Universidade de Vigo

Presmoothing Aalen-Johansen estimator of transition probabilities

Ana Moreira, Jacobo de Uña Álvarez and Luís Meira Machado

Report 11/03

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://webs.uvigo.es/depc05/
E-mail: depc05@uvigo.es

Universidade de Vigo

Presmoothing Aalen-Johansen estimator of transition probabilities

Ana Moreira, Jacobo de Uña Álvarez and Luís Meira Machado

Report 11/03

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

Presmoothing Aalen-Johansen estimator of transition probabilities

Ana Moreira
Department of Mathematics and Applications
University of Minho
School of Sciences
Campus de Azurém, 4800-058 Guimarães, Portugal
Telephone: (+351) 253510400
Fax: (+351) 253510401
E-mail: a.moreira.cris@gmail.com

Jacobo de Uña-Álvarez
Department of Statistics and O.R.
University of Vigo, Spain

Luís Meira-Machado
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. The usual nonparametric estimator of the transition matrix for non-homogeneous Markov processes is the Aalen-Johansen estimator (Aalen and Johansen (1978)). In this paper we propose a modification of the Aalen-Johansen estimator in the illness-death model based on presmoothing. The consistency of the proposed estimators is formally established. Simulations show that the presmoothed estimators may be much more efficient than the Aalen-Johansen estimator. A real data illustration is included.

Keywords: Aalen-Johansen, Kaplan-Meier, Markov condition, Multi-state models and semiparametric censorship

1 Introduction

The analysis of survival data may be described by the Markov process with two states, 'alive' and 'dead' and a single transition between them. This is known as the multi-state mortality model. Multi-state models (Andersen et al. 1993; Meira-Machado et al. 2009) may be considered a generalization of survival analysis where survival is the ultimate outcome of interest but where intermediate (transient) states are identified. For example, in cancer studies more than one endpoint may be defined such as 'local recurrence', 'distant metastasis' and 'dead'. A simple multi-state model is obtained by splitting the 'alive' state of the mortality model into two transient states. For example, the illness-death model is fully characterized by three states and three transitions between them, see Figure 1. Graphically, multi-state models are represented by diagrams with rectangular boxes and arrows between them indicating respectively possible states and possible transitions for a given patient.

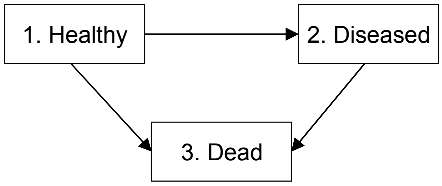

Figure 1: Illness-death model.

A multi-state model is a stochastic process $(X(t), t \in \mathcal{T})$ with a finite state space, where $X(t)$ represents the state occupied by the process at time $t \geq 0$. For two states i,j and $s < t$, introduce the so-called transition probabilities

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

Estimating these quantities is interesting, since they allow for long-term predictions of the process. The inference in multi-state models is traditionally performed under the Markov assumption, which states that past and future are independent given the present state. Aalen and Johansen (1978) introduced a nonparametric estimator of $p_{ij}(s,t)$ for non-homogeneous Markov models. Their estimation method extends the time-honored Kaplan-Meier estimator (Kaplan and Meier 1958) to Markov chains. As for the Kaplan-Meier, the standard error of the Aalen-Johansen estimator may be large when the censoring is heavy, particularly with a small sample size.

Interestingly, the variance of the Kaplan-Meier estimator may be reduced by presmoothing. The idea of presmoothing (Dikta 1998) involves replacing the censoring indicators by some smooth fit before the Kaplan-Meier formula is applied. This preliminary smoothing may be based on a certain parametric family such as the logistic (thus leading to a semiparametric estimator), or on a nonparametric estimator of the binary regression curve. Successful applications of presmoothed estimators include nonparametric curve estimation (Cao and Jácome 2004), regression analysis (de Uña Álvarez and Rodríguez-Campos 2004; Yuan 2005), and the estimation of the bivariate distribution of censored gap times (de Uña-Álvarez and Amorim 2011). The main goal of the present work is to propose a presmoothed version of the Aalen-Johansen estimator for the transition matrix of an illness-death model, and to investigate its statistical properties.

The rest of the paper is organized as follows. In Section 2 we introduce the new estimator and we formally establish its consistency. In Section 3 we compare by simulations the proposed estimator to the original Aalen-Johansen curve. In Section 4 we illustrate the proposed method using data from the Stanford heart transplant study. Main conclusions are reported in Section 5. The Appendix contains the technical proofs.

2 The estimator: main result

In this paper we consider the (progressive) illness-death model depicted in Figure 1. We assume that all subjects are in State 1 ('healthy') at time $t = 0$, and that they may either visit State 2 ('diseased') at some time point; or not, going directly to the absorbing ('dead') state. Given two time points $s < t$, there are in essence three different transition probabilities to estimate: $p_{11}(s,t)$, $p_{12}(s,t)$, and $p_{22}(s,t)$. The two other transition probabilities ($p_{13}(s,t)$ and $p_{23}(s,t)$) can be obtained from $p_{13}(s,t) = 1-p_{11}(s,t)-p_{12}(s,t)$ and $p_{23}(s,t) = 1 - p_{22}(s,t)$.

The irreversible illness-death model is fully characterized by three transitions represented by the arrows in Figure 1. Let T_{ij} denote the potential transition time from state i to state j. In this model we have two competing transitions $1 \to 2$ and $1 \to 3$. Therefore, we denote by $\rho = I(T_{12} \leq T_{13})$ the indicator of visiting state 2 at some time, and introduce $Z = T_{12} \wedge T_{13}$, the sojourn time in state 1. A subject visiting State 2 will arrive at the absorbing 'dead' state at time $T_{12}+T_{23}$, while this time will be T_{13} for those not visiting State 2 (i.e. $\rho = 0$). Finally, let $T = Z + \rho T_{23}$ denote the total survival time of the process. However, because of follow-up limitations, lost cases and so on, rather than (Z, T, ρ) one observes $(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta, \Delta_1\rho)$ where $\widetilde{Z} = Z \wedge C$, $\widetilde{T} = T \wedge C$, $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$. Here C denotes the potential censoring time, which we assume to be independent of the process (that is, C and (Z,T) are assumed to be independent). Under continuity, the information provided by $\triangle_1\rho$ is superfluous since we have $\triangle_1\rho = I(\widetilde{Z} < \widetilde{T})$. With this notation, the transition probabilities are written as

$$
\begin{aligned}
p_{11}(s,t) &= \frac{P(Z > t)}{P(Z > s)}, \qquad p_{12}(s,t) = \frac{P(s < Z \leq t, T > t)}{P(Z > s)}, \\
p_{22}(s,t) &= \frac{P(Z \leq s, T > t)}{P(Z \leq s, T > s)}.
\end{aligned}
$$

Under the Markov assumption, all these quantities are estimated non-parametrically using Aalen-Johansen estimators. Explicit formulae of the Aalen-Johansen estimator for the illness-death model are available (Borgan 1998). Here we give alternative expressions for this estimator suitable to motivate our method of presmoothing below.

Assume that we have a sample of n individuals from the population under study. Let $(\widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i, \Delta_{1i}\rho_i)$, $i = 1, ..., n$ be the corresponding sampling information. The Aalen-Johansen estimate of the transition probability $p_{11}(s,t)$ is the Kaplan-Meier estimator

$$\widehat{p}_{11}^{AJ}(s,t) = \prod_{s<\widetilde{Z}_i\leq t}\left[1 - \frac{\Delta_{1i}}{n\widetilde{M}_{0n}(\widetilde{Z}_i)}\right] \quad (1)$$

where

$$\widetilde{M}_{0n}(y) = \frac{1}{n}\sum_{i=1}^{n} I(\widetilde{Z}_i \geq y).$$

Then, Aalen-Johansen estimate of the transition probability $p_{22}(s,t)$ is the Kaplan-Meier estimator

$$\widehat{p}_{22}^{AJ}(s,t) = \prod_{s<\widetilde{T}_i\leq t, \widetilde{Z}_i<\widetilde{T}_i}\left[1 - \frac{\Delta_i}{n\widetilde{M}_{1n}(\widetilde{T}_i)}\right] \quad (2)$$

where

$$\widetilde{M}_{1n}(y) = \frac{1}{n}\sum_{i=1}^{n} I(\widetilde{Z}_i < y \leq \widetilde{T}_i).$$

Finally, the estimator for $p_{12}(s,t)$ is given by

$$\widehat{p}_{12}^{AJ}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{p}_{11}^{AJ}(s,\widetilde{Z}_i^-)\widehat{p}_{22}^{AJ}(\widetilde{Z}_i,t)I(s<\widetilde{Z}_i\leq t, \widetilde{Z}_i<\widetilde{T}_i)}{n\widetilde{M}_{0n}(\widetilde{Z}_i)} \quad (3)$$

where

$$\widehat{p}_{11}^{AJ}(s,t^-) = lim_{u\uparrow t}\widehat{p}_{11}^{AJ}(s,u)$$

Now, we discuss how to introduce modified estimators based on presmoothing. Presmoothing the Aalen-Johansen (AJ) involves replacing the censoring indicators (in the transition probabilities $p_{11}(s,t)$ and $p_{22}(s,t)$) by a smooth fit. The presmoothed version of $p_{11}(s,t)$ is obtained by replacing the Δ_{1i}'s in (1) by some smooth fit to the binary regression function

$m_0(z) = P\left(\Delta_1 = 1 | \widetilde{Z} = z\right)$ (see e.g. Dikta 1998). Then, the corresponding presmoothed Aalen-Johansen (P-AJ) estimator is given by

$$\widetilde{p}_{11}^{PAJ}(s,t) = \prod_{s<\widetilde{Z}_i \le t} \left[1 - \frac{m_{0n}(\widetilde{Z}_i)}{n\widetilde{M}_{0n}(\widetilde{Z}_i)}\right] \tag{4}$$

where $m_{0n}(z)$ stands for an estimator of the binary regression function $m_0(z)$. Then, $m_0(\widetilde{Z})$ is the conditional probability of censoring on Z given $\widetilde{Z}$. Since the pair $\widetilde{Z}, \Delta_1$ is observable, the function $m_0(z)$ can be estimated by standard methods. For example, logistic regression may be performed. Consider now the presmoothed version of (2) given by

$$\widetilde{p}_{22}^{PAJ}(s,t) = \prod_{s<\widetilde{T}_i \le t, \widetilde{Z}_i < \widetilde{T}_i} \left[1 - \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)}\right] \tag{5}$$

where $m_{1n}(z,t)$ stands for an estimator of the binary regression function $m_1(z,t) = P\left(\Delta = 1 | \widetilde{Z} = z, \widetilde{T} = t, \Delta_1\rho = 1\right)$. Then, $m_1(\widetilde{Z}, \widetilde{T})$ is the conditional probability of censoring on T given $(\widetilde{Z}, \widetilde{T})$ and given that transition $1 \to 2$ is observed ($\Delta_1\rho = 1$). Amorim et al. (2011) discussed the role of the function $m_1(z,t)$ as a suitable presmoothing strategy for $p_{22}(s,t)$; although these authors considered a different context in which the Markov assumption may not hold, their discussion on the presmoothing issue remains valid here. As before, $\widetilde{Z}, \widetilde{T}, \Delta$ and $\Delta_1\rho$ are observable, allowing the function $m_1(z,t)$ to be estimated by standard methods. Finally the transition probability $p_{12}(s,t)$ can be estimated by plugging (4) and (5) into equation (3).

The estimator $m_{0n}(z)$ is based on the whole sample, while $m_{1n}(z,t)$ is based on the subsample $i : \Delta_{1i}\rho_i = 1$. We assume that these two empirical functions approximate well their targets in a uniform sense; more specifically, set

$$U_1: \ \sup_z |m_{0n}(z) - m_0(z)| \to 0 \qquad \text{w. p. 1},$$

and

$$U_2: \ \sup_{z,t} |m_{1n}(z,t) - m_1(z,t)| \to 0 \qquad \text{w. p. 1}.$$

Conditions under which U_1 and U_2 can be fulfilled were investigated in a number of papers, including Dikta (1998, 2000), Devroye (1978a,b), Mack and Silverman (1982) and Härdle and Luckhaus (1984). The uniform consistency of $\widehat{p}_{11}^{PAJ}(s,t)$ will hold on $0 \le s < t \le \tau$, where τ is strictly smaller than the upper bound of the support of $\widetilde{Z}$. Put $\widetilde{M}_1(y) = P(\widetilde{Z} < y \le \widetilde{T})$. For the uniform consistency of $\widehat{p}_{22}^{PAJ}(s,t)$ and $\widehat{p}_{12}^{PAJ}(s,t)$ we will refer to the following assumption:

$$M: \quad \widetilde{M}_1 \text{ is bounded from below on } [\tau_0, \tau_1].$$

This condition allows to handle some denominators which appear in the proofs. It can be interpreted as a 'non empty risk set' assumption for the transition from State 2 to State 3. By force, $\tau_0 > 0$, while τ_1 is (similarly as for τ) strictly smaller than the upper bound of the support of $\widetilde{T}$. We have the following result. The proof is deferred to the Appendix.

Theorem 1. (a) Under U_1 we have w. p. 1

$$\sup_{0 \le s < t \le \tau} \left| \widehat{p}_{11}^{PAJ}(s,t) - p_{11}(s,t) \right| \to 0.$$

(b) Besides, under U_2 and M, we have w. p. 1

$$\sup_{\tau_0 \le s < t \le \tau_1} \left| \widehat{p}_{22}^{PAJ}(s,t) - p_{22}(s,t) \right| \to 0.$$

(c) Finally, under U_1, U_2 and M we have w. p. 1

$$\sup_{\tau_0 \le s < t \le \tau} \left| \widehat{p}_{12}^{PAJ}(s,t) - p_{12}(s,t) \right| \to 0.$$

3 Simulation study

In this section, we compare by simulations the presmoothed Aalen-Johansen estimator for the transition probabilities to the original Aalen-Johansen estimator. More specifically, the AJ and P-AJ type estimators $\widehat{p}_{11}(s,t)$, $\widehat{p}_{12}(s,t)$ and $\widehat{p}_{22}(s,t)$ introduced in Section 2 are considered. As presmoothing function we always take a parametric (logistic) family, so we actually have a semiparametric Aalen-Johansen estimator.

To simulate the data in the illness-death model, we followed the work of Amorim et al. (2011), which contains, among other things, a comprehensive simulation study. We assume that all individuals are in State 1 ("healthy") at time $t = 0$. Therefore, the patients history (or course) may be divided into two groups according to whether the disease occurred (that is, passing through State 2) ($1 \to 2 \to 3$) or not ($1 \to 3$). We separately consider these two possible subgroups of individuals. For the first subgroup of individuals ($\rho = 1$), the successive gap times $(Z, T - Z)$ are simulated according to the bivariate distribution

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

with unit exponential margins. The parameter θ controls for the amount of dependency between the gap times $(Z, T - Z)$ and was set to 0 and 1, corresponding to 0 and 0.25 correlation between Z and $T - Z$. For the second subgroup of individuals ($\rho = 0$), the value of Z is simulated according to an exponential with rate parameter 1. We also consider that 70% of the individuals were in the first group. Details about the simulation procedure can be obtained in Amorim et al. (2011).

The follow-up time was subjected to right censoring, C, according to uniform 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

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

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, namely 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. For $m_0(z) = P\left(\Delta_1 = 1 | \widetilde{Z} = t\right)$, we have

$$m_0(z) = \frac{1}{1 + \eta_0(t)}, \qquad \text{where} \;\; \eta_0(t) = \frac{\lambda_G(t)}{\lambda_Z(t)}$$

and where $\lambda_Z(t)$ stands for the hazard function of Z.

Similarly as above, we also perform logistic presmoothing for the function $m_0(z)$, with the variable $\widetilde{Z}$ transformed by $-ln(\tau_G - \widetilde{Z})$. This function belongs to the logistic family with some preliminary transformation. To estimate the function $m_0(z)$ in the simulations, we fit the logistic model

$$m_0(z;\gamma) = \frac{1}{1+\exp(\gamma_0 + \gamma_1 \ln(\eta_0(t)))}$$

which contains the true presmoothing function m_0 as a special case ($\gamma_0 = 0$,$\gamma_1 = 1$).

The β parameter in model $m_1(.;\beta)$ is estimated via maximization of the conditional likelihood of the Δ_i's given the $(\widetilde{Z}_i, \widetilde{T}_i)$'s, for those subjects with $\Delta_1\rho = 1$ (see Dikta (1998, 2000)). Similarly, the γ parameter in model $m_0(.;\gamma)$ is estimated via maximization of the conditional likelihood of the Δ_{1i}'s given the $\widetilde{Z}_i$'s. Note that the β parameter is needed for estimating $p_{22}(s,t)$ and $p_{12}(s,t)$, while γ enters the estimation of $p_{11}(s,t)$ and (again) $p_{12}(s,t)$. The aim of this simulation study is to compare the Aalen-Johansen estimator (1978) and the new estimator based on presmoothing (P-AJ). Again, for measuring the estimates relative performance, we followed the work of Amorim et al. (2011). As in Amorim et al. (2011), we computed the integrated absolute bias, integrated variance and the integrated MSE of the estimates. For each simulated setting ($\theta = 0$ and $\theta = 1$) we derived the analytic expression of $p_{ij}(s,t)$ so that the bias and the MSE of the estimator could be examined. K = 1000 data sets were generated, with three different sample sizes $n = 50$, $n = 100$ and $n = 200$.

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

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

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

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

To summarize the results we also calculated the integrated absolute bias (BIAS), integrated variance (VAR) and the integrated MSE (IMSE), 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 given in Tables 2 to 5 were obtained by numerical integration on the interval $[s, t_1]$ with $t_1 = 4$, taking a grid of step $\delta = 0.05$.

In Tables 2 to 5 we report the results for the summary statistics attained by the proposed estimator when based on several presmoothing functions (P-AJ), for all scenarios. In all tables, the row labeled with m corresponds to presmoothing with the true function which is unrealistic in practice, because this function will be typically unknown. However, this row represents

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.

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 $\widetilde{Z} = z, \widetilde{T} = t$ shown before. 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 gap times. This is labeled with $m(.,\xi)$. Note that the true m does not belong to this parametric family. Finally, we also report the results pertaining to the Aalen-Johansen estimator, which corresponds to the situation with no presmoothing at all. This is labeled in the Tables as AJ.

It is obvious from the analysis of Tables 2 to 5, that presmoothing leads to estimators with smaller variance and thus attaining better results with regard to the integrated MSE. As expected, the (integrated) MSE, bias and variance of the estimated transition probabilities always decrease with an increasing sample size, while they increase with the censoring degree. The estimator which makes use of the true m is the one with the best performance. However, this estimator is unrealistic since in practice one has to estimate the function m. In general, the lowest errors among the realistic versions of the estimators correspond to the estimator based on the correctly specified parametric family, $m(.;\beta,\gamma)$. However, the presmoothed estimator based on the wrong parametric model $m(.;\xi)$ is still (much) better than AJ. This means that it is worthwhile doing some presmoothing even when we are not completely sure about the parametric family.

Results shown in the Tables 2 to 5 support the idea that presmoothing leads to variance improvement. When compared to the estimators based on presmoothing, the relative efficiency (defined as the quotient between the two integrated MSEs) of the Aalen-Johansen estimator is always below 1. For higher values of s, where the censoring effects are stronger, the relative efficiency can drop below 50%. These findings agree with the results obtained by Amorim et al. (2011) and support the intuition that the use of presmoothing for the estimation of transition probabilities will be more clearly seen in the presence of large censoring degrees.

Tables 2 and 3 show a systematic bias for all estimators of the transition probabilities $p_{12}(s,t)$ and $p_{22}(s,t)$. This is because these tables report the results attained when generating data from a dependency scenario and therefore reflects the failure of the Markov assumption. To illustrate these features we present in Figures 2 to 7 the graphical average results for the two methods (AJ and P-AJ corresponding to presmoothing via standard logistic models, $m(.,\xi)$). These figures plot the data generating functions and pointwise 95% oscillation limits of the estimates $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{22}(s,t)$, for sample sizes of $n = 200$ with percentages of censored data obtained using $C \sim U[0,3]$. The good performance of the resulting estimates (for both methods) is evident for independent gap times ($\theta = 0$), recovering the functional forms of the corresponding true curves very successfully. However, a systematic bias of $p_{12}(s,t)$ and $p_{22}(s,t)$ in the dependent scenario ($\theta = 1$) is also clear, see Figures 4 and 6. This bias is much more evident when s is large, in agreement with the amount of false information introduced by the Markov condition (which increases with s). In all scenarios, the use of the presmoothing yields estimators with less variability.

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

	n	50			100			200		
$P_{ij}(s,t)$	*Method*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*
$P_{11}(0.2877,t)$	$m(.;\beta,\gamma)$	0.01864	0.04079	0.01769	0.00878	0.01909	0.00855	0.00452	0.01357	0.00443
	$m(.;\xi)$	0.01878	0.04246	0.01800	0.00883	0.02126	0.00868	0.00460	0.01582	0.00452
	AJ	0.02123	0.02158	0.02092	0.01028	0.00955	0.01022	0.00537	0.00800	0.00533
	m	0.01312	0.02146	0.01280	0.00665	0.01079	0.00656	0.00344	0.00671	0.00342
$P_{12}(0.2877,t)$	$m(.;\beta,\gamma)$	0.02174	0.03026	0.02141	0.01121	0.02500	0.01100	0.00612	0.02916	0.00584
	$m(.;\xi)$	0.02269	0.02669	0.02243	0.01170	0.02092	0.01153	0.00632	0.02470	0.00609
	AJ	0.02702	0.02891	0.02677	0.01393	0.02727	0.01370	0.00732	0.03171	0.00701
	m	0.01881	0.02859	0.01857	0.00994	0.02612	0.00972	0.00547	0.03169	0.00516
$P_{22}(0.2877,t)$	$m(.;\beta,\gamma)$	0.04065	0.18028	0.03067	0.02499	0.18808	0.01403	0.01759	0.18551	0.00678
	$m(.;\xi)$	0.04094	0.17961	0.03104	0.02509	0.18810	0.01419	0.01752	0.18515	0.00682
	AJ	0.04237	0.16216	0.03398	0.02599	0.18096	0.01567	0.01812	0.18317	0.00752
	m	0.03502	0.16667	0.02628	0.02215	0.18047	0.01192	0.01635	0.18258	0.00577
$P_{11}(0.6931,t)$	$m(.;\beta,\gamma)$	0.03168	0.05996	0.02947	0.01404	0.02708	0.01352	0.00734	0.01871	0.00713
	$m(.;\xi)$	0.03197	0.06149	0.03016	0.01416	0.03022	0.01455	0.00747	0.01962	0.00734
	AJ	0.03750	0.03148	0.03677	0.01738	0.01321	0.01724	0.00907	0.01069	0.00898
	m	0.02099	0.03053	0.02026	0.01061	0.01558	0.01040	0.00540	0.00855	0.00534
$P_{12}(0.6931,t)$	$m(.;\beta,\gamma)$	0.03353	0.05172	0.03256	0.01739	0.05133	0.01644	0.00994	0.05500	0.00882
	$m(.;\xi)$	0.03502	0.04245	0.03435	0.01803	0.04047	0.01740	0.01003	0.04502	0.00926
	AJ	0.04290	0.05486	0.04186	0.02212	0.05527	0.02104	0.01204	0.05855	0.01080
	m	0.02989	0.05345	0.02886	0.01623	0.05223	0.01526	0.00934	0.05884	0.00810
$P_{22}(0.6931,t)$	$m(.;\beta,\gamma)$	0.04377	0.16461	0.03463	0.02471	0.15916	0.01617	0.01634	0.15786	0.00791
	$m(.;\xi)$	0.04482	0.16395	0.03568	0.02515	0.15866	0.01656	0.01652	0.15726	0.00806
	AJ	0.05003	0.14281	0.04313	0.02702	0.14921	0.01949	0.01738	0.15153	0.00956
	m	0.03403	0.14646	0.02685	0.02029	0.15018	0.01264	0.01438	0.15295	0.00641
$P_{11}(1.3863,t)$	$m(.;\beta,\gamma)$	0.07510	0.10977	0.06691	0.03363	0.05112	0.03160	0.01740	0.03539	0.01659
	$m(.;\xi)$	0.07165	0.09970	0.06577	0.03213	0.04383	0.03119	0.01680	0.02807	0.01647
	AJ	0.09922	0.06458	0.09584	0.04451	0.02572	0.04387	0.02321	0.01962	0.02288
	m	0.04581	0.06145	0.04256	0.02268	0.03088	0.02176	0.01152	0.01697	0.01126
$P_{12}(1.3863,t)$	$m(.;\beta,\gamma)$	0.06659	0.07348	0.06401	0.03530	0.08320	0.03225	0.02043	0.08684	0.01714
	$m(.;\xi)$	0.06926	0.06745	0.06722	0.03643	0.07357	0.03415	0.02048	0.07735	0.01789
	AJ	0.08594	0.08094	0.08282	0.04449	0.08388	0.04140	0.02468	0.08903	0.02121
	m	0.06411	0.07731	0.06128	0.03538	0.07969	0.03259	0.02058	0.08970	0.01706
$P_{22}(1.3863,t)$	$m(.;\beta,\gamma)$	0.07104	0.15190	0.05960	0.03372	0.12455	0.02667	0.01881	0.11085	0.01328
	$m(.;\xi)$	0.07763	0.16072	0.06520	0.03687	0.13391	0.02867	0.02084	0.12039	0.01417
	AJ	0.09115	0.11812	0.08482	0.04292	0.10587	0.03798	0.02227	0.09872	0.01794
	m	0.04746	0.11902	0.04076	0.02412	0.10993	0.01875	0.01413	0.09979	0.00972

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

	n	50			100			200		
$P_{ij}(s,t)$	*Method*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*
$P_{11}(0.2877,t)$	$m(.;\beta,\gamma)$	0.02953	0.10624	0.02315	0.01473	0.07644	0.01102	0.00789	0.05496	0.00581
	$m(.;\xi)$	0.02632	0.09326	0.02210	0.01188	0.05811	0.01025	0.00571	0.03641	0.00514
	AJ	0.03275	0.07520	0.02880	0.01738	0.05731	0.01481	0.00960	0.04389	0.00799
	m	0.01576	0.06984	0.01220	0.00829	0.05236	0.00603	0.00476	0.04450	0.00316
$P_{12}(0.2877,t)$	$m(.;\beta,\gamma)$	0.03195	0.07673	0.02826	0.01770	0.06396	0.01549	0.01073	0.05867	0.00875
	$m(.;\xi)$	0.03225	0.06543	0.02951	0.01670	0.04196	0.01565	0.00923	0.03519	0.00859
	AJ	0.04214	0.07597	0.03878	0.02353	0.06286	0.02163	0.01424	0.05894	0.01241
	m	0.02367	0.06837	0.02104	0.01404	0.06356	0.01204	0.00913	0.06058	0.00727
$P_{22}(0.2877,t)$	$m(.;\beta,\gamma)$	0.05085	0.23753	0.03434	0.02842	0.21001	0.01528	0.02056	0.20695	0.00784
	$m(.;\xi)$	0.05044	0.23121	0.03479	0.02757	0.20100	0.01533	0.01920	0.19292	0.00777
	AJ	0.05321	0.20781	0.04065	0.02933	0.19422	0.01801	0.02112	0.19866	0.00934
	m	0.03757	0.20957	0.02484	0.02325	0.19957	0.01140	0.01729	0.19522	0.00594
$P_{11}(0.6931,t)$	$m(.;\beta,\gamma)$	0.05636	0.16091	0.04151	0.02877	0.11555	0.02023	0.01530	0.08120	0.01059
	$m(.;\xi)$	0.04874	0.14156	0.03894	0.02217	0.08628	0.01845	0.01026	0.05163	0.00902
	AJ	0.06414	0.11352	0.05495	0.03437	0.08654	0.02848	0.01884	0.06642	0.01514
	m	0.02725	0.10585	0.01898	0.01502	0.07888	0.00985	0.00876	0.06510	0.00515
$P_{12}(0.6931,t)$	$m(.;\beta,\gamma)$	0.04722	0.08383	0.04321	0.02693	0.07881	0.02411	0.01647	0.07472	0.01388
	$m(.;\xi)$	0.04795	0.06495	0.04518	0.02577	0.05767	0.02435	0.01470	0.05209	0.01351
	AJ	0.06564	0.08976	0.06141	0.03744	0.08014	0.03469	0.02259	0.07569	0.02003
	m	0.03907	0.08358	0.03571	0.02342	0.08059	0.02059	0.01528	0.07817	0.01260
$P_{22}(0.6931,t)$	$m(.;\beta,\gamma)$	0.07295	0.25545	0.04772	0.04069	0.22121	0.02272	0.02646	0.20766	0.01088
	$m(.;\xi)$	0.07299	0.24931	0.04976	0.03866	0.21119	0.02291	0.02316	0.18830	0.01075
	AJ	0.07732	0.20713	0.06053	0.04333	0.19808	0.02917	0.02816	0.19499	0.01456
	m	0.04427	0.20789	0.02782	0.02715	0.19887	0.01286	0.01935	0.18761	0.00665
$P_{11}(1.3863,t)$	$m(.;\beta,\gamma)$	0.15828	0.29155	0.10488	0.08415	0.21623	0.05218	0.04715	0.15201	0.02922
	$m(.;\xi)$	0.11915	0.22087	0.08857	0.05278	0.12627	0.04157	0.02464	0.06960	0.02130
	AJ	0.20944	0.23542	0.16998	0.11095	0.17679	0.08648	0.06199	0.13227	0.04708
	m	0.07598	0.21746	0.04090	0.04199	0.15815	0.02099	0.02568	0.12626	0.01136
$P_{12}(1.3863,t)$	$m(.;\beta,\gamma)$	0.08819	0.07494	0.08539	0.05252	0.07908	0.04934	0.03167	0.07506	0.02885
	$m(.;\xi)$	0.08883	0.07613	0.08580	0.05214	0.09063	0.04745	0.03181	0.09885	0.02603
	AJ	0.12562	0.07176	0.12305	0.07381	0.08038	0.07056	0.04413	0.07355	0.04143
	m	0.09009	0.07455	0.08731	0.05502	0.07825	0.05194	0.03535	0.07559	0.03251
$P_{22}(1.3863,t)$	$m(.;\beta,\gamma)$	0.16509	0.35887	0.09329	0.08891	0.25952	0.04873	0.05575	0.23567	0.02312
	$m(.;\xi)$	0.16617	0.34178	0.10439	0.07845	0.23345	0.04957	0.04181	0.19140	0.02324
	AJ	0.20923	0.29425	0.15625	0.10352	0.21472	0.07417	0.06376	0.21087	0.03698
	m	0.08661	0.28003	0.03956	0.04982	0.22075	0.01907	0.03531	0.20314	0.00957

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

	n	50			100			200		
$P_{ij}(s,t)$	$Method$	$IMSE$	$BIAS$	VAR	$IMSE$	$BIAS$	VAR	$IMSE$	$BIAS$	VAR
$P_{11}(0.1438410,t)$	$m(.;\beta,\gamma)$	0.00838	0.02707	0.00809	0.00402	0.01428	0.00393	0.00199	0.00884	0.00196
	$m(.;\xi)$	0.00834	0.02676	0.00807	0.00400	0.01280	0.00393	0.00198	0.00754	0.00196
	AJ	0.00919	0.01602	0.00910	0.00442	0.00933	0.00438	0.00219	0.00603	0.00217
	m	0.00712	0.01665	0.00701	0.00360	0.00924	0.00357	0.00178	0.00589	0.00177
$P_{12}(0.1438410,t)$	$m(.;\beta,\gamma)$	0.01373	0.02980	0.01327	0.00695	0.01800	0.00681	0.00332	0.00827	0.00327
	$m(.;\xi)$	0.01388	0.02675	0.01353	0.00705	0.01811	0.00695	0.00338	0.00883	0.00335
	AJ	0.01509	0.01858	0.01494	0.00771	0.01066	0.00766	0.00375	0.00444	0.00374
	m	0.01096	0.01962	0.01079	0.00587	0.01081	0.00581	0.00291	0.00480	0.00290
$P_{22}(0.1438410,t)$	$m(.;\beta,\gamma)$	0.03485	0.03926	0.03406	0.01570	0.02428	0.01547	0.00816	0.01279	0.00808
	$m(.;\xi)$	0.03524	0.03541	0.03460	0.01587	0.02268	0.01570	0.00821	0.01549	0.00815
	AJ	0.03825	0.02212	0.03803	0.01761	0.01110	0.01756	0.00907	0.00601	0.00906
	m	0.02648	0.02187	0.02625	0.01260	0.01044	0.01254	0.00666	0.00686	0.00663
$P_{11}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.01361	0.04000	0.01295	0.00651	0.02182	0.00631	0.00315	0.01291	0.00309
	$m(.;\xi)$	0.01354	0.03944	0.01292	0.00648	0.01945	0.00631	0.00314	0.01082	0.00309
	AJ	0.01526	0.02288	0.01505	0.00724	0.01392	0.00716	0.00355	0.00833	0.00352
	m	0.01121	0.02422	0.01095	0.00572	0.01372	0.00564	0.00279	0.00816	0.00276
$P_{12}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.01897	0.03421	0.01836	0.00941	0.01940	0.00923	0.00452	0.00872	0.00446
	$m(.;\xi)$	0.01926	0.02938	0.01881	0.00959	0.02009	0.00945	0.00461	0.00891	0.00458
	AJ	0.02111	0.02110	0.02090	0.01053	0.01259	0.01045	0.00513	0.00495	0.00512
	m	0.01563	0.02167	0.01540	0.00815	0.01290	0.00807	0.00404	0.00487	0.00402
$P_{22}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.03453	0.04611	0.03336	0.01648	0.02898	0.01612	0.00836	0.01563	0.00824
	$m(.;\xi)$	0.03496	0.04375	0.03398	0.01673	0.02831	0.01644	0.00848	0.01709	0.00838
	AJ	0.03879	0.02603	0.03845	0.01874	0.01389	0.01865	0.00959	0.00743	0.00956
	m	0.02506	0.02703	0.02468	0.01251	0.01212	0.01241	0.00659	0.00781	0.00655
$P_{11}(0.6931472,t)$	$m(.;\beta,\gamma)$	0.03292	0.07945	0.03021	0.01543	0.04093	0.01460	0.00703	0.02497	0.00676
	$m(.;\xi)$	0.03237	0.07819	0.02985	0.01521	0.03796	0.01453	0.00694	0.02117	0.00675
	AJ	0.03878	0.04530	0.03786	0.01758	0.02712	0.01724	0.00823	0.01612	0.00812
	m	0.02502	0.04954	0.02393	0.01280	0.02740	0.01244	0.00605	0.01610	0.00594
$P_{12}(0.6931472,t)$	$m(.;\beta,\gamma)$	0.03348	0.04312	0.03259	0.01716	0.02217	0.01688	0.00796	0.00992	0.00788
	$m(.;\xi)$	0.03406	0.03520	0.03345	0.01751	0.02291	0.01733	0.00814	0.00865	0.00812
	AJ	0.03699	0.03089	0.03663	0.01911	0.01463	0.01900	0.00905	0.00719	0.00903
	m	0.02800	0.03097	0.02760	0.01559	0.01603	0.01545	0.00736	0.00645	0.00733
$P_{22}0.6931472,t)$	$m(.;\beta,\gamma)$	0.04656	0.06709	0.04407	0.02035	0.03591	0.01967	0.01041	0.02166	0.01016
	$m(.;\xi)$	0.04775	0.06316	0.04563	0.02092	0.03790	0.02037	0.01078	0.02540	0.01056
	AJ	0.05389	0.03560	0.05318	0.02449	0.01699	0.02431	0.01249	0.01136	0.01242
	m	0.03105	0.03742	0.03025	0.01475	0.01626	0.01455	0.00791	0.01171	0.00783

Table 5: Integrated absolute bias, integrated variance and the integrated MSE of $\hat{p}_{ij}(s,.)$ along 1,000 trials, case $\theta = 0$ and $C \backsim U[0,3]$.

	n	50			100			200		
$P_{ij}(s,t)$	*Method*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*	*IMSE*	*BIAS*	*VAR*
$P_{11}(0.1438410,t)$	$m(.;\beta,\gamma)$	0.01011	0.05216	0.00903	0.00465	0.02690	0.00431	0.00232	0.01836	0.00218
	$m(.;\xi)$	0.00987	0.05049	0.00890	0.00448	0.02401	0.00422	0.00222	0.01329	0.00214
	AJ	0.01114	0.03199	0.01072	0.00523	0.01876	0.00506	0.00264	0.01474	0.00255
	m	0.00721	0.03439	0.00671	0.00358	0.01954	0.00339	0.00188	0.01396	0.00178
$P_{12}(0.14384410,t)$	$m(.;\beta,\gamma)$	0.01794	0.06344	0.01539	0.01053	0.05551	0.00851	0.00562	0.03775	0.00458
	$m(.;\xi)$	0.01721	0.05431	0.01536	0.00933	0.04606	0.00819	0.00465	0.02610	0.00429
	AJ	0.02001	0.04733	0.01853	0.01182	0.04105	0.01052	0.00653	0.03120	0.00571
	m	0.01256	0.05168	0.01079	0.00712	0.03777	0.00591	0.00401	0.03306	0.00312
$P_{22}(0.14384410,t)$	$m(.;\beta,\gamma)$	0.04573	0.09080	0.04120	0.02252	0.07677	0.01925	0.01074	0.04930	0.00916
	$m(.;\xi)$	0.04442	0.08009	0.04103	0.02068	0.06389	0.01878	0.00936	0.03394	0.00880
	AJ	0.05071	0.05839	0.04834	0.02515	0.05380	0.02319	0.01271	0.04037	0.01150
	m	0.03039	0.06363	0.02755	0.01440	0.05175	0.01255	0.00727	0.03926	0.00602
$P_{11}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.01766	0.07802	0.01519	0.00813	0.04089	0.00733	0.00388	0.02710	0.00354
	$m(.;\xi)$	0.01716	0.07535	0.01495	0.00777	0.03661	0.00716	0.00364	0.01933	0.00346
	AJ	0.01946	0.04699	0.01853	0.00938	0.02837	0.00898	0.00450	0.02088	0.00430
	m	0.01185	0.05129	0.01071	0.00601	0.02955	0.00557	0.00309	0.02069	0.00288
$P_{12}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.02448	0.07199	0.02112	0.01451	0.06295	0.01180	0.00767	0.04143	0.00628
	$m(.;\xi)$	0.02359	0.05989	0.02117	0.01296	0.05208	0.01142	0.00634	0.02957	0.00587
	AJ	0.02802	0.05509	0.02599	0.01651	0.04770	0.01472	0.00898	0.03519	0.00786
	m	0.01787	0.05939	0.01548	0.01006	0.04377	0.00840	0.00558	0.03673	0.00437
$P_{22}(0.3465736,t)$	$m(.;\beta,\gamma)$	0.04739	0.10555	0.04069	0.02563	0.09178	0.02073	0.01244	0.05873	0.01007
	$m(.;\xi)$	0.04573	0.09499	0.04068	0.02300	0.07892	0.02004	0.01041	0.04130	0.00953
	AJ	0.05282	0.07153	0.04933	0.02863	0.06451	0.02571	0.01447	0.04638	0.01268
	m	0.02883	0.07802	0.02464	0.01509	0.05921	0.01235	0.00787	0.04693	0.00599
$P_{11}(0.6931472,t)$	$m(.;\beta,\gamma)$	0.05074	0.16039	0.04031	0.02131	0.08110	0.01808	0.00968	0.05098	0.00834
	$m(.;\xi)$	0.04802	0.15437	0.03876	0.01959	0.07031	0.01720	0.00862	0.03673	0.00789
	AJ	0.05762	0.10097	0.05343	0.02562	0.05671	0.02394	0.01161	0.03948	0.01081
	m	0.02894	0.10975	0.02389	0.01380	0.05862	0.01200	0.00720	0.03949	0.00637
$P_{12}(0.6931472,t)$	$m(.;\beta,\gamma)$	0.04507	0.08466	0.04044	0.02650	0.07508	0.02238	0.01397	0.04836	0.01182
	$m(.;\xi)$	0.04357	0.06638	0.04038	0.02376	0.05956	0.02157	0.01158	0.03460	0.01090
	AJ	0.05378	0.06815	0.05080	0.03093	0.05686	0.02813	0.01627	0.04390	0.01452
	m	0.03633	0.07501	0.03270	0.01927	0.05530	0.01666	0.01094	0.04572	0.00899
$P_{22}(0.6931472,t)$	$m(.;\beta,\gamma)$	0.06945	0.15333	0.05520	0.03880	0.12909	0.02881	0.01897	0.08283	0.01421
	$m(.;\xi)$	0.06660	0.13952	0.05570	0.03408	0.11298	0.02785	0.01517	0.06032	0.01330
	AJ	0.07825	0.10236	0.07088	0.04380	0.09094	0.03791	0.02278	0.06504	0.01923
	m	0.03859	0.10977	0.02990	0.02029	0.08309	0.01481	0.01138	0.06728	0.00761

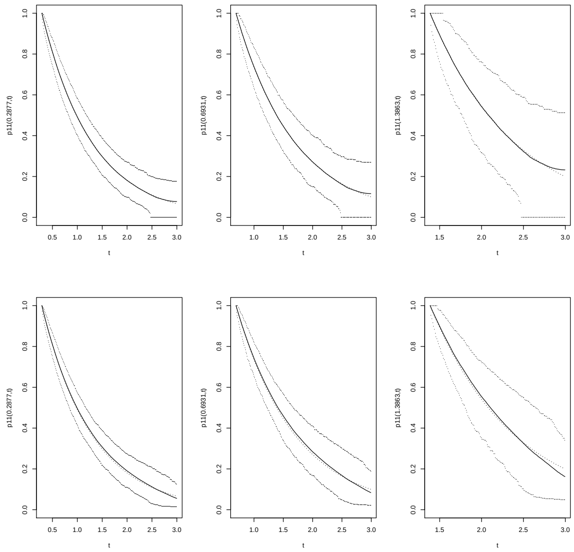

Figure 2: True $p_{11}(s, t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.2877$, $s = 0.6931$ and $s = 1.3863$. Estimates with $n = 200$ and U[0,3] censoring. Dependency scenario.

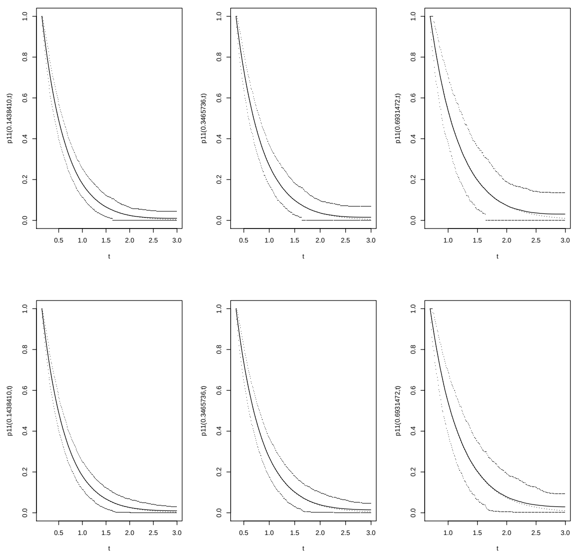

Figure 3: True $p_{11}(s, t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.1438$, $s = 0.3466$ and $s = 0.6931$. Estimates with $n = 200$ and U[0,3] censoring. Independency scenario.

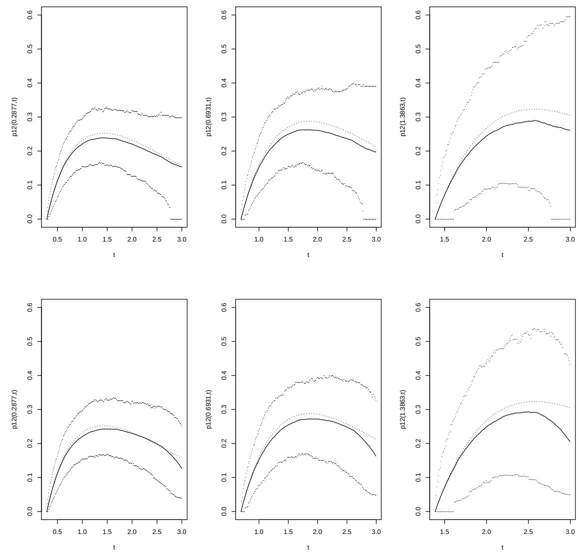

Figure 4: True $p_{12}(s,t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.2877$, $s = 0.6931$ and $s = 1.3863$. Estimates with $n = 200$ and U[0,3] censoring. Dependency scenario.

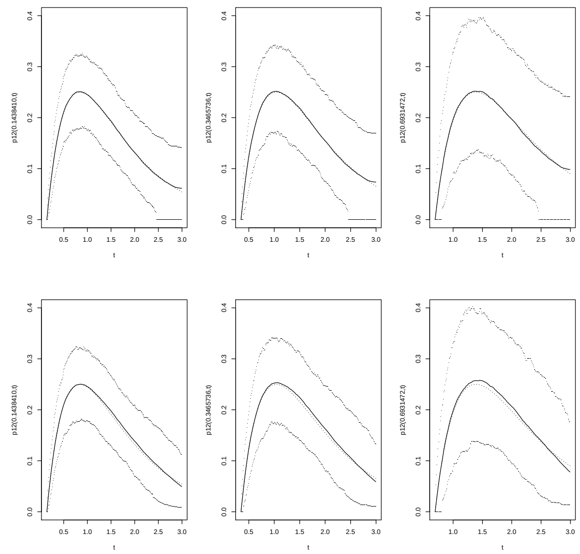

Figure 5: True $p_{12}(s,t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.1438$, $s = 0.3466$ and $s = 0.6931$. Estimates with $n = 200$ and U[0,3] censoring. Independency scenario.

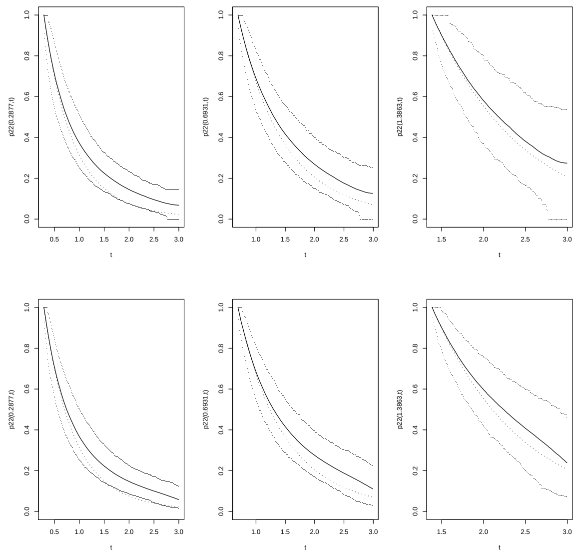

Figure 6: True $p_{22}(s,t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.2877$, $s = 0.6931$ and $s = 1.3863$. Estimates with $n = 200$ and U[0,3] censoring. Dependency scenario.

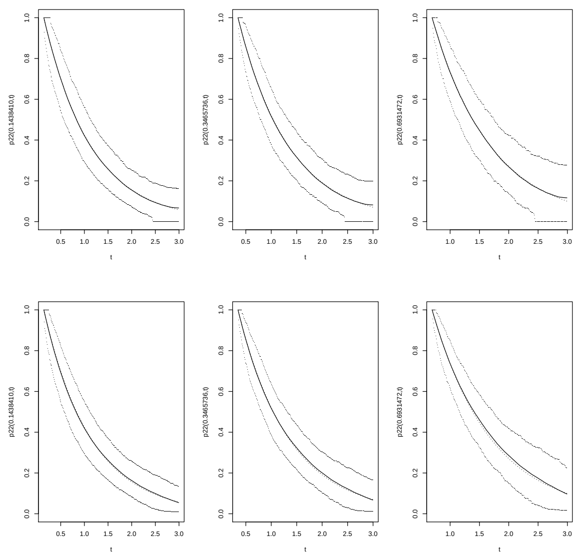

Figure 7: True $p_{22}(s,t)$ (dotted line), average estimator (solid line), and 95% oscillation limits of the AJ estimates (first row) and P-AJ (second row) for $s = 0.1438$, $s = 0.3466$ and $s = 0.6931$. Estimates with $n = 200$ and U[0,3] censoring. Independency scenario.

4 Stanford Heart Transplant data

For illustration purposes, we apply the proposed methods of Section 2 to data from the Stanford Heart Transplant Study. The data are available as part of the R survival package, and they are also reported in Crowley and Hu (1977). This study covers the period from October 1967 to April 1974. It includes 103 patients enrolled in the Stanford Heart transplant program, from which 69 received a heart transplant and among these 45 died. The total number of deaths was 75 (30 without transplantation); the remaining 28 patients contributed with censored survival times. The transplant can be considered as an associated state of risk, and we may use the so-called illness-death model with states “own heart”, “new heart” (or transplant) and “dead”. In most applications, a Markov model is often assumed for the multi-state model. A Cox model (Cox 1972) can be used to test this assumption (Hougaard 1999; Andersen et al. 2000). This is usually performed by including covariates depending on the history, such as the time of transition to the current state or the time since entry into the current state. This assumption was verified for the Stanford Transplant Study, e.g. by Hougaard (1999), which conclude that there is no effect of time since transplant on mortality, and thus that the Markov model is satisfactory. This is important, because otherwise, the consistency of the Aalen-Johansen estimator and the new estimator based on presmoothing cannot be ensured. On the other hand, if markovianity is fulfilled, the use of these methods is a wise choice. To deal with ties, a re-definition of the empiricals $M_{0n}(y)$ and $M_{1n}(y)$ is needed. Put $\widetilde{Z}_{i:n}$ for the i-th ordered Z-statistics. Similarly, put $\widetilde{T}_{i:n}$ for the i-th ordered T-statistics. For $y = \widetilde{Z}_{k:n}$ we define $\widetilde{M}_{0n}(y) = \frac{1}{n}\sum_{i=k}^{n} I(\widetilde{Z}_{i:n} \geq y)$ while for $y = \widetilde{T}_{k:n}$ we define $\widetilde{M}_{1n}(y) = \frac{1}{n}\sum_{i=k}^{n} I(\widetilde{Z}_{[i:n]} < y \leq \widetilde{T}_{i:n})$ where $\widetilde{Z}_{[i:n]}$ is the i-th concomitant (i.e. the Z-value attached to $\widetilde{T}_{i:n}$). When there are no ties, these empiricals reduce to those introduced in section 2.

Our aim with this application is to illustrate the differences between the estimated transition probabilities from Aalen-Johansen estimator (AJ) and the semiparametric estimator based on presmoothing (P-AJ). Figures 8 and 9 plot, for the two methods, the estimated transition probabilities $p_{ij}(s,t)$, $1 \leq i \leq j \leq 3$ together with pointwise confidence bands based on the bootstrap. The bootstrap estimates were obtained for $B = 1000$ replicates, by randomly sampling the n items from the original data set with replacement. The bootstrap estimates were used to obtain the 95% limits for the confidence interval of $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{22}(s,t)$. The semiparametric estimator was obtained using standard logistic regression for m_0 and m_1. The values s were chosen to be the percentile 25 and 50 of the total time ($s = 32$ and $s = 90$ days respectively). As expected, the P-AJ estimator has less variability than the AJ estimator, which has fewer jump points as s increases. For example, the extra jump points of the presmoothed AJ esti-

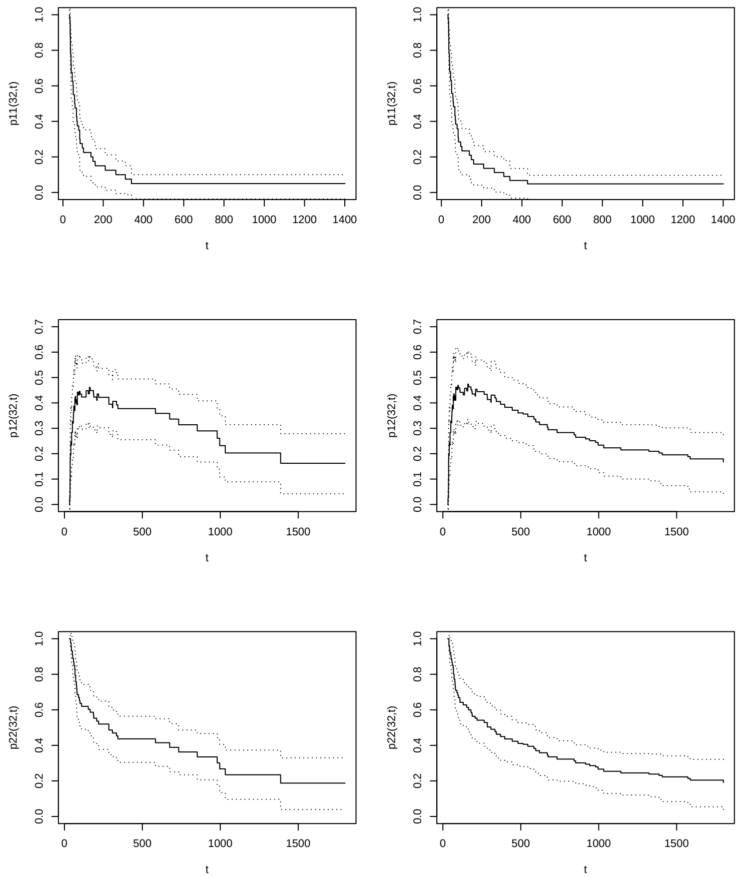

Figure 8: Estimated transition probabilities for $p_{ij}(s,t)$ with $s = 32$ based on the Aalen-Johansen estimator (on the left) and based on the presmoothed Aalen-Johansen estimator (on the right) with the corresponding 95% pointwise confidence bands. Stanford Heart Transplant data.

mator of $p_{22}(s,t)$ correspond to transplanted patients with censored values of the total time. However, both methods provide similar point estimates for all values of time. In sum, the new approach provides more reliable curves with less variability and accordingly narrower pointwise confidence bands.

5 Conclusions and final remarks

There has been several recent contributions for the estimation of the transition probabilities in the context of multi-state models. However, the Aalen-Johansen estimator is still the standard method for estimating these quan-

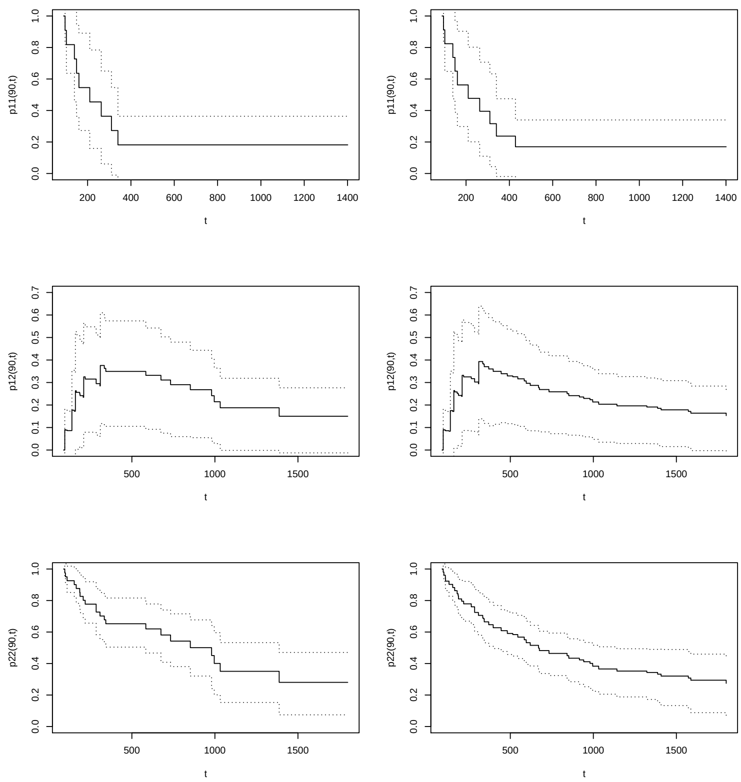

Figure 9: Estimated transition probabilities for $p_{ij}(s,t)$ with $s = 90$ based on the Aalen-Johansen estimator (on the left) and based on the presmoothed Aalen-Johansen estimator (on the right) with the corresponding 95% pointwise confidence bands. Stanford Heart Transplant data.

tities in Markov models. In this paper we propose a modification of Aalen-Johansen estimator in the illness-death model, based on a preliminary estimation (presmoothing) of the censoring probability for the total time (respectively, of the sojourn time in state 1), given the available information. An interesting open question is if this idea can be generalized (and how) to more complex multi-state models.

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. We verified through simulations that the method based on the presmoothing may be much more efficient than the original Aalen-Johansen estimators, even when there is some miss-specification in the chosen parametric family. The relative benefits of presmoothing are more clearly seen in heavily censored scenarios. We illustrated the proposed methodology using data from the

Stanford heart transplant study.

The original and the presmoothed AJ estimators are consistent in Markov models. If the Markov property is violated, then the consistency of the time-honored Aalen-Johansen estimator and of its presmoothed version can not be ensured in general. Exceptions to this are the estimator for $p_{11}(s,t)$ (for which the Markov assumption is empty) or for $p_{ij}(0,t)$ (the so-called stage occupation probabilities, see Datta and Satten 2001). Alternative estimators of the transition probabilities not relying on the Markov condition were recently proposed (Meira-Machado et al. (2006); Amorim et al. (2011)). As a drawback, these alternative methods will suffer from a larger variance in estimation, particularly when the sample size is small and there is a large censoring degree. Consequently, AJ-type estimators will be preferred when there is no strong evidence against the Markov condition.

Acknowledgement. This work was supported by the Grants MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovación and 10PXIB300068PR of the Xunta de Galicia. 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 Moreira acknowledges financial support by grant SFRH/BD/62284/2009 of the Portuguese Ministry of Science, Technology and Higher Education. This research was financed by FEDER Funds through Programa Operacional Factores de Competitividade COMPETE and by Portuguese Funds through FCT - Fundação para a Ciência e a Tecnologia, within the Project $Est-C/MAT/UI0013/2011$.

References

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non homogeneous markov and chains based on censored observations, *Scandinavian Journal of Statistics* **5**: 141–150.

Amorim, A., de Uña Álvarez, J. and Meira-Machado, L. (2011). Presmoothing the transition probabilities in the illness-death model, *Statistics & Probability Letters* **81(7)**: 797–806.

Andersen, P., Esbjerj, S. and Sorensen, T. (2000). Multistate models for bleeding episodes and mortality in liver cirrhosis, *Statistics in Medicine* **19**: 587–599.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Borgan, Ø. (1998). Aalen-johansen estimator, *Encyclopedia of Biostatistics* **1**: 5–10.

Cao, R. and Jácome, M. (2004). Presmoothed kernel density estimator for censored data, *Journal of Nonparametric Statistics* **16**: 289–309.

Cox, D. (1972). Regression models and life tables, *Journal of the Royal Statistical Society Series B* **34**: 187–220.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data, *Journal of the American Statistical Association* **72**: 27–36.

Datta, S. and Satten, G. (2001). Validity of the aalenjohansen estimators of stage occupation probabilities and nelson aalen integrated transition hazards for non-markov models, *Statistics & Probability Letters* **55**: 403–411.

de la Peña, V. and Giné, E. (1999). *Decoupling from Dependence to Independence*, Springer, New York.

de Uña-Álvarez, J. and Amorim, A. P. (2011). A semiparametric estimator of the bivariate distribution function for censored gap times, *Biometrical Journal* **53(1)**: 113–127.

de Uña Álvarez, J. and Rodríguez-Campos, C. (2004). Strong consistency of presmoothed kaplan-meier integrals when covariables are present, *Statistics* **38**: 483–496.

Devroye, L. (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. (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.

Härdle, W. and Luckhaus, S. (1984). Uniform consistency of a class of regression function estimators, *Annals of Statistics* **12**: 612–623.

Hougaard, P. (1999). Multi-state models: A review, *Lifetime Data Analysis* **5**: 239–264.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**: 457–481.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates, *Probability Theory and Related Fields* **61**: 405–415.

Meira-Machado, L., de Uña-Álvarez, J. and Cadarso-Suárez, C. (2006). Non-parametric estimation of transition probabilities in a non-markov illness-death model, *Lifetime Data Analysis* **12**: 325–344.

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

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

6 Appendix: Technical proofs

In this Section we give the proof to Theorem 1. Throughout this Section $\widehat{p}_{ij}(s,t)$ stands for the presmoothed Aalen-Johansen estimator $\widehat{p}_{ij}^{PAJ}(s,t)$. Theorem 1(a) is a consequence of Dikta 1998. Now we prove Theorem 1(b), that is, the uniform strong consistency of

$$\widehat{p}_{22}(s,t) = \prod_{s<\widetilde{T}_i\le t}\left[1 - \frac{m_{1n}(\widetilde{Z}_i,\widetilde{T}_i)I(\widetilde{Z}_i<\widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)}\right]$$

where (recall) $m_{1n}(z,t)$ is an estimator of $m_1(z,t) = P(\Delta = 1|\widetilde{Z} = z, \widetilde{T} = t, \widetilde{Z} < \widetilde{T})$ and where (recall) $\widetilde{M}_{1n}(y) = n^{-1}\sum_{i=1}^{n} I(\widetilde{Z}_i < y \le \widetilde{T}_i)$ is the empirical counterpart of $\widetilde{M}_1(y) = P(\widetilde{Z} < y \le \widetilde{T})$. Since continuity is assumed throughout, note that $\Delta_1\rho = I(\widetilde{Z} < \widetilde{T})$. The following notation will be used: $I(s,t) = \left\{i : s < \widetilde{T}_i \le t, \widetilde{Z}_i < \widetilde{T}_i\right\}$ and $I^*(s,t) = \left\{i : s < \widetilde{T}_i \le t, \widetilde{Z}_i < \widetilde{T}_i, m_{1n}(\widetilde{Z}_i,\widetilde{T}_i) > 0\right\}$. With this notation, we have

$$\widehat{p}_{22}(s,t) = \prod_{i\in I(s,t)}\left[1 - \frac{m_{1n}(\widetilde{Z}_i,\widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)}\right] = \prod_{i\in I^*(s,t)}\left[1 - \frac{m_{1n}(\widetilde{Z}_i,\widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)}\right].$$

Note that $\widehat{p}_{22}(s,t) = 0$ may happen; indeed, this is the case whenever $n\widetilde{M}_{1n}(\widetilde{T}_i) = 1$ and $m_{1n}(\widetilde{Z}_i,\widetilde{T}_i) = 1$ for some $i \in I(s,t)$. In order to avoid problems when taking logarithms, introduce the following approximation to $\widehat{p}_{22}(s,t)$:

$$\overline{p}_{22}(s,t) = \prod_{i\in I(s,t)} \frac{n\widetilde{M}_{1n}(\widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i,\widetilde{T}_i)}.$$

Since $\left|\prod_j a_j - \prod b_j\right| \le \sum_j |a_j - b_j|$ for $|a_j|, |b_j| \le 1$, we have

$$|\widehat{p}_{22}(s,t) - \overline{p}_{22}(s,t)| \le \sum_{i\in I(s,t)} \frac{m_{1n}(\widetilde{Z}_i,\widetilde{T}_i)^2}{n^2\widetilde{M}_{1n}(\widetilde{T}_i)^2}.$$

We will refer to the following Lemma, which follows from e.g. Corollary 5.2.3 in de la Peña and Giné (1999).

Lemma 1. We have w.p. 1 $\sup_y \left|\widetilde{M}_{1n}(y) - \widetilde{M}_1(y)\right| \to 0.\blacksquare$

Under condition M, from Lemma 1 we have eventually for $y \in [\tau_0, \tau_1]$ and some constant $c > 0$

$$\widetilde{M}_{1n}(y) \geq \inf_{\tau_0 \leq y \leq \tau_1} \widetilde{M}_1(y) - \sup_{\tau_0 \leq y \leq \tau_1} \left|\widetilde{M}_{1n}(y) - \widetilde{M}_1(y)\right| \geq c.$$

Hence we have w.p. 1

$$\sup_{\tau_0 \leq s < t \leq \tau_1} |\widehat{p}_{22}(s,t) - \overline{p}_{22}(s,t)| = O(n^{-1}). \tag{6}$$

Now write

$$\begin{aligned} \overline{p}_{22}(s,t) - p_{22}(s,t) &= \exp(\ln \overline{p}_{22}(s,t)) - \exp(-\Psi_n(s,t)) \\ &\quad + \exp(-\Psi_n(s,t)) - \exp(-\Psi(s,t)) \end{aligned}$$

where

$$\Psi(s,t) = \int_s^t \frac{H^1(dy)}{\widetilde{M}_1(y)}, \qquad \text{with } H^1(y) = P(\widetilde{T} \leq y, \Delta = 1, \widetilde{Z} < \widetilde{T}),$$

and

$$\Psi_n(s,t) = \sum_{i \in I(s,t)} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)}.$$

Note that $p_{22}(s,t) = \exp(-\Psi(s,t))$ because of the Markov condition, and that

$$\Psi(s,t) = E\left[\frac{I(s < \widetilde{T} \leq t)\Delta I(\widetilde{Z} < \widetilde{T})}{\widetilde{M}_1(\widetilde{T})}\right] = E\left[\frac{I(s < \widetilde{T} \leq t) m_1(\widetilde{Z}, \widetilde{T}) I(\widetilde{Z} < \widetilde{T})}{\widetilde{M}_1(\widetilde{T})}\right].$$

It will be shown that $p_{22}(s,t) = \exp(-\Psi(s,t))$ is indeed the limit of $\exp(-\Psi_n(s,t))$. This will follow from the mean-value theorem after proving the uniform strong consistency of $\Psi_n(s,t)$, which is the goal of the following Lemma.

Lemma 2. Under U_2 and M we have w.p. 1 $\sup_{\tau_0 \leq s < t \leq \tau_1} |\Psi_n(s,t) - \Psi(s,t)| \to 0$.

Proof: Write

$$\Psi_n(s,t) = \sum_{i \in I(s,t)} \frac{m_1(\widetilde{Z}_i, \widetilde{T}_i)}{n\widetilde{M}_1(\widetilde{T}_i)} + \frac{1}{n} \sum_{i \in I(s,t)} \left[\frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{\widetilde{M}_{1n}(\widetilde{T}_i)} - \frac{m_1(\widetilde{Z}_i, \widetilde{T}_i)}{\widetilde{M}_1(\widetilde{T}_i)}\right]$$

$$\equiv \Psi_n^0(s,t) + R_n(s,t).$$

By the SLLN we have $\Psi_n^0(s,t) \to \Psi(s,t)$ w.p. 1. Furthermore, under M we have w.p. 1

$$\sup_{\tau_0 \le s < t \le \tau_1} \left| \Psi_n^0(s,t) - \Psi(s,t) \right| \to 0. \tag{7}$$

To see this, note that for $s,t \in [\tau_0, \tau_1]$ we have under M

$$\Psi(s,t) \le \frac{1}{\inf_{\tau_0 \le y \le \tau_1} \widetilde{M}_1(y)} E\left[I(\tau_0 < \widetilde{T} \le \tau_1) \Delta I(\widetilde{Z} < \widetilde{T}) \right] < \infty.$$

Introduce

$$\varphi_{s,t}(u,v) = \frac{I(s < v \le t) m_1(u,v) I(u < v)}{\widetilde{M}_1(v)}.$$

Now, $\{\varphi_{s,t} : \tau_0 \le s < t \le \tau_1\}$ is a VC-subgraph class (see Proposition 5.1.12 and comments following Definition 5.1.14 in de la Peña and Giné (1999)), and φ_{τ_0,τ_1} is an integrable envelope for that class. Hence, (7) follows from Corollary 5.2.3 in de la Peña and Giné (1999).

Now,

$$\frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{\widetilde{M}_{1n}(\widetilde{T}_i)} - \frac{m_1(\widetilde{Z}_i, \widetilde{T}_i)}{\widetilde{M}_1(\widetilde{T}_i)} = \frac{1}{\widetilde{M}_{1n}(\widetilde{T}_i)} \left[m_{1n}(\widetilde{Z}_i, \widetilde{T}_i) - m_1(\widetilde{Z}_i, \widetilde{T}_i) \right]$$

$$+ \frac{m_1(\widetilde{Z}_i, \widetilde{T}_i)}{\widetilde{M}_{1n}(\widetilde{T}_i) \widetilde{M}_1(\widetilde{T}_i)} \left[\widetilde{M}_1(\widetilde{T}_i) - \widetilde{M}_{1n}(\widetilde{T}_i) \right].$$

Under U_2 and M we have

$$\sup_{\tau_0 \le s < t \le \tau_1} |R_n(s,t)| \le \left[\frac{\sup_{z<t, \tau_0 \le t \le \tau_1} |m_{1n}(z,t) - m_1(z,t)|}{c} + \frac{\sup_{\tau_0 \le y \le \tau_1} \left| \widetilde{M}_{1n}(y) - \widetilde{M}_1(y) \right|}{c'} \right] \times$$

$$\times \frac{1}{n} \sum_{i=1}^n I(\tau_0 < \widetilde{T}_i \le \tau_1) I(\widetilde{Z}_i < \widetilde{T}_i) = o(1) \text{ w.p. } 1.$$

Then the assertion of Lemma 2 follows.■

By the mean-value theorem,

$$\exp(\ln \overline{p}_{22}(s,t)) - \exp(-\Psi_n(s,t))$$

$$= (\Psi_n(s,t) + \ln \overline{p}_{22}(s,t)) \exp(-\xi_n^*(s,t))$$

for some ξ_n^* between Ψ_n and $-\ln \overline{p}_{22}$. Now:

$$\begin{aligned}\ln \overline{p}_{22}(s,t) &= \sum_{i \in I^*(s,t)} \ln \left[\frac{n\widetilde{M}_{1n}(\widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)} \right] \\ &= \sum_{i \in I^*(s,t)} \ln \left[1 - \frac{1}{x_i} \right]\end{aligned}$$

where

$$x_i = \frac{n\widetilde{M}_{1n}(\widetilde{T}_i)}{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)} + 1.$$

Note that x_i is well defined for $i \in I^*(s,t)$ and that $x_i > 1$ (because $n\widetilde{M}_{1n}(\widetilde{T}_i) \geq 1$ for $i \in I^*(s,t)$). Use

$$\ln(1 - \frac{1}{x}) = -\sum_{k=1}^{\infty} \frac{1}{kx^k}, \quad x > 1,$$

to write

$$\ln \overline{p}_{22}(s,t) = -\sum_{i \in I^*(s,t)} \sum_{k=1}^{\infty} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)^k}{k(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))^k}.$$

Hence

$$\begin{aligned}\Psi_n(s,t) + \ln \overline{p}_{22}(s,t) &= \sum_{i \in I^*(s,t)} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)} \\ &- \sum_{i \in I^*(s,t)} \sum_{k=1}^{\infty} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)^k}{k(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))^k} \\ &= \sum_{i \in I^*(s,t)} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)}{n\widetilde{M}_{1n}(\widetilde{T}_i)(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))} \\ &- \sum_{i \in I^*(s,t)} \sum_{k=2}^{\infty} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)^k}{k(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))^k} \equiv I + II.\end{aligned}$$

Under M we have, uniformly in $\tau_0 \leq s < t \leq \tau_1$, $I = O(n^{-1})$ w.p. 1. Besides, by noting

$$\sum_{k=2}^{\infty} x^k = \frac{1}{1-x} - 1 - x = \frac{x^2}{1-x}, \quad x < 1,$$

we have that the absolute value of II is bounded by (take $x = m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)/(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))$)

$$\sum_{i \in I^*(s,t)} \sum_{k=2}^{\infty} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)^k}{(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))^k}$$

$$= \sum_{i \in I^*(s,t)}^{n} \frac{m_{1n}(\widetilde{Z}_i, \widetilde{T}_i)^2}{n\widetilde{M}_{1n}(\widetilde{T}_i)(n\widetilde{M}_{1n}(\widetilde{T}_i) + m_{1n}(\widetilde{Z}_i, \widetilde{T}_i))} = O(n^{-1})$$

w.p. 1. uniformly in $\tau_0 \leq s < t \leq \tau_1$. This shows that

$$\sup_{\tau_0 \leq s < t \leq \tau_1} |\Psi_n(s,t) + \ln \overline{p}_{22}(s,t)| = O(n^{-1}) \quad \text{w.p. } 1$$

and consequently

$$\sup_{\tau_0 \leq s < t \leq \tau_1} |\exp(\ln \overline{p}_{22}(s,t)) - \exp(-\Psi_n(s,t))| = O(n^{-1}) \quad \text{w.p. } 1. \qquad (8)$$

Now, use the mean-value theorem to write

$$\exp(-\Psi(s,t)) - \exp(-\Psi_n(s,t)) = [\Psi_n(s,t) - \Psi(s,t)] \exp(-\xi_n(s,t))$$

from which

$$\sup_{\tau_0 \leq s < t \leq \tau_1} |\exp(-\Psi(s,t)) - \exp(-\Psi_n(s,t))| \leq \sup_{\tau_0 \leq s < t \leq \tau_1} |\Psi_n(s,t) - \Psi(s,t)|.$$

Then Theorem 1(b) follows from Lemma 2, (8), (6), and the decomposition

$$\begin{aligned} \widehat{p}_{22}(s,t) - p_{22}(s,t) &= \widehat{p}_{22}(s,t) - \overline{p}_{22}(s,t) \\ &\quad + \exp(\ln \overline{p}_{22}(s,t)) - \exp(-\Psi_n(s,t)) \\ &\quad + \exp(-\Psi_n(s,t)) - \exp(-\Psi(s,t)). \end{aligned}$$

In order to prove Theorem 1(c) write, with $J(s,t) = \left\{ i : s < \widetilde{Z}_i \leq t, \widetilde{Z}_i < \widetilde{T}_i \right\}$,

$$\widehat{p}_{12}(s,t) = \frac{1}{n} \sum_{i \in J(s,t)} \frac{\widehat{p}_{11}(s, \widetilde{Z}_i^-)\widehat{p}_{22}(\widetilde{Z}_i, t)}{\widetilde{M}_{0n}(\widetilde{Z}_i)}$$

$$\begin{aligned} &= \frac{1}{n} \sum_{i \in J(s,t)} \left[\widehat{p}_{11}(s, \widetilde{Z}_i^-) - p_{11}(s, \widetilde{Z}_i) \right] \frac{\widehat{p}_{22}(\widetilde{Z}_i, t)}{\widetilde{M}_{0n}(\widetilde{Z}_i)} \\ &\quad + \frac{1}{n} \sum_{i \in J(s,t)} \left[\widehat{p}_{22}(\widetilde{Z}_i, t) - p_{22}(\widetilde{Z}_i, t) \right] \frac{p_{11}(s, \widetilde{Z}_i)}{\widetilde{M}_{0n}(\widetilde{Z}_i)} \\ &\quad + \frac{1}{n} \sum_{i \in J(s,t)} p_{11}(s, \widetilde{Z}_i) p_{22}(\widetilde{Z}_i, t) \left[\frac{1}{\widetilde{M}_{0n}(\widetilde{Z}_i)} - \frac{1}{\widetilde{M}_0(\widetilde{Z}_i)} \right] \\ &\quad + \frac{1}{n} \sum_{i \in J(s,t)} \frac{p_{11}(s, \widetilde{Z}_i) p_{22}(\widetilde{Z}_i, t)}{\widetilde{M}_0(\widetilde{Z}_i)} \\ &\equiv I(s,t) + II(s,t) + III(s,t) + IV(s,t) \end{aligned}$$

where $\widetilde{M_0}(y) = P(\widetilde{Z} \geq y)$. Since, because of the Markov condition,

$$E\left[\frac{p_{11}(s, \widetilde{Z}_i)p_{22}(\widetilde{Z}_i, t)}{\widetilde{M_0}(\widetilde{Z}_i)} I(s < \widetilde{Z}_i \leq t, \widetilde{Z}_i < \widetilde{T}_i)\right] = p_{12}(s, t),$$

the SLLN gives $IV(s, t) \to p_{12}(s, t)$ w.p. 1. Furthermore, by using Proposition 5.1.12 in de la Peña and Giné (1999) as in Lemma 2 above we get w.p. 1

$$\sup_{0 \leq s < t \leq \tau} |IV(s, t) - p_{12}(s, t)| \to 0.$$

It remains to show that $I(s, t)$, $II(s, t)$, and $III(s, t)$ go to zero w.p. 1 uniformly on $[0, \tau]$. But this is easily seen by using Theorem 1(a),(b), Glivenko-Cantelli, and the fact that $\widetilde{M_0}$ is bounded away from zero on $[0, \tau]$.■