 | 0.02462 | 0.01590 | 0.06653 | 0.01252 |

Table 3: Integrated absolute bias, integrated variance and the integrated MSE of $\hat{p}_{23}(s,.)$ along 1,000 trials, case $\theta = 1$ and $C \backsim U[0,3]$.

| | $n$ | 50 | | | 100 | | | 200 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_{23}(s,t)$ | $Method$ | $MSE$ | $BIAS$ | $VAR$ | $MSE$ | $BIAS$ | $VAR$ | $MSE$ | $BIAS$ | $VAR$ |
| $P_{23}(0.2877,t)$ | $m(.;\beta,\gamma)$ | 0.12218 | 0.40839 | 0.06439 | 0.09002 | 0.39064 | 0.03768 | 0.06763 | 0.36926 | 0.02120 |
| | $m(.;\xi)$ | 0.12935 | 0.43071 | 0.06386 | 0.09588 | 0.40568 | 0.03868 | 0.07262 | 0.38376 | 0.02171 |
| | $KM$ | 0.15415 | 0.45271 | 0.08112 | 0.11203 | 0.43394 | 0.04620 | 0.08408 | 0.41669 | 0.02407 |
| | $m$ | 0.12240 | 0.41434 | 0.06244 | 0.09067 | 0.40199 | 0.03473 | 0.06909 | 0.38448 | 0.01825 |
| $P_{23}(0.6931,t)$ | $m(.;\beta,\gamma)$ | 0.09040 | 0.32719 | 0.05399 | 0.05770 | 0.28552 | 0.03070 | 0.04003 | 0.25597 | 0.01879 |
| | $m(.;\xi)$ | 0.10076 | 0.35268 | 0.05780 | 0.06438 | 0.29891 | 0.03461 | 0.04519 | 0.26506 | 0.02230 |
| | $KM$ | 0.13399 | 0.38893 | 0.08022 | 0.08955 | 0.35406 | 0.04595 | 0.06444 | 0.33193 | 0.02684 |
| | $m$ | 0.08702 | 0.32171 | 0.05196 | 0.05645 | 0.29396 | 0.02767 | 0.04075 | 0.27371 | 0.01620 |
| $P_{23}(1.3863,t)$ | $m(.;\beta,\gamma)$ | 0.11456 | 0.35334 | 0.06039 | 0.07326 | 0.29224 | 0.03377 | 0.04903 | 0.24384 | 0.01876 |
| | $m(.;\xi)$ | 0.12753 | 0.37918 | 0.06622 | 0.08083 | 0.31168 | 0.03800 | 0.05308 | 0.25858 | 0.02137 |
| | $KM$ | 0.19063 | 0.45280 | 0.09973 | 0.14137 | 0.41735 | 0.06497 | 0.10093 | 0.37783 | 0.03788 |
| | $m$ | 0.10898 | 0.33798 | 0.05898 | 0.07232 | 0.29907 | 0.03151 | 0.05035 | 0.26725 | 0.01641 |