Biometrics 000, 000–000 DOI: 000
000 0000

Nonparametric Estimation of Transition Probabilities in the non-Markov Illness-Death Model: a Comparative Study

Jacobo de Uña-Álvarez

Department of Statistics and Operations Research,

Facultad de Ciencias Económicas y Empresariales & Centro de Investigaciones Biomédicas (CINBIO),

University of Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain

email: jacobo@uvigo.es

and

Luís Meira-Machado

Centre of Mathematics & Department of Mathematics and Applications, University of Minho, Campus de Azurém

4800-058 Guimarães,Portugal

email: lmachado@math.uminho.pt

Summary: Multi-state models are often used for modeling complex event history data. In these models the estimation of the transition probabilities is of particular interest, since they allow for long-term predictions of the process. These quantities have been traditionally estimated by the Aalen-Johansen estimator, which is consistent if the process is Markov. Several non-Markov estimators have been proposed in the recent literature, and their superiority with respect to the Aalen-Johansen estimator has been proved in situations in which the Markov condition is strongly violated. However, the existing estimators have the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution, which is not often the case. In this paper we propose two new methods for estimating the transition probabilities in the progressive illness-death model. Some asymptotic results are derived. The proposed estimators are consistent regardless the Markov condition and the referred assumption about the censoring support. We explore the finite sample behavior of the estimators through simulations. The main conclusion of this piece of research is that the proposed estimators are much more efficient than the existing non-Markov estimators in most cases. An application to a clinical trial on colon cancer is included. Extensions to progressive processes beyond the three-state illness-death model are discussed.

Key words: Aalen-Johansen estimator; Kaplan-Meier; Markov condition; Multi-state model; Survival Analysis.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Multi-state models are models for stochastic processes which represent the evolution of an item (a patient, a device, and so on) along time. The item may visit a number of states along its progress and, hence, the multi-state model indicates the existing states as long as the allowed transitions among them. Multi-state models have been widely used in biomedicine to investigate the progression of patients undergoing a given illness or surgery (Commenges, 1999; Hougaard, 1999, 2000; Andersen et al., 2000; Meira-Machado et al., 2009); usually, the states represent the occurrence of an event which may be related to survival prognosis, such as complications after surgery, recurrences, or non-fatal episodes.

In this setting, the estimation of the so-called transition hazards and transition probabilities is of particular interest. From the transition hazard, it is possible to evaluate the probability of one event ahead. On the other hand, in order to consider several transitions ahead, the transition probabilities must be evaluated and, hence, they are essential for long-term survival prognosis (Hougaard, 2000). The occupation probabilities, often used in practice, are particular examples of transition probabilities in which the time origin is taken as the current time.

The standard nonparametric method to estimate a transition probability matrix is the time-honored Aalen-Johansen estimator (Aalen and Johansen, 1978). This estimator is adapted to censoring, and it benefits from the assumption of Markovianity on the underlying stochastic process. This assumption states that the relevant information for the future evolution of the process is provided by its current state, independently of the states previously visited and the transition times among them. The Markov assumption may be violated in practice; for example, the arrival time to the current state of the process often influences the transition hazard, leading to non-Markov structures (e.g. Andersen et al., 2000). See the application to the colon cancer study in Section 4 for further illustration. Although the Aalen-Johansen

estimator may be used to consistently estimate occupation probabilities for non-Markov multi-state models (Datta and Satten, 2001, 2002; Glidden, 2002), in general it provides biased estimators if the Markov assumption does not hold (Meira-Machado et al., 2006; Allignol et al., 2014).

Meira-Machado et al. (2006) introduced for the first time fully non-Markov estimators for the transition probabilities in the progressive illness-death model (see Section 2), which is a special multi-state model with plenty of applications in biomedicine and other areas. These authors showed the practical superiority of their estimators relative to the Aalen-Johansen in situations in which the Markov condition is strongly violated. Meira-Machado et al. (2006)'s approach was recently revisited by Allignol et al. (2014), who proposed a closely related non-Markov estimator too. Both Meira-Machado et al. (2006) and Allignol et al. (2014) proposals have the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution, otherwise they only report valid estimators for truncated transition probabilities. Unfortunately, this assumption is often not fulfilled in medical applications due to limitations in the patients' following-up. A practical example of this situation is considered in Section 4, where patients with colon cancer have a maximum follow-up time of about 9 years. In this paper we overcome this issue by introducing alternative estimators which are consistent regardless the Markov condition and the referred assumption on the censoring support. Nonparametric estimation of occupancy probabilities and, more generally, of transition probabilities in the progressive illness-death model and in more complicated multi-state models has been considered in a number of papers, including (besides the aforementioned ones) Pepe et al. (1991), Pepe (1991), Andersen et al. (1991), Borgan (2002), Frydman and Szarek (2009), Lan and Datta (2010), Amorim et al. (2011), Spitoni et al. (2012), Moreira et al. (2013), Mostajabi and Datta (2013), Frydman et al. (2013), and Meira-Machado et al. (2013).

This paper is organized as follows. In Section 2 we revisit Meira-Machado et al. (2006)'s approach and we introduce a simple modification of their estimators which makes the censoring support condition unnecessary. Some new alternative estimators based on subsamples are also introduced. The asymptotic properties of the proposed estimators are discussed. The performance of three sets of estimators is investigated through simulations in Section 3, while in Section 4 the methods are compared through the re-analysis of medical data from a clinical trial on colon cancer. Extensions to progressive processes beyond the three-state illness-death model are discussed in Section 5. The main conclusions of our research are deferred to Section 6. The technical results on the consistency and asymptotic normality of the introduced estimators are given in the online supplementary material.

2. Nonparametric Estimators

The non-Markov illness-death model is characterized by the joint distribution of (Z^0, T^0), where Z^0 is the sojourn time in the initial state 1 and T^0 is the total survival time (i.e. the absorption time). Note that (Z^0, T^0) falls on a line with a strictly positive probability, since $T^0 = Z^0$ for those individuals undergoing a direct transition from state 1 to the absorbing state 3. On the other hand, $Z^0 < T^0$ indicates that the individual visits the intermediate state 2 at some time. Under censoring, only the censored versions of Z^0 and T^0, along with their corresponding censoring indicators, are available. Define $Z = \min(Z^0, C)$, $T = \min(T^0, C)$, $\rho = I(Z^0 \leqslant C)$, and $\delta = I(T^0 \leqslant C)$, where C is the potential censoring time, which is assumed to be independent of (Z^0, T^0). Finally, the data are $(Z_i, T_i, \rho_i, \delta_i)$, $1 \leqslant i \leqslant n$, iid copies of (Z, T, ρ, δ).

The target is each of the five transition probabilities $p_{ij}(s,t)$, where $1 \leqslant i \leqslant j \leqslant 3$ and $s < t$ are two pre-specified time points. As noted by Meira-Machado et al. (2006), these transition probabilities are functions involving expectations of particular transformations of

the pair (Z^0, T^0); namely,

$$\begin{aligned} p_{11}(s,t) &= \frac{E\left[I(Z^0 > t)\right]}{E\left[I(Z^0 > s)\right]}, \qquad p_{12}(s,t) = \frac{E\left[I(s < Z^0 \leqslant t < T^0)\right]}{E\left[I(Z^0 > s)\right]}, \\ p_{13}(s,t) &= \frac{E\left[I(s < Z^0, T^0 \leqslant t)\right]}{E\left[I(Z^0 > s)\right]}; \end{aligned}$$

and

$$p_{22}(s,t) = \frac{E\left[I(Z^0 \leqslant s, t < T^0)\right]}{E\left[I(Z^0 \leqslant s < T^0)\right]}, \qquad p_{23}(s,t) = \frac{E\left[I(Z^0 \leqslant s < T^0 \leqslant t)\right]}{E\left[I(Z^0 \leqslant s < T^0)\right]}.$$

The transition probability $p_{11}(s,t)$ depends just on the marginal distribution of the sojourn time in state 1 and, therefore, it may be efficiently estimated from the Kaplan-Meier estimator of the survival function $S_{Z^0}(z) = P(Z^0 > z)$, given by $\widehat{S}_{Z^0}(z) = 1 - \sum_{i=1}^{n} w_i I(Z_{(i)} \leqslant z) \equiv 1 - \widehat{F}_{Z^0}(z)$, where $Z_{(1)} \leqslant ... \leqslant Z_{(n)}$ denotes the ordered Z-sample and

$$w_i = \frac{\rho_{[i]}}{n-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\rho_{[j]}}{n-j+1}\right]$$

is the Kaplan-Meier weight attached to $Z_{(i)}$ (i.e. the jump of $\widehat{F}_{Z^0}(z)$ at $z = Z_{(i)}$). In the expression of w_i, $\rho_{[i]}$ denotes the i-th concomitant value of the censoring indicator (that is, $\rho_{[i]} = \rho_j$ if $Z_{(i)} = Z_j$). Explicitly, $\widehat{p}_{11}(s,t) = \widehat{S}_{Z^0}(t)/\widehat{S}_{Z^0}(s)$ is a consistent estimator of $p_{11}(s,t)$ for $t \leqslant \tau_Z \equiv \inf\{z : F_Z(z) = 1\}$, where F_Z stands for the cumulative distribution function (cdf) of Z. The inconsistency of $t \mapsto \widehat{p}_{11}(s,t)$ beyond τ_Z is a consequence of the complete absence of information on Z^0 on the interval (τ_Z, ∞). Note that this problem vanishes when the support of C contains the support of Z^0. The estimator $\widehat{p}_{11}(s,t)$ is just the Aalen-Johansen estimator as introduced by Aalen and Johansen (1978) for a general Markov multi-state model; note that the Markov assumption is irrelevant for state 1 (since there is only one possible past for the individuals in that state) and, consequently, non-Markov and Markov methods coincide in this case.

The estimation of the other four transition probabilities is more involving, since they depend on both Z^0 and T^0. Meira-Machado et al. (2006) considered the following estimators:

$$\widehat{p}_{12}(s,t) = \frac{\sum_{i=1}^{n} W_i I(s < Z_{[i]} \leqslant t < T_{(i)})}{\widehat{S}_{Z^0}(s)}, \qquad \widehat{p}_{13}(s,t) = \frac{\sum_{i=1}^{n} W_i I(s < Z_{[i]}, T_{(i)} \leqslant t)}{\widehat{S}_{Z^0}(s)};$$

and

$$\widehat{p}_{22}(s,t) = \frac{\sum_{i=1}^{n} W_i I(Z_{[i]} \leqslant s, t < T_{(i)})}{\sum_{i=1}^{n} W_i I(Z_{[i]} \leqslant s < T_{(i)})}, \qquad \widehat{p}_{23}(s,t) = \frac{\sum_{i=1}^{n} W_i I(Z_{[i]} \leqslant s < T_{(i)} \leqslant t)}{\sum_{i=1}^{n} W_i I(Z_{[i]} \leqslant s < T_{(i)})}.$$

Here, W_i denotes the Kaplan-Meier weight attached to the i-th ordered T-datum, $T_{(i)}$; and $\left(Z_{[i]}, \delta_{[i]}\right)$ stands for its concomitant. Explicitly, we have

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

Note that the Kaplan-Meier estimator of $S_{T^0}(t) = P(T^0 > t)$ is given by $\widehat{S}_{T^0}(t) = 1 - \sum_{i=1}^{n} W_i I(T_{(i)} \leqslant t) \equiv 1 - \widehat{F}_{T^0}(t)$. The idea of weighting (Z_i, T_i) by the marginal weight of T_i^0 works because C is assumed to be independent of the process. The assumption of independent censoring implies that the identifiability conditions in Stute (1993) for multivariate Kaplan-Meier integrals hold, where the sojourn time in state 1 plays the role of a covariate. However, potential inconsistencies at the right tail of T^0 may appear. In order to verify this note that, in general (cfr. Theorem 1 in Meira-Machado et al., 2006)

$$\sum_{i=1}^{n} W_i \varphi(Z_i, T_i) \to E\left[\varphi(Z^0, T^0) I(T^0 \leqslant \tau_T)\right]$$

with probability 1, where φ stands for any transformation (provided that the limit exists) and τ_T is the upper bound of the support of T. Therefore, the estimator $\widehat{p}_{13}(s,t)$ is consistent only for $t \leqslant \tau_T$, while there is no hope that $\widehat{p}_{12}(s,t)$, $\widehat{p}_{22}(s,t)$ and $\widehat{p}_{23}(s,t)$ will converge to their respective targets for any t when the support of C is strictly contained in the support of T^0. In practice, this means that these transition probabilities may only be intepreted for the subpopulation of individuals with absorption time smaller than τ_T, which is hardly of interest. A similar problem occurs with a closely related estimator recently proposed by Allignol et al. (2014).

To avoid this issue, alternative estimators may be introduced. First, since $E\left[I(Z^0 \leqslant s < T^0)\right] = S_{T^0}(s) - S_{Z^0}(s)$, the denominator in the definition of $\widehat{p}_{22}(s,t)$ and $\widehat{p}_{23}(s,t)$ can be replaced by $\widehat{S}_{T^0}(s) - \widehat{S}_{Z^0}(s)$, which is consistent for $s \leqslant \tau_Z$. Second, $\widehat{p}_{12}(s,t)$, $\widehat{p}_{22}(s,t)$ and $\widehat{p}_{23}(s,t)$ are

redefined as

$$\widehat{p}^{*}_{12}(s,t) = 1 - \widehat{p}_{11}(s,t) - \widehat{p}_{13}(s,t) = \frac{\sum_{i=1}^{n} w_i I(s < Z_{(i)} \leqslant t) - \sum_{i=1}^{n} W_i I(s < Z_{[i]}, T_{(i)} \leqslant t)}{\widehat{S}_{Z^0}(s)},$$

$$\widehat{p}^{*}_{22}(s,t) = 1 - \widehat{p}^{*}_{23}(s,t), \qquad \text{and } \widehat{p}^{*}_{23}(s,t) = \frac{\sum_{i=1}^{n} W_i I(Z_{[i]} \leqslant s < T_{(i)} \leqslant t)}{\widehat{S}_{T^0}(s) - \widehat{S}_{Z^0}(s)},$$

respectively. In this manner, $\widehat{p}^{*}_{12}(s,t)$, $\widehat{p}^{*}_{22}(s,t)$ and $\widehat{p}^{*}_{23}(s,t)$ are consistent when $t \leqslant \tau_Z$. Moreover, $\widehat{p}^{*}_{22}(s,t)$ and $\widehat{p}^{*}_{23}(s,t)$ are consistent for $s \leqslant \tau_Z$ and $t \leqslant \tau_T$. These alternative estimators can be found in e.g. de Uña-Álvarez (2010). On the other hand, by using the iid representation for multivariate Kaplan-Meier integrals (Stute, 1996) and the delta method together with the multivariate CLT, the asymptotic normality of the estimators $\widehat{p}^{*}_{ij}(s,t)$ is immediately obtained. See the online supplementary material for details.

Another way to introduce nonparametric estimators for the transition probabilities is by considering specific subsamples or portions of the data at hand. For example, given the time point s, to estimate $p_{1j}(s,t)$ for $j = 1,2,3$ the analysis can be retricted to the individuals observed in state 1 at time s. This set is just $\mathcal{S}_1 = \{i : Z_i > s\}$. Note that an individual in $\mathcal{S}_1$ reports information not only about his/her sojourn time in state 1 ($Z^0 > s$) but also about the potential censoring time ($C > s$); however, as long as C is independent of Z^0, a subject in $\mathcal{S}_1$ is representative of those individuals for which Z^0 exceeds s. On the other hand, for the subpopulation $Z > s$, the censoring time C is still independent of the pair (Z^0, T^0) and, therefore, Kaplan-Meier-based estimation will be consistent. The same applies to the analysis restricted to the individuals observed in state 2 at time s, say $\mathcal{S}_2 = \{i : Z_i \leqslant s < T_i\}$, which serves to introduce new estimators for $p_{2j}(s,t)$, $j = 2,3$.

To formalize things, let n_{1s} be the cardinal of $\mathcal{S}_1$, and let $\widehat{S}^{(s)}_{Z^0}(z) = 1 - \sum_{i=1}^{n_{1s}} w_i^{(s)} I(Z^{(s)}_{(i)} \leqslant z)$ and $\widehat{S}^{(s)}_{T^0}(t) = 1 - \sum_{i=1}^{n_{1s}} W_i^{(s)} I(T^{(s)}_{(i)} \leqslant t)$ be the Kaplan-Meier survival functions of Z^0 and T^0 computed from such a subsample. That is, $w_i^{(s)}$ and $W_i^{(s)}$ are defined through

$$w_i^{(s)} = \frac{\rho^{(s)}_{[i]}}{n_{1s} - i + 1} \prod_{j=1}^{i-1} \left[1 - \frac{\rho^{(s)}_{[j]}}{n_{1s} - j + 1} \right], \qquad 1 \leqslant i \leqslant n_{1s};$$

and

$$W_i^{(s)} = \frac{\delta_{[i]}^{(s)}}{n_{1s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j]}^{(s)}}{n_{1s}-j+1} \right], \qquad 1 \leqslant i \leqslant n_{1s},$$

where $\left(Z_{(i)}^{(s)}, \rho_{[i]}^{(s)}\right)$, $i = 1, ..., n_{1s}$, is the (Z, ρ)-sample in $\mathcal{S}_1$ ordered with respect to Z, and $\left(T_{(i)}^{(s)}, \delta_{[i]}^{(s)}\right)$, $i = 1, ..., n_{1s}$, is the (T, δ)-sample in $\mathcal{S}_1$ ordered with respect to T. Introduce $\widehat{p}_{11}^{new}(s,t) = \widehat{S}_{Z^0}^{(s)}(t)$, $\widehat{p}_{12}^{new}(s,t) = 1 - \widehat{p}_{11}^{new}(s,t) - \widehat{p}_{13}^{new}(s,t)$, and $\widehat{p}_{13}^{new}(s,t) = 1 - \widehat{S}_{T^0}^{(s)}(t)$. Note that $\widehat{p}_{12}^{new}(s,t) = \widehat{S}_{T^0}^{(s)}(t) - \widehat{S}_{Z^0}^{(s)}(t)$; in a certain sense, $\widehat{p}_{12}^{new}(s,t)$ is a Pepe-type estimator (Pepe, 1991) when conditioning on the individuals observed in state 1 at time s. On the other hand, $\widehat{p}_{11}^{new}(s,t)$ is just the Aalen-Johansen estimator $\widehat{p}_{11}(s,t)$ referred above, as demonstrated in the online supplementary material.

Similarly, let n_{2s} be the cardinal of $\mathcal{S}_2$, and let $\widehat{S}_{T^0}^{[s]}(t) = 1 - \sum_{i=1}^{n_{2s}} W_i^{[s]} I(T_{(i)}^{[s]} \leqslant t)$ be the Kaplan-Meier survival function of T^0 computed from such a subsample. That is,

$$W_i^{[s]} = \frac{\delta_{[i]}^{[s]}}{n_{2s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j]}^{[s]}}{n_{2s}-j+1} \right], \qquad 1 \leqslant i \leqslant n_{2s},$$

where $\left(T_{(i)}^{[s]}, \delta_{[i]}^{[s]}\right)$, $i = 1, ..., n_{2s}$, is the (T, δ)-sample in $\mathcal{S}_2$ ordered with respect to T. Introduce $\widehat{p}_{22}^{new}(s,t) = \widehat{S}_{T^0}^{[s]}(t)$ and $\widehat{p}_{23}^{new}(s,t) = 1 - \widehat{S}_{T^0}^{[s]}(t)$. The estimators $\widehat{p}_{ij}^{new}(s,t)$ have the simple form of a Kaplan-Meier estimator, based on a certain subsample which is determined by the time point s. An asymptotic analysis of these new estimators is reported in the online supplementary material; in particular, their asymptotic normality is established, and expressions for the limit variances are provided.

In practice, estimation of the variance is needed for inference purposes. To this end, plug-in methods or resampling techniques such as the jackknife or the bootstrap can be used. Among these, resampling methods are often preferred due to its simplicity and good relative behavior. In Section 4 we use the bootstrap to evaluate the variance of the estimated transition probabilities and to construct confidence limits based on the percentile bootstrap and on the normal approximation. Like in previous papers in the area (Moreira et al., 2013;

Allignol et al., 2014), the bootstrap provides here a practical solution to the problem of variance estimation and inference.

3. Simulation Study

In this Section we investigate the performance of the proposed estimators through simulations. More specifically, the estimators for the transition probabilities $p_{12}(s,t)$ and $p_{23}(s,t)$ introduced in Section 2 are considered.

To simulate the data in the illness-death model, we separately consider the subjects passing through the intermediate state 2 at some time (that is, those cases with $\gamma = I(Z^0 < T^0) = 1$), and those who directly go to the absorbing state 3 ($\gamma = 0$). For the first subgroup of individuals ($\gamma = 1$), the successive gap times $(Z^0, T^0 - Z^0)$ 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 θ controls for the amount of dependence between the gap times. The parameter θ is set to 0 (independent gap times), and also to 1 (0.25 correlation between Z^0 and $T^0 - Z^0$). For the second subgroup of individuals ($\gamma = 0$), the value of Z^0 is simulated according to an exponential with rate parameter 1. To control the amount of individuals undergoing each transition ($1 \to 2 \to 3$ and $1 \to 3$) we simulate γ independently of Z^0, according to a Bernoulli distribution with parameter $p = 0.7$. The simulation procedure is as follows:

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

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

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

(4) $Z^0 = \log\left(1/\left(1-U_1\right)\right)$;

(5) $\gamma \sim Bernoulli(p)$, with $p = 0.7$;

(6) $T^0 = Z^0 + \gamma \log\left(1/\left(1-U_2\right)\right)$.

This algorithm (scenario 1 in our simulations) can be seen as arising from a latent transition times approach, in which first the potential transition times $1 \to 2$ (T_{12} say) and $1 \to 3$ (T_{13}) are generated according to independent exponentials with rate 0.7 and 0.3 respectively, and then $Z^0 = min(T_{12}, T_{13})$ and $\gamma = I(T_{12} \leqslant T_{13})$ are recorded.

In order to consider a situation with nonhomogeneous transition intensities (scenario 2), we introduce a frailty term F in the algorithm above, acting multiplicatively on the transition hazards. Specifically, F is generated according to an exponential model with rate 1, and steps (4) and (6) above are replaced by $Z^0 = \log(1/(1-U_1))/F$ and $T^0 = Z^0 + \gamma \log(1/(1-U_2))/F$ respectively. This relates the simulation of a mixed-Markov model as considered in Satten (1999); indeed, in the independent case ($\theta = 0$), the simulated model is Markov conditionally on the frailty F. Unconditionally, this algorithm simulates loglogistic times Z^0 and $T^0 - Z^0$, leading to a non-constant transition hazards model.

An independent uniform censoring time C is generated, according to models $Uniform(0, 4)$ and $Uniform(0, 3)$ in scenario 1, and models $Uniform(0, 8)$ and $Uniform(0, 5)$ in scenario 2, to introduce lighter or heavier censoring. To be more specific, in the case $\theta = 1$ the censoring proportion in the first gap time Z^0 is 24%-32% (scenario 1) or 27%-35% (scenario 2), while for the second gap time these percentages increase to 47%-57% (scenario 1) or 35%-48% (scenario 2). Similar censoring proportions hold for $\theta = 0$.

For each simulated setting we derive the analytic expression of $p_{12}(s,t)$ and $p_{23}(s,t)$ for several (s,t) pairs, corresponding to combinations of the percentiles 20%, 40%, 60% and 80% of the marginal distributions of the gap times. Sample sizes $n = 50$, $n = 100$ and $n = 250$ are considered. In each simulation, 1000 samples are generated. For each fixed pair (s,t) and

each of the three estimators we obtain the mean bias, the standard deviation (SD), and the mean square error (MSE) based on the 1000 Monte Carlo replicates. Tables 1 (scenario 1) and 2 (scenario 2) report the results for the case with dependent gap times ($\theta = 1$); the results for the independent case (not shown) are similar.

From Tables 1 and 2 it can be seen that the performance of the methods is poorer at the right tail (i.e. larger values of s and t) where the censoring effects are stronger. As expected, an increasing sample size n yields to smaller SD. The SD increases with the censoring percentage, which was also expected.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

Results in Tables 1 and 2 reveal that the new estimators based on subsamples ($\widehat{p}_{12}^{new}(s,t)$ and $\widehat{p}_{23}^{new}(s,t)$) may be much more efficient than other available estimators. This is particularly clear when comparing these estimators to the original Meira-Machado et al. (2006) estimators $\widehat{p}_{12}(s,t)$ and $\widehat{p}_{23}(s,t)$, hereafter referred to as LIDA estimators; the new method achieves much better results with less bias and less variability. The two methods proposed in this paper ($\widehat{p}_{ij}^{*}(s,t)$ and $\widehat{p}_{ij}^{new}(s,t)$) obtain in all settings a negligible bias (decreasing as n increases), while the LIDA estimators show a systematic bias which does not decrease with an increasing sample size. Indeed, for a sample size of $n = 250$ some cases in which the bias of $\widehat{p}_{ij}(s,t)$ has the order of magnitude of the SD may be found. On the other hand, the variance dominates the performance of the two estimators proposed in this paper.

The relative efficiency between the new estimator based on subsamples ($\widehat{p}_{ij}^{new}(s,t)$) and the modified LIDA estimator ($\widehat{p}_{ij}^{*}(s,t)$) can be measured by the ratio between their corresponding MSEs. Tables 1 and 2 show that $\widehat{p}_{ij}^{*}(s,t)$ is very competitive; this is particularly true for the transition probability $p_{12}(s,t)$. Even so, the new estimators reported a smaller MSE with

only a few exceptions, corresponding to small values of s and t. The relative advantages of the new estimator are apparently greater when estimating the transition probability $p_{23}(s,t)$. For example, for $(s,t) = (0.2231, 1.6094)$ and $n = 250$, the relative efficiency of $\widehat{p}^{*}_{ij}(s,t)$ was as low as 62% in scenario 1 (53% in scenario 2). It can be verified that the MSE ratio decreases as the target $p_{ij}(s,t)$ gets larger, when the estimator based on subsamples provides a clear improvement over the modified LIDA estimator. There is some (relatively minor) influence of the sample size on the MSE ratio too, which tends to decrease as n increases. Results for $n = 5000$ (not shown) indicate that the asymptotic behavior of both methods is different, and the limit variance of the one based on subsamples is strictly smaller in most situations.

We have also investigated the influence of the proportion of individuals visiting the intermediate state (p) on the results, by repeating the simulations with $p = 0.5$ in step (5) of the algorithm. The results (not shown) were roughly the same, with a somehow poorer performance of each method when estimating the transition probability $p_{23}(s,t)$, according to the relative smaller sample size in the intermediate state. We have investigated the impact of possible dependencies between γ and Z^0 on the estimators accuracy too. To this end, the parameter p in step (5) was allowed to depend on the Z^0 generated in step (4) through the equation $p(Z^0) = exp(aZ^0)$, for the cases $a = -1$ (50% of visitors for state 2) and $a = -3/7$ (70% of visitors). The results for $a = -3/7$ are provided in Table 3, and they are similar to those in Table 1, with only some slight improvements of all of them in the estimation of $p_{23}(s,t)$. The results corresponding to the case $a = -1$ (not shown) were close to those of the independent setting with $p = 0.5$.

4. Clinical Trial on Duke's Stage III Patients

In this Section we re-analyze data of 929 patients 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, 468 patients developed recurrence and among these 414

died, while 38 patients died without recurrence. The rest of the patients (423) remained alive and disease-free up to the end of the follow-up. We use an illness-death multi-state model, with "recurrence" (local recurrence or distant metastasis) as a transient state and "death" as absorbing state. We have studied the influence of the intermediate event time on the mortality transition of recurrent patients using a Cox proportional hazards model. This model reported a significant coefficient (p-value < 0.001) of negative sign for the recurrence time, according to an increased risk of death shortly after relapse. This suggests that the Markov assumption may be unsatisfactory and, therefore, non-Markov estimators may be a sensible choice to analyze these data. In this section we present plots for the three different methods to estimate the transition probabilities described in Section 2. As mentioned, none of these methods require the process to be Markov.

Figure 1 reports estimated transition probabilities for $p_{12}(s,t)$, for fixed values $s = 365$ and $s = 1095$ (days), along time t. Hence, this Figure 1 allows for an inspection along time of the probability of being alive with recurrence for the individuals who are disease free one year (Figure 1, top) and three years (Figure 1, bottom) after surgery. Since the recurrence state is transient, this curve is first increasing and then decreasing. It is also evident that the probabilities for $s = 1095$ are smaller than those for $s = 365$. In Figure 1 the estimators $\widehat{p}_{12}^{new}(s,t)$ and $\widehat{p}_{12}^{*}(s,t)$ (labelled as New and LIDA* respectively) report roughly the same estimates. However, the values of the original LIDA estimator $\widehat{p}_{12}(s,t)$ are systematically smaller along most of the support of the observed survival time T, which suggests a negative bias. The reason for this is found in the truncated limits of $\widehat{p}_{12}(s,t)$ given in Theorem 1 in the online supplementary material; since the maximum value for T in the sample is censored ($\hat{\tau}_T = 3329$ days), $\widehat{p}_{12}(s,t)$ may be only interpreted as a probability restricted to that value, thus it underestimates the target of ultimate interest.

Similar conclusions can be obtained from Figure 2, in which the transition probability

$p_{23}(s,t)$ is estimated through the three methods again for $s = 365$ and $s = 1095$. Thus, these plots report one minus the survival fraction along time, among the individuals in the recurrence state one year (Figure 2, top) and three years (Figure 2, bottom) after surgery. The survivorship is smaller for the first group, suggesting a negative impact of an earlier recurrence time. For this transition probability the bias of the original LIDA estimator is more visible for the largest value of s. In this case, the estimator $\widehat{p}_{23}(s,t)$ provides a less optimistic prognosis than that of its competitors. This was expected since $\widehat{p}_{23}(s,t)$ is consistent only for the subpopulation dying before time $\hat{\tau}_T$. Summarising, it becomes clear from this application that the proposed methods overcome the bias of Meira-Machado et al. (2006)'s estimators when the support of the censoring variable is strictly contained in that of the lifetime variable, which often happens in medical applications due to time limitations in the follow-up.

In order to evaluate the accuracy of the estimators, Table 4 reports point estimates for $p_{12}(s,t)$ and $p_{23}(s,t)$ for the three methods, along with bootstrap variance (SD^2) and 95% confidence intervals (CI). We consider $s = 365$ (Figures 1 and 2, top) and $s = 1095$ (Figures 1 and 2, bottom) and several specific values for t. The CIs are based on a normal approximation and on the percentile bootstrap (NCI and PCI, respectively). The values of SD^2 in Table 4 indicate that the new estimator and the modified LIDA estimator have roughly the same accuracy, with some visible differences only for $p_{23}(s,t)$ and large values of t, when the efficiency of $\widehat{p}^{*}_{23}(s,t)$ relative to the new method $\widehat{p}^{new}_{23}(s,t)$ is ocasionally smaller than 95%. On the other hand, the variance of the original LIDA estimator is often much larger, particularly in the right tail of $p_{12}(s,t)$, where it is even one order of magnitude above the variance of its competitors. This is in agreement with the simulation results in the previous Section. Accordingly, the 95% CIs of the modified LIDA estimator and of the new method are clearly narrower than those of the original estimator $\widehat{p}_{ij}(s,t)$.

When comparing the normal approximation method with the percentile method, it is seen that both methods produce similar CIs and no method dominates the other in the sense of reporting systematically wider or narrower confidence intervals. For example, for the new estimator $\widehat{p}_{ij}^{new}(s,t)$, NCI are narrower than PCI about half of the times (10 out of the 18 cases in Table 4), although the differences between the two methods are small in any case. Similar comments apply to $\widehat{p}_{ij}(s,t)$ and $\widehat{p}_{ij}^{*}(s,t)$, which indicates that the normal distribution works well when approximating the sampling distribution of the estimators.

[Figure 1 about here.]

[Figure 2 about here.]

[Table 4 about here.]

5. Extension to a General Multi-State Model

The new method $\widehat{p}_{ij}^{new}(s,t)$ introduced in this paper to estimate the transition probabilities in a progressive non-Markov illness-death model can be easily extended to a more general multi-state model. In this Section we explore this possibility and we give the needed details.

Let $\{X(t), t \geqslant 0\}$ be a general multi-state model with state space $\mathcal{E}$, where $X(t)$ denotes the state occupied by the process at time t, $t \geqslant 0$. Assume that, for two specific states i, j in $\mathcal{E}$, one is interested in estimating the transition probability $p_{ij}(s,t) = P(X(t) = j | X(s) = i)$, $s \leqslant t$. Let $\mathcal{X}_j$ be the set of states from which j is reachable, and let Z^j and T^j denote the sojourn times in $\mathcal{X}_j$ and $\mathcal{X}_j \cup \{j\}$, respectively. Like for the progressive illness-death model, $Z^j = T^j$ may occur; indeed, this will be the case for the subpopulation not visiting state j. On the contrary, $Z^j < T^j$ indicates that the individual is visiting state j at some time. Moreover, with this notation

$$p_{ij}(s,t) = P(Z^j \leqslant t < T^j | X(s) = i).$$

Under independent censoring, this probability can be estimated from the subset of individuals observed to be in state i at time s, say $\mathcal{S}_i = \{k : X_k(s) = i, C_k > s\}$, where $X_k(.)$ is the trajectory of the k-th individual, and C_k denotes his/her potential censoring time. Interestingly, since $Z^j \leqslant T^j$, the following decomposition holds:

$$p_{ij}(s,t) = P(Z^j \leqslant t | X(s) = i, C > s) - P(T^j \leqslant t | X(s) = i, C > s).$$

Both terms at the right-hand side may be consistently estimated from the subsample S_i by computing the Kaplan-Meier estimators corresponding to Z^j and T^j, respectively, in such a subsample. Like for the illness-death model, standard software can be used to this end, once the subset $\mathcal{S}_i$ is constructed and the observed sojourn times Z^j and T^j (together with their corresponding censoring indicators) are computed. On the other hand, the asymptotic properties of this new estimator, say $\widehat{p}^g_{ij}(s,t)$, are easily derived from the properties of the Kaplan-Meier curve, similarly as for the progressive illness-death model. In particular, the estimator is consistent on the support of the censoring time. For particular states i, j, $\widehat{p}^g_{ij}(s,t)$ actually involves only one Kaplan-Meier estimator. This is the case for example when j is the initial state of the process; then $\mathcal{X}_j$ is empty, and $Z^j = 0$ with probability 1 by convention, from which $p_{ij}(s,t) = 1 - P(T^j \leqslant t | X(s) = i, C > s)$. The same occurs when $i = j$, since in such a case $X(s) = i$ implies $Z^j = Z^i < s$ and therefore $Z^j \leqslant t$. Finally, when the state j is the unique absorbing state of the process, one rather gets $T^j = \infty$ and $p_{ij}(s,t) = P(Z^j \leqslant t | X(s) = i, C > s)$, so again only one Kaplan-Meier estimator is involved in the estimation of $p_{ij}(s,t)$. When j is one out of several possible absorbing states, an empirical cumulative incidence function based on the subsample $\mathcal{S}_i$ may be used to estimate $p_{ij}(s,t)$ instead. In the uncensored case, $\widehat{p}^g_{ij}(s,t)$ is just the fraction of individuals in state j at time t among those being in state i at time s. Also, it is easily seen that $\widehat{p}^g_{ij}(s,t)$ reduces to the estimator $\widehat{p}^{new}_{ij}(s,t)$ when the multi-state model is the illness-death progressive model presented in Section 2; in that case, Z^j and T^j are respectively the sojourn time in state 1

and the total survival time if $j = 2$, while T^j is the sojourn time in state 1 and $Z^j = 0$ when $j = 1$.

It should be noted that the representation given for $p_{ij}(s,t)$ is valid for a general multi-state model as long as $\mathcal{X}_j$ does not contain j itself. If the process allows for more than one visit to state j, the equality $\{X(t) = j\} = \{Z^j \leqslant t < T^j\}$ is no longer true; this is the case for example in the three-state illness-death model with recovery (Hougaard, 2000), where the individual can go back to the initial state 1 from the intermediate state 2. However, the given formulation works for a fairly general multi-state model and, in particular, for any progressive multi-state model one may have in mind. Finally, for consistency purposes it is assumed that the potential censoring time C is independent of Z^j and T^j conditionally on $\{X(s) = i, C > s\}$; this holds in particular for every time point s and every pair of states (i,j) if C is independent of the process.

6. Main Conclusions

In this paper the problem of estimating a transition probability matrix in a non-Markov illness-death model has been reviewed, and two new sets of estimators have been proposed. The first one comes from a simple modification of the estimators in Meira-Machado et al. (2006) and looks for a correction of the 'bias problem' due to censoring. The second method proceeds by considering specific subsets of individuals (namely, those observed to be in a given state at a pre-specified time point s) for which the ordinary Kaplan-Meier survival function leads to a consistent estimator of the target. The provided simulations suggest that both approaches are preferable to the original estimators in Meira-Machado et al. (2006), since they often have less variance and, more importantly, they are able to remove the systematic bias in heavily censored scenarios. Between the two new methods, the one based on subsamples is recommended; it may be computed by using standard software (since it only involves two Kaplan-Meier survival functions) and, for most of the considered simulated

scenarios, its mean squared error is smaller than that corresponding to the modified Meira-Machado et al. (2006)'s estimators. More specifically, the relative efficiency of the approach based on subsamples may be as large as 1.9, depending on the sample size, the censoring percentage, and the particular target to be estimated.

Asymptotic normality of the proposed estimators has been established. The relative performance of the methods has been evaluated through the analysis of data from a clinical trial on colon cancer too. In this application, the original estimator in Meira-Machado et al. (2006) when estimating transition probabilities is clearly biased. The new methods overcome this issue and provide estimates with smaller variances. To estimate the variance, the simple bootstrap has been proposed. Besides, confidence intervals for the transition probabilities based on the normal approximation and on the percentile bootstrap have been provided for the colon cancer study. Both methods reported similar results.

Finally, the extension of the proposed estimators to multi-state models beyond the progressive illness-death model has also been discussed. In particular, it has been shown how the new method based on the computation of Kaplan-Meier curves in subsamples can be extended to any progressive multi-state model. Due to the good properties and simplicity of this approach, we believe that it may become a new standard for the estimation of transition probabilities in non-Markov multi-state models.

Acknowledgements

We thank an Associate Editor and a referee for comments and suggestions which greatly improved the presentation of this paper. Jacobo de Uña-Álvarez was supported by Grant MTM2011-23104 (FEDER support included) of the Spanish Ministry of Science. Luís Meira-Machado acknowledges financial support from FEDER Funds through "Programa Operacional Factores de Competitividade - COMPETE" and by Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia", in the form of grant PEst-OE/MAT/UI0013/2014.

References

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

Allignol, A., Beyersmann, J., Gerds, T., and Latouche, A. (2014). A competing risks approach for nonparametric estimation of transition probabilities in an non-Markov illness-death model. *Lifetime Data Analysis*, **20**, 495–513.

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

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

Andersen, P. K., Hansen, L. S., and Keiding, N. (1991). Nonparametric and semiparametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, **18**, 153–167.

Borgan, O. (2002). Estimation of covariate-dependent Markov transition probabilities from nested case-control data. *Statistical Methods in Medical Research*, **11**, 183–202.

Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, **5**, 315–327.

Datta, S. and Satten, G. A. (2001). Validity of the Aalen-Johansen estimators of stage occupancy probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics & Probability Letters*, **55**, 403–411.

Datta, S. and Satten, G. A. (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring. *Biometrics*, **58**, 792–802.

de Uña-Álvarez, J. (2010). Recent Developments in Censored, non-Markov Multi-State Models. In. Combining Soft Computing and Statistical Methods in Data Analysis (Borgelt et al., Eds.) Advances in Intelligent and Soft Computing, Springer, pp. 173–179.

Frydman, H., Gerds, T., Gron, R., and Keiding, N. (2013). Nonparametric estimation in a Markov 'illness-death' model when all transition times are interval censored. *Biometrical Journal*, **55**, 823–843.

Frydman, H. and Szarek, M. (2009). Nonparametric estimation in a Markov 'illness-death' process from interval censored observations with missing intermediate transition status. *Biometrics*, **65**, 143–151.

Glidden, D. (2002). Robust inference for event probabilities with non-Markov event data. *Biometrics*, **58**, 361–368

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

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

Lan, L. and Datta, S. (2010). Non-parametric estimation of state occupation, entry and exit times with multistate current status data. *Statistical Methods in Medical Research*, **19**, 147–165.

Meira-Machado, L. F., de Uña-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric 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. K. (2009). Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, **18**, 195–222.

Meira-Machado, L., Roca-Pardiñas, J., Van Keilegom, I., and Cadarso-Suárez, C. (2013). Bandwidth selection for the estimation of transition probabilities in the location-scale progressive three-state model. *Computational Statistics*, **28**, 2185–2210.

Moreira, A., de Uña-Álvarez, J., and Machado, L. (2013) Presmoothing the Aalen-Johansen estimator in the illness-death model. *Electronic Journal of Statistics*, **7**, 1491–1516.

Moertel, C. G., Fleming, T. R., Macdonald, J.S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., Veeder, M. H., and Mailliard, J. A. (1990) Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, **332**, 352–358.

Mostajabi, F. and Datta, S. (2013) Nonparametric regression of state occupation, entry, exit, and waiting times with multistage right-censored data. *Statistics in Medicine*, **32**, 3006–3019.

Pepe, M. S. (1991) Inference for events with dependent risks in multiple end-point studies. *Journal of the American Statistical Association*, **86**, 770–778.

Pepe, M. S., Longton, G., and Thornquist, M. (1991) A qualifier Q for the survival function to describe the prevalence of a transient condition. *Statistics in Medicine*, **10**, 413–421.

Satten, G. A. (1999). Estimating the extent of tracking in interval-censored chain-of-events data. *Biometrics*, **55**, 1228–1231.

Spitoni, C., Verduijn, M., and Putter, H. (2012). Estimation and asymptotic theory for transition probabilities in Markov renewal multi-state models. *International Journal of Biostatistics*, **8**, Number 1, Article 23.

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

Stute, W. (1996) Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics*, **23**, 461–471.

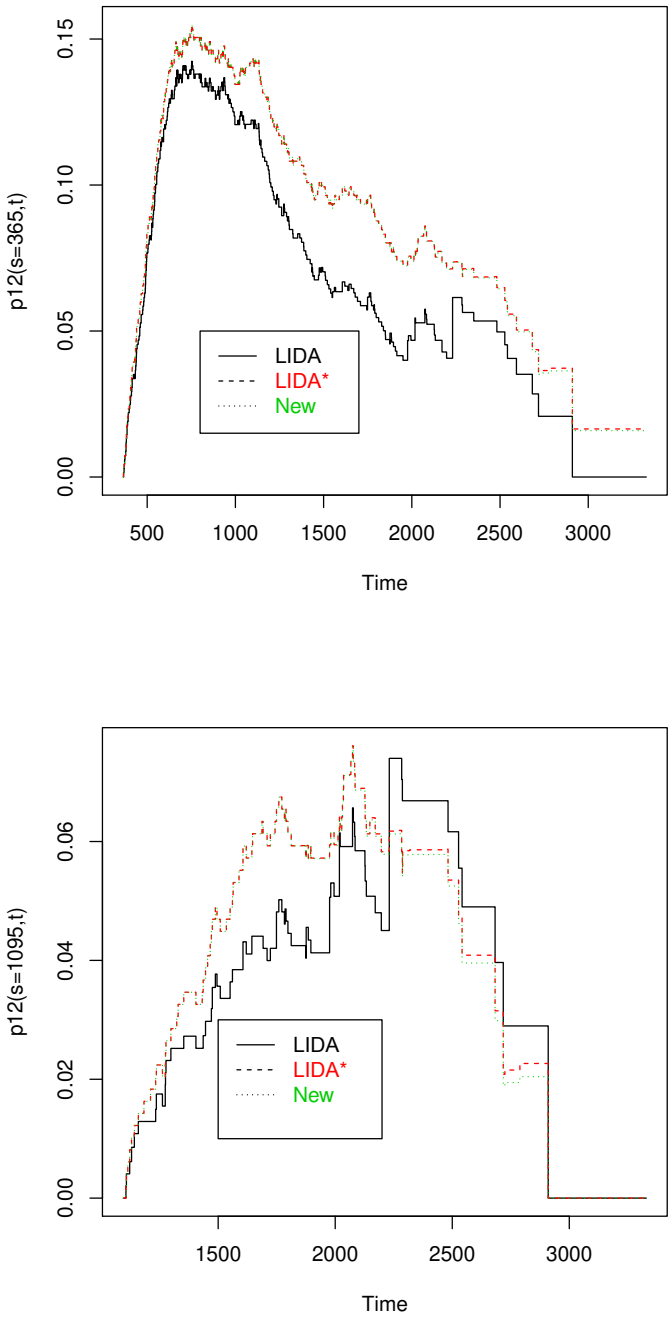

Figure 1. Estimated transition probabilities for $p_{12}(s,t)$, $s = 365$ (top) and $s = 1095$ bottom. Colon cancer data.

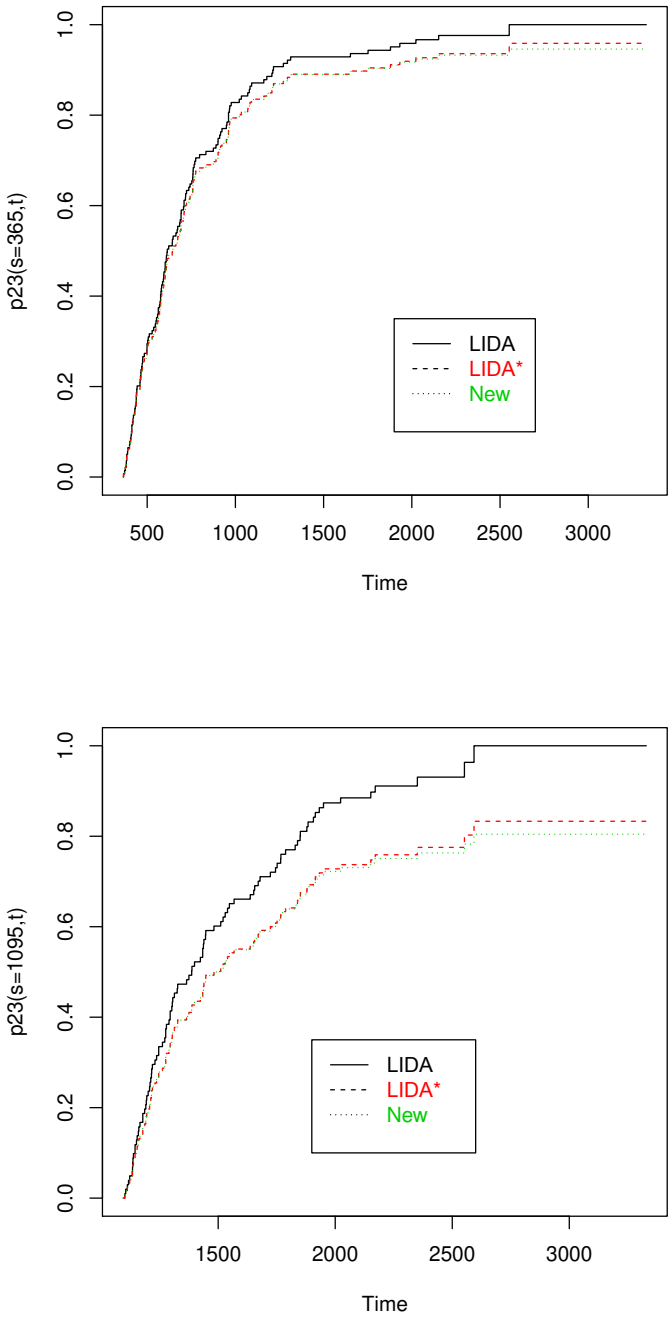

Figure 2. Estimated transition probabilities for $p_{23}(s,t)$, $s = 365$ (top) and $s = 1095$ bottom. Colon cancer data.

Table 1

Bias and standard deviation (SD) for the three estimators of $p_{ij}(s,t)$. The MSE of $\widehat{p}_{ij}^{new}(s,t)$ relative to $\widehat{p}_{ij}^{*}(s,t)$ is also given. Scenario 1: correlated exponential gap times, $p = 0.7$.

		$\widehat{p}_{12}(s,t)$		$\widehat{p}_{12}^{*}(s,t)$		$\widehat{p}_{12}^{new}(s,t)$		
		bias	SD	bias	SD	bias	SD	MSE^{new}/MSE^{*}
(s,t)=	(.2231,.5108)							
n=50	$C \sim U[0,4]$	-0.0086	0.0705	-0.0020	0.0577	-0.0021	0.0577	1.0021
	$C \sim U[0,3]$	-0.0106	0.0791	-0.0025	0.0593	-0.0025	0.0593	1.0006
n=100	$C \sim U[0,4]$	-0.0037	0.0506	-0.0012	0.0401	-0.0012	0.0401	1.0019
	$C \sim U[0,3]$	-0.0115	0.0547	0.0007	0.0429	0.0007	0.0429	1.0011
n=250	$C \sim U[0,4]$	-0.0094	0.0325	-0.0001	0.0255	-0.0001	0.0256	1.0031
	$C \sim U[0,3]$	-0.0106	0.0356	0.0002	0.0260	0.0001	0.0260	1.0030
(s,t)=	(.2231,1.6094)							
n=50	$C \sim U[0,4]$	-0.0396	0.1103	0.0010	0.0820	0.0010	0.0814	0.9859
	$C \sim U[0,3]$	-0.0726	0.1292	0.0023	0.0925	0.0019	0.0922	0.9936
n=100	$C \sim U[0,4]$	-0.0319	0.0816	0.0021	0.0593	0.0019	0.0588	0.9798
	$C \sim U[0,3]$	-0.0759	0.0933	-0.0011	0.0634	-0.0008	0.0623	0.9662
n=250	$C \sim U[0,4]$	-0.0258	0.0557	0.0012	0.0379	0.0012	0.0377	0.9882
	$C \sim U[0,3]$	-0.0700	0.0644	0.0001	0.0423	0.0001	0.0419	0.9782
(s,t)=	(.9163,1.6094)							
n=50	$C \sim U[0,4]$	-0.0533	0.1697	0.0057	0.1292	0.0057	0.1276	0.9737
	$C \sim U[0,3]$	-0.1094	0.1842	-0.0051	0.1381	-0.0042	0.1361	0.9706
n=100	$C \sim U[0,4]$	-0.0405	0.1275	-0.0023	0.0894	-0.0024	0.0886	0.9814
	$C \sim U[0,3]$	-0.0902	0.1414	0.0038	0.0954	0.0028	0.0947	0.9843
n=250	$C \sim U[0,4]$	-0.0345	0.0678	-0.0001	0.0538	0.0001	0.0531	0.9751
	$C \sim U[0,3]$	-0.0924	0.0938	-0.0039	0.0574	-0.0038	0.0565	0.9677
		$\widehat{p}_{23}(s,t)$		$\widehat{p}_{23}^{*}(s,t)$		$\widehat{p}_{23}^{new}(s,t)$		
(s,t)=	(.2231,.5108)							
n=50	$C \sim U[0,4]$	0.0314	0.2718	-0.0063	0.2471	0.0034	0.2453	0.9850
	$C \sim U[0,3]$	0.0331	0.2872	-0.0048	0.2497	-0.0094	0.2471	0.9800
n=100	$C \sim U[0,4]$	0.0134	0.1651	0.0026	0.1651	0.0031	0.1641	0.9179
	$C \sim U[0,3]$	0.0188	0.1875	-0.0036	0.1664	-0.0028	0.1639	0.9702
n=250	$C \sim U[0,4]$	-0.0013	0.1059	-0.0074	0.1011	-0.0077	0.0995	0.9223
	$C \sim U[0,3]$	0.0048	0.1107	-0.0071	0.1029	-0.0075	0.1005	0.9551
(s,t)=	(.2231,1.6094)							
n=50	$C \sim U[0,4]$	0.0304	0.1551	-0.0429	0.1819	-0.0184	0.1781	0.9882
	$C \sim U[0,3]$	0.0399	0.1665	-0.0589	0.2090	-0.0410	0.2099	0.9702
n=100	$C \sim U[0,4]$	0.0180	0.1257	-0.0205	0.1333	-0.0046	0.1176	0.7619
	$C \sim U[0,3]$	0.0444	0.1228	-0.0333	0.1481	-0.0129	0.1335	0.7809
n=250	$C \sim U[0,4]$	0.0087	0.0900	-0.0065	0.0871	-0.0013	0.0725	0.8786
	$C \sim U[0,3]$	0.0295	0.0933	-0.0079	0.0961	0.0003	0.0760	0.6213
(s,t)=	(.9163,1.6094)							
n=50	$C \sim U[0,4]$	0.0589	0.2410	-0.0025	0.2087	-0.0045	0.2003	0.9696
	$C \sim U[0,3]$	0.1453	0.2738	0.0034	0.2309	0.0082	0.2263	0.9613
n=100	$C \sim U[0,4]$	0.0515	0.1656	0.0079	0.1420	0.0067	0.1331	0.6895
	$C \sim U[0,3]$	0.0963	0.2028	0.0019	0.1596	-0.0042	0.1457	0.8332
n=250	$C \sim U[0,4]$	0.0306	0.1078	-0.0015	0.0930	-0.0013	0.0872	0.8793
	$C \sim U[0,3]$	0.0856	0.1295	0.0013	0.1023	0.0005	0.0939	0.8416

Table 2

Bias and standard deviation (SD) for the three estimators of $p_{ij}(s,t)$. The MSE of $\widehat{p}_{ij}^{new}(s,t)$ relative to $\widehat{p}_{ij}^{*}(s,t)$ is also given. Scenario 2: correlated loglogistic gap times, $p = 0.7$.

		$\widehat{p}_{12}(s,t)$		$\widehat{p}_{12}^{*}(s,t)$		$\widehat{p}_{12}^{new}(s,t)$		
		bias	SD	bias	SD	bias	SD	MSE^{new}/MSE^{*}
(s,t)=	(.2500,.6667)							
n=50	$C \sim U[0,8]$	-0.0044	0.0604	-0.0012	0.0538	-0.0012	0.0538	0.9983
	$C \sim U[0,5]$	-0.0096	0.0652	-0.0016	0.0550	-0.0015	0.0550	0.9974
n=100	$C \sim U[0,8]$	-0.0033	0.0462	-0.0007	0.0392	-0.0007	0.0392	1.0000
	$C \sim U[0,5]$	-0.0076	0.0483	-0.0011	0.0403	-0.0011	0.0403	1.0000
n=250	$C \sim U[0,8]$	-0.0025	0.0287	-0.0003	0.0242	-0.0003	0.0242	0.9983
	$C \sim U[0,5]$	-0.0065	0.0308	0.0006	0.0246	-0.0006	0.0246	1.0000
(s,t)=	(.2500,4.0000)							
n=50	$C \sim U[0,8]$	-0.0385	0.0887	0.0033	0.0731	0.0037	0.0719	0.9695
	$C \sim U[0,5]$	-0.0992	0.0815	0.0071	0.0969	0.0069	0.0960	0.9823
n=100	$C \sim U[0,8]$	-0.0344	0.0673	0.0030	0.0507	0.0029	0.0502	0.9764
	$C \sim U[0,5]$	-0.0943	0.0682	0.0033	0.0715	0.0033	0.0704	0.9718
n=250	$C \sim U[0,8]$	-0.0339	0.0469	0.0012	0.0322	0.0011	0.0318	0.9721
	$C \sim U[0,5]$	-0.0904	0.0526	0.0013	0.0443	0.0012	0.0438	0.9766
(s,t)=	(1.5000,4.0000)							
n=50	$C \sim U[0,8]$	-0.0579	0.1550	0.0063	0.1204	0.0073	0.1183	0.9671
	$C \sim U[0,5]$	-0.1505	0.1327	0.0120	0.1641	0.0126	0.1612	0.9664
n=100	$C \sim U[0,8]$	-0.0520	0.1221	0.0054	0.0875	0.0056	0.0859	0.9636
	$C \sim U[0,5]$	-0.1430	0.1143	0.0085	0.1207	0.0078	0.1198	0.9858
n=250	$C \sim U[0,8]$	-0.0537	0.0795	0.0021	0.0562	0.0022	0.0554	0.9696
	$C \sim U[0,5]$	-0.1376	0.0881	0.0025	0.0760	0.0027	0.0748	0.9690
		$\widehat{p}_{23}(s,t)$		$\widehat{p}_{23}^{*}(s,t)$		$\widehat{p}_{23}^{new}(s,t)$		
(s,t)=	(.2500,.6667)							
n=50	$C \sim U[0,8]$	0.0148	0.2832	-0.0036	0.2634	-0.0009	0.2661	1.0210
	$C \sim U[0,5]$	0.0275	0.2998	-0.0114	0.2668	-0.0041	0.2737	1.0504
n=100	$C \sim U[0,8]$	0.0106	0.1862	0.0013	0.1780	0.0009	0.1769	0.9869
	$C \sim U[0,5]$	0.0247	0.1994	0.0000	0.1821	0.0016	0.1822	1.0018
n=250	$C \sim U[0,8]$	0.0056	0.1134	-0.0014	0.1059	-0.0019	0.1042	0.9686
	$C \sim U[0,5]$	0.0146	0.1220	-0.0012	0.1098	-0.0017	0.1074	0.9581
(s,t)=	(.2500,4.0000)							
n=50	$C \sim U[0,8]$	0.0072	0.1083	-0.0560	0.1427	-0.0295	0.1337	0.7977
	$C \sim U[0,5]$	0.0138	0.1095	-0.0853	0.1679	-0.0507	0.1597	0.7914
n=100	$C \sim U[0,8]$	0.0118	0.0650	-0.0320	0.0850	-0.0104	0.0714	0.6313
	$C \sim U[0,5]$	0.0207	0.0607	-0.0527	0.1115	-0.0288	0.0963	0.6651
n=250	$C \sim U[0,8]$	0.0090	0.0506	-0.0165	0.0546	-0.0044	0.0414	0.5305
	$C \sim U[0,5]$	0.0197	0.0513	-0.0271	0.0665	-0.0137	0.0529	0.5792
(s,t)=	(1.5000,4.0000)							
n=50	$C \sim U[0,8]$	0.0734	0.2131	-0.0350	0.2105	-0.0189	0.2062	0.9413
	$C \sim U[0,5]$	0.1285	0.2231	-0.0741	0.2651	-0.0657	0.2632	0.9711
n=100	$C \sim U[0,8]$	0.0596	0.1586	-0.0097	0.1485	-0.0037	0.1350	0.8238
	$C \sim U[0,5]$	0.1441	0.1477	-0.0273	0.1900	-0.0178	0.1789	0.8772
n=250	$C \sim U[0,8]$	0.0441	0.1116	-0.0015	0.1005	-0.0009	0.0835	0.6898
	$C \sim U[0,5]$	0.1291	0.1223	-0.0067	0.1330	-0.0050	0.1134	0.7263

Table 3

Bias and standard deviation (SD) for the three estimators of $p_{ij}(s,t)$. The MSE of $\widehat{p}_{ij}^{new}(s,t)$ relative to $\widehat{p}_{ij}^{*}(s,t)$ is also given. Scenario 3: correlated exponential gap times, $p(Z^0) = exp(-(3/7)Z^0)$.

		$\widehat{p}_{ij}(s,t)$		$\widehat{p}_{ij}^{*}(s,t)$		$\widehat{p}_{ij}^{new}(s,t)$		
		bias	SD	bias	SD	bias	SD	MSE^{new}/MSE^{*}
(s,t)=	(.2231,.5108)							
n=50	$C \sim U[0,4]$	-0.0090	0.0741	-0.0010	0.0634	-0.0011	0.0633	0.9985
	$C \sim U[0,3]$	-0.0106	0.0791	-0.0025	0.0593	-0.0025	0.0593	1.0006
n=100	$C \sim U[0,4]$	-0.0065	0.0549	-0.0012	0.0452	-0.0012	0.0452	0.9990
	$C \sim U[0,3]$	-0.0115	0.0547	0.0007	0.0429	0.0007	0.0429	1.0011
n=250	$C \sim U[0,4]$	-0.0057	0.0349	-0.0002	0.0273	-0.0002	0.0273	0.9987
	$C \sim U[0,3]$	-0.0106	0.0356	0.0002	0.0260	0.0001	0.0260	1.0030
(s,t)=	(.2231,1.6094)							
n=50	$C \sim U[0,4]$	-0.0344	0.1044	0.0036	0.0854	0.0029	0.0846	0.9788
	$C \sim U[0,3]$	-0.0726	0.1292	0.0023	0.0925	0.0019	0.0922	0.9936
n=100	$C \sim U[0,4]$	-0.0295	0.0786	0.0025	0.0602	0.0020	0.0599	0.9887
	$C \sim U[0,3]$	-0.0759	0.0933	-0.0011	0.0634	-0.0008	0.0623	0.9662
n=250	$C \sim U[0,4]$	-0.0257	0.0521	0.0019	0.0386	0.0019	0.0379	0.9618
	$C \sim U[0,3]$	-0.0700	0.0644	0.0001	0.0423	0.0001	0.0419	0.9782
(s,t)=	(.9163,1.6094)							
n=50	$C \sim U[0,4]$	-0.0403	0.1544	0.0035	0.1230	0.0037	0.1205	0.9604
	$C \sim U[0,3]$	-0.1094	0.1842	-0.0051	0.1381	-0.0042	0.1361	0.9706
n=100	$C \sim U[0,4]$	-0.0327	0.0963	0.0027	0.0849	0.0023	0.0838	0.9734
	$C \sim U[0,3]$	-0.0902	0.1414	0.0038	0.0954	0.0028	0.0947	0.9843
n=250	$C \sim U[0,4]$	-0.0282	0.0632	0.0033	0.0529	0.0038	0.0518	0.9577
	$C \sim U[0,3]$	-0.0924	0.0938	-0.0039	0.0574	-0.0038	0.0565	0.9677
		$\widehat{p}_{23}(s,t)$		$\widehat{p}_{23}^{*}(s,t)$		$\widehat{p}_{23}^{new}(s,t)$		
(s,t)=	(.2231,.5108)							
n=50	$C \sim U[0,4]$	0.0135	0.2241	-0.0045	0.1981	-0.0058	0.1976	0.9950
	$C \sim U[0,3]$	0.0230	0.2372	-0.0026	0.2033	-0.0019	0.2056	1.0226
n=100	$C \sim U[0,4]$	0.0045	0.1439	-0.0038	0.1322	-0.0044	0.1301	0.9695
	$C \sim U[0,3]$	0.0129	0.1549	-0.0042	0.1361	-0.0042	0.1344	0.9749
n=250	$C \sim U[0,4]$	0.0061	0.0899	0.0013	0.0840	0.0013	0.0831	0.9796
	$C \sim U[0,3]$	0.0138	0.0957	0.0011	0.0866	0.0017	0.0853	0.9696
(s,t)=	(.2231,1.6094)							
n=50	$C \sim U[0,4]$	0.0097	0.1597	-0.0361	0.1565	-0.0201	0.1533	0.9270
	$C \sim U[0,3]$	0.0226	0.1664	-0.0441	0.1743	-0.0287	0.1654	0.8722
n=100	$C \sim U[0,4]$	0.0059	0.1144	-0.0184	0.1127	-0.0107	0.1031	0.8236
	$C \sim U[0,3]$	0.0207	0.1190	-0.0253	0.1245	-0.0146	0.1103	0.7667
n=250	$C \sim U[0,4]$	0.0055	0.0744	-0.0055	0.0739	-0.0029	0.0606	0.6701
	$C \sim U[0,3]$	0.0192	0.0808	-0.0103	0.0818	-0.0051	0.0661	0.6453
(s,t)=	(.9163,1.6094)							
n=50	$C \sim U[0,4]$	0.0486	0.2174	-0.0050	0.1967	-0.0048	0.1865	0.8988
	$C \sim U[0,3]$	0.1158	0.2606	-0.0063	0.2156	-0.0062	0.2044	0.8986
n=100	$C \sim U[0,4]$	0.0441	0.1549	0.0035	0.1382	0.0030	0.1285	0.8638
	$C \sim U[0,3]$	0.1047	0.1827	0.0046	0.1518	0.0026	0.1384	0.8306
n=250	$C \sim U[0,4]$	0.0379	0.1001	0.0054	0.0852	-0.0029	0.0606	0.9056
	$C \sim U[0,3]$	0.0876	0.1193	0.0049	0.0934	0.0069	0.0881	0.8922

Table 4

Point estimates for $p_{12}(s,t)$ and $p_{23}(s,t)$ and bootstrap variances (SD^2) for the three methods, along with 95% bootstrap confidence intervals based on the normal approximation (NCI) and the percentile method (PCI). Colon cancer study.

		$\widehat{p}_{ij}(s,t)$	NCI	$\widehat{p}^{*}_{ij}(s,t)$	NCI	$\widehat{p}^{new}_{ij}(s,t)$	NCI
		$(SD^2 \times 10^4)$	PCI	$(SD^2 \times 10^4)$	PCI	$(SD^2 \times 10^4)$	PCI
	$p_{12}(s,t)$	$s = 365$					
$t = 500$		0.0768	[0.0570; 0.0966]	0.0832	[0.0629; 0.1035]	0.0833	[0.0630; 0.1036]
		(1.0206)	[0.0567; 0.0964]	(1.0727)	[0.0633; 0.1033]	(1.0728)	[0.0633; 0.1033]
$t = 1000$		0.1206	[0.0945; 0.1467]	0.1345	[0.1089; 0.1601]	0.1344	[0.1088; 0.1600]
		(1.7733)	[0.0937; 0.1482]	(1.0706)	[0.1098; 0.1616]	(1.0709)	[0.1098; 0.1016]
$t = 1500$		0.0703	[0.0481; 0.0925]	0.0994	[0.0768; 0.1220]	0.0993	[0.0767; 0.1219]
		(1.2830)	[0.0498; 0.0918]	(1.3296)	[0.0757; 0.1222]	(1.3339)	[0.0756; 0.1221]
$t = 2000$		0.0468	[0.0220; 0.0716]	0.0740	[0.0574; 0.0933]	0.0739	[0.0546; 0.0932]
		(1.6011)	[0.0243; 0.0720]	(0.9658)	[0.0560; 0.0939]	(0.9682)	[0.0558; 0.0938]
$t = 2500$		0.0497	[0.0003; 0.0991]	0.0649	[0.0429; 0.0869]	0.0644	[0.0423; 0.0865]
		(6.3530)	[0.0088; 0.1084]	(1.2599)	[0.0429; 0.0871]	(1.2702)	[0.0427; 0.0867]
	$p_{12}(s,t)$	$s = 1095$					
$t = 1200$		0.0129	[0.0028; 0.0230]	0.0163	[0.0053; 0.0273]	0.0163	[0.0053; 0.0273]
		(0.2636)	[0.0040; 0.0236]	(0.3150)	[0.0060; 0.0284]	(0.3170)	[0.0060; 0.0284]
$t = 1500$		0.0357	[0.0166; 0.0548]	0.0469	[0.0266; 0.0672]	0.0469	[0.0266; 0.0672]
		(0.9497)	[0.0181; 0.0567]	(1.0727)	[0.0284; 0.0693]	(1.0732)	[0.0284; 0.0693]
$t = 2000$		0.0508	[0.0207; 0.0809]	0.0595	[0.0374; 0.0816]	0.0593	[0.0372; 0.0814]
		(2.3545)	[0.0228; 0.0830]	(1.2714)	[0.0391; 0.0815]	(1.2718)	[0.0389; 0.0815]
$t = 2500$		0.0616	[0.0000; 0.1272]	0.0535	[0.0290; 0.0780]	0.0525	[0.0279; 0.0771]
		(11.2024)	[0.0070; 0.1358]	(1.5626)	[0.0290; 0.0787]	(1.5736)	[0.0275; 0.0779]
	$p_{23}(s,t)$	$s = 365$					
$t = 500$		0.2950	[0.2148; 0.3752]	0.2829	[0.2060; 0.3598]	0.2828	[0.2059; 0.3597]
		(16.7437)	[0.2148; 0.3836]	(15.3942)	[0.2055; 0.3668]	(15.4342)	[0.2055; 0.3667]
$t = 1000$		0.8280	[0.7573; 0.8987]	0.7939	[0.7268; 0.8610]	0.7931	[0.7262; 0.8600]
		(13.0120)	[0.7558; 0.8935]	(11.7206)	[0.7218; 0.8561]	(11.6508)	[0.7214; 0.8542]
$t = 1500$		0.9288	[0.8740; 0.9836]	0.8906	[0.8408; 0.9404]	0.8897	[0.8400; 0.9394]
		(7.8174)	[0.8690; 0.9792]	(6.4660)	[0.8356; 0.9338]	(6.4244)	[0.8356; 0.9329]
$t = 2000$		0.9586	[0.9092; 1.0000]	0.9191	[0.8746; 0.9636]	0.9172	[0.8730; 0.9614]
		(6.3527)	[0.9032; 1.0000]	(5.1549)	[0.8684; 0.9620]	(5.0847)	[0.8671; 0.9592]
$t = 2500$		0.9761	[0.9310; 1.0000]	0.9359	[0.8939; 0.9778]	0.9326	[0.8910; 0.9742]
		(5.3549)	[0.9228; 1.0000]	(4.5840)	[0.8871; 0.9734]	(4.5050)	[0.8852; 0.9696]
	$p_{23}(s,t)$	$s = 1095$					
$t = 1200$		0.2266	[0.1415; 0.3117]	0.1888	[0.1192; 0.2584]	0.1885	[0.1189; 0.2581]
		(18.8492)	[0.1486; 0.3172]	(12.6102)	[0.1251; 0.2631]	(12.6100)	[0.1250; 0.2628]
$t = 1500$		0.6014	[0.4903; 0.7125]	0.5011	[0.4096; 0.5926]	0.5000	[0.4086; 0.5914]
		(32.1316)	[0.4818; 0.7104]	(21.7945)	[0.4094; 0.5922]	(21.7468)	[0.4071; 0.5918]
$t = 2000$		0.8737	[0.7699; 0.9775]	0.7279	[0.6432; 0.8126]	0.7222	[0.6384; 0.8060]
		(28.0478)	[0.7639; 0.9677]	(18.6754)	[0.6433; 0.8100]	(18.2807)	[0.6374; 0.8017]
$t = 2500$		0.9307	[0.8397; 1.0000]	0.7754	[0.6924; 0.8584]	0.7630	[0.6824; 0.8436]
		(21.5469)	[0.8327; 1.0000]	(17.9333)	[0.6878; 0.8658]	(16.9112)	[0.6769; 0.8496]