Supplementary Material
Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Acyclic Multi-State Models: Application to LTC Insurance

Quentin Guibert*[1,2] and Frédéric Planchet†[1,2]

[1]Univ Lyon, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances (ISFA), Laboratoire SAF EA2429, F-69366, Lyon, France
[2]Prim'Act, 42 avenue de la Grande Armée, 75017 Paris, France

April 30, 2017

*Email: q.guibert@hotmail.fr.
†Email: frederic.planchet@univ-lyon1.fr.

The aim of this document is to provide additional information on the paper "Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Acyclic Multi-State Models: Application to LTC Insurance". This document is organized as follows. Section 1 presents a general estimation framework with right-censoring and theoretical results for competing risks data which is quite close to those introduced by Suzukawa (2002). To obtain asymptotic results for transition probabilities, it is needed to extend his framework with the addition of covariates. Section 2 explains how we applied these results to transition probabilities of interest. Inclusion of left-truncation is considered in Section 3. Some additional simulation results are reported in Section 4. Finally, the proofs for results of Section 1 are presented in Section 5.

1 Non-parametric estimation, asymptotic properties and applications

Our aim here is to explain rigorously how the non-parametric estimation framework introduced in Section 3 of the paper is built and to prove asymptotic properties for our estimators. In Section 1.1, we recall some notations, introduce basic assumptions and then define general non-parametric estimators for integrals of the type $\int \varphi dF_0^{(v)}$, where φ is $F_0^{(v)}$-measurable with the cumulative incidence function $F_0^{(v)}$ associated to competing risks data. Section 1.2 contains the asymptotic results for such integrals.

1.1 Setup

The problem that we consider entails a unique right-censoring process C with a continuous distribution function G.

Assumption 1. C is independent of the vector (S, T, V).

Thus, the following variables are available

$$\begin{cases} Y = \min(S, C) \text{ and } \gamma = \mathbb{1}_{\{S \leqslant C\}}, \\ Z = \min(T, C) \text{ and } \delta = \mathbb{1}_{\{T \leqslant C\}}. \end{cases}$$

For the sake of generality, we incorporate here a vector $\boldsymbol{\Theta} = (\Theta_i)_{i=1,\dots,p}$ of p-covariates. Following Stute (1993), we only assume[1] that these covariates do not provide any further information as to whether censoring will take place or not, i.e. we have

Assumption 2.

i. $\mathbb{P}(S \leqslant C \mid S, \boldsymbol{\Theta}, V_1) = \mathbb{P}(S \leqslant C \mid S, V_1)$,

ii. $\mathbb{P}(T \leqslant C \mid S, T, \boldsymbol{\Theta}, V) = \mathbb{P}(T \leqslant C \mid T, V)$.

Equality **ii.** of Assumption 2 is explained by the fact that the pair (S, T) is, by construction, subject to censoring and S is uncensored whenever T is. Let $\mathcal{V}$ be the set of values taken by $V = (V_1, V_2)$. For estimation purposes, we introduce the distribution function of $(S, \boldsymbol{\Theta})$, noted $H_0(s, \boldsymbol{\theta}) = \mathbb{P}(S \leqslant s, \boldsymbol{\Theta} \leqslant \boldsymbol{\theta})$, and $F_0^{(v)}(s, t, \boldsymbol{\theta}) = \mathbb{P}(S \leqslant s, T \leqslant t, \boldsymbol{\Theta} \leqslant \boldsymbol{\theta}, V = v)$, the sub-distribution function[2] of $(S, T, \boldsymbol{\Theta})$, where the cause is $V = v$ with $v \in \mathcal{V}$. Moreover, we have $F_0 = \sum_v F_0^{(v)}$.

[1] No assumption is made about the dependence structure between $(C, S, T, \boldsymbol{\Theta}, V)$.
[2] Also called cumulative incidence function.

The theoretical results presented below can be applied both for discrete (e.g. gender, geographical location, social status) and continuous (e.g. biomedical measures) covariates. However, continuous covariates should be beforehand transformed into categorical variables for practical applications. Another solution could be to smooth the multivariate cumulative incidence function $F_0^{(v)}$ for each competing cause. Note in addition that this condition is satisfied if C is independent of $(S, T, V, \boldsymbol{\Theta})$, which would also be a realistic assumption in our insurance application.

In this context, the observation of the i-th individual of a sample of length $n \geqslant 1$ is characterized by

$$(Y_i, \gamma_i, \gamma_i V_{1,i}, Z_i, \delta_i, \delta_i V_{2,i}, \boldsymbol{\Theta}_i) \ \ 1 \leqslant i \leqslant n \ ,$$

which are assumed to be i.i.d. replications of the variable $(Y, \gamma, \gamma V_1, Z, \delta, \delta V_2, \boldsymbol{\Theta})$. If $\delta = 1$, then obviously $\gamma = 1$. Consider the ordered Y-values $Y_{1:n} \leqslant Y_{2:n} \leqslant \ldots \leqslant Y_{n:n}$ and $\left(\gamma_{[i:n]}, \boldsymbol{\Theta}_{[i:n]}\right)$ the concomitant of the i-th order statistic (i.e. the value of $(\gamma_j, \boldsymbol{\Theta}_j)_{1 \leqslant j \leqslant n}$ paired with $Y_{i:n}$). An estimator for H_0 is simply obtained from the multivariate Kaplan-Meier estimator considered by Stute (1993)

$$\widehat{H}_{0n}(s, \boldsymbol{\theta}) = \sum_{i=1}^{n} W_{in} \mathbb{1}_{\left\{Y_{i:n} \leqslant s, \boldsymbol{\Theta}_{[i:n]} \leqslant \boldsymbol{\theta}\right\}}. \tag{1.1}$$

Kaplan-Meier integrals, taking the form $S(\varphi) = \int \varphi \, dH_0$ with some generic function φ, are estimated with

$$\widehat{S}_n(\varphi) = \int \varphi(s, \boldsymbol{\theta}) \ \widehat{H}_{0n}(ds, d\boldsymbol{\theta}) = \sum_{i=1}^{n} W_{in} \varphi\left(Y_{i:n}, \boldsymbol{\Theta}_{[i:n]}\right).$$

Let $Z_{1:n} \leqslant Z_{2:n} \leqslant \ldots \leqslant Z_{n:n}$ be the ordered Z-values and $\left(Y_{[i:n]}, \delta_{[i:n]}, J_{[i:n]}^{(v)}, \boldsymbol{\Theta}_{[i:n]}\right)$ be the concomitant of the i-th order statistic with $J_i^{(v)} = \mathbb{1}_{\{V=v,\}}$ and $v \in \mathcal{V}$. Since S is assimilated to an uncensored covariate when T is not censored, the cumulative incidence function $F_0^{(v)}$ can be estimated with the so-called Aalen-Johansen estimator for competing risks data with covariates

$$\begin{aligned} \widehat{F}_{0n}^{(v)}(y, z, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \mathbb{1}_{\left\{Y_{[i:n]} \leqslant y, Z_{i:n} \leqslant z, \boldsymbol{\Theta}_{[i:n]} \leqslant \boldsymbol{\theta}\right\}} \\ &= \sum_{i=1}^{n} \widetilde{W}_{in} J_{[i:n]}^{(v)} \mathbb{1}_{\left\{Y_{[i:n]} \leqslant y, Z_{i:n} \leqslant z, \boldsymbol{\Theta}_{[i:n]} \leqslant \boldsymbol{\theta}\right\}}, \end{aligned} \tag{1.2}$$

Based on the representation as a sum of (1.2), we want to obtain estimators of general quantities $S^{(v)}(\varphi) = \int \varphi \, dF_0^{(v)}$ with φ a generic function. In absence of censoring variable, non-parametric estimation is straightforward, resulting to integrals under the empirical multivariate distribution function of $(S, T, \boldsymbol{\Theta})$. In this context, complete information is available and each observation has the same weight into the empirical process. Since the joint distribution of (T, V) has the aspect of a competing risks model, we estimate $S^{(v)}(\varphi)$, for each $v \in \mathcal{V}$ by computing the Aalen-Johansen integral of the form

$$\widehat{S}_n^{(v)}(\varphi) = \int \varphi(s, t, \boldsymbol{\theta}) \ \widehat{F}_{0n}^{(v)}(ds, dt, d\boldsymbol{\theta}) = \sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{\Theta}_{[i:n]}\right). \tag{1.3}$$

These estimators are not so hard to handle and are similar to those exhibited by Suzukawa (2002), but we refine his approach by the introduction of covariates[3]. In this context, we need to derive new asymptotic properties.

[3] $\boldsymbol{\Theta}$ is covariate and S plays the same role as a covariate in (1.3).

1.2 Asymptotic properties

Let τ_Y and τ_Z be the least upper bounds of the distribution functions of Y and Z. Under Assumptions 1 and 2, the consistency and weak convergence of estimator (1.1) on $[0, \tau_Y]$ can be easily demonstrated, since Stute (1993) and Stute (1996) conditions are satisfied. Note that this result is also verified if H and G have no jump in common which is a less restrictive condition than continuity. For the rest of this subsection, we focus on the consistency and weak convergence properties of estimator (1.3).

Theorem 1. *Under Assumptions 1 and 2 and assuming that φ is an F_0-integrable function, we have with probability 1*

$$\widehat{S}_n^{(v)}(\varphi) \longrightarrow S_\infty^{(v)}(\varphi) = \int \mathbb{1}_{\{t<\tau_Z\}} \varphi(s,t,\boldsymbol{\theta}) \; F_0^{(v)}(ds, dt, d\boldsymbol{\theta}), \; v \in \mathcal{V}. \tag{1.4}$$

In addition, if the support of Z is included in that of C, we have obviously $\widehat{S}_n^{(v)}(\varphi) \to S^{(v)}(\varphi)$ w.p.1.

By addition of covariates, this result constitutes an extension of the those demonstrated by Suzukawa (2002, Theorem 1), which are directly based on the proof of Stute and Wang (1993) theorem's. For the sake of completeness, more details about the proof is given in Subsection 5.1.

To obtain weak convergence properties, we adapt the approach followed by Stute (1995) for Kaplan-Meier integrals and Stute (1996) for the version with covariates. We define similar integrability conditions for any function φ F_0-integrable to prove a general convergence result. These conditions are given below.

Assumption 3. $\displaystyle\int \frac{\varphi(S,T,\boldsymbol{\Theta})^2 \delta}{(1-G(T))^2} d\mathbb{P} = \int \frac{\varphi(S,T,\boldsymbol{\Theta})^2}{1-G(T)} d\mathbb{P} < \infty.$

Assumption 4. $\displaystyle\int |\varphi(S,T,\boldsymbol{\Theta})| \sqrt{C_0(T)} \mathbb{1}_{\{T<\tau_Z\}} d\mathbb{P} < \infty.$

Of course, under these assumptions, similar conditions replacing F_0 by $F_0^{(v)}$, for all $v \in \mathcal{V}$, are necessarily satisfied. We consider

$$M(z) = \mathbb{P}(Z \leqslant z), M_0(z) = \mathbb{P}(Z \leqslant z, \delta = 0),$$
$$M^{(v)}(y,z,\boldsymbol{\theta}) = \mathbb{P}(Y \leqslant y, Z \leqslant z, \boldsymbol{\Theta} \leqslant \boldsymbol{\theta}, \delta = 1, V = v),$$

and

$$C_0(x) = \int_0^{x-} \frac{G(dy)}{(1-M(y))(1-G(y))}.$$

We also introduce the functions

$$\lambda_1^{(v)}(x) = \frac{1}{1-M(x)} \int \frac{\varphi(s,t,\boldsymbol{\theta}) \mathbb{1}_{\{x<t<\tau_Z\}}}{(1-G(t))} M^{(v)}(ds, dt, d\boldsymbol{\theta}),$$

and

$$\lambda_2^{(v)}(x) = \int \frac{\lambda_1^{(v)}(t) \mathbb{1}_{\{t<x\}}}{1-M(t)} M_0(dt).$$

Assumption 3 corresponds to a variance type assumption on φ, guaranteeing the existence of a finite second moment. The second assumption is nothing but a technical condition to control the bias of $S_n^{(v)}$. More details are given by Stute (1995). Some additional notations are needed to give our results. Introduce the vectors of size $Card\left(\mathcal{V}\right)$, $\widehat{\boldsymbol{S}}_n\left(\varphi\right) = \left(\widehat{S}_n^{(v)}\left(\varphi\right)\right)_{v\in\mathcal{V}}^{\top}$, $\boldsymbol{S}_\infty\left(\varphi\right) = \left(S_\infty^{(v)}\left(\varphi\right)\right)_{v\in\mathcal{V}}^{\top}$ and $\boldsymbol{S}\left(\varphi\right) = \left(S^{(v)}\left(\varphi\right)\right)_{v\in\mathcal{V}}^{\top}$. The following theorem gives asymptotic properties for $\widehat{\boldsymbol{S}}_n\left(\varphi\right)$.

Theorem 2. *Suppose that the assumptions of Theorem 1 are satisfied. Under Assumptions 3 and 4, we have*

$$\sqrt{n}\left\{\widehat{\boldsymbol{S}}_n\left(\varphi\right) - \boldsymbol{S}_\infty\left(\varphi\right)\right\} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma}\left(\varphi\right)\right), \tag{1.5}$$

where $\boldsymbol{\Sigma}\left(\varphi\right)$ is a symmetric matrix associated to the covariance matrix of the vector $\boldsymbol{a}\left(\varphi\right) = \left(a_v\left(\varphi\right)\right)_{v\in\mathcal{V}}$ where

$$a_v\left(\varphi\right) = \frac{\varphi\left(Y, Z, \boldsymbol{\Theta}\right)\delta J^{(v)}}{1 - G\left(Z\right)} + \lambda_1^{(v)}\left(Z\right)\left(1-\delta\right) - \lambda_2^{(v)}\left(Z\right), \ v \in \mathcal{V}.$$

$\boldsymbol{S}_\infty\left(\varphi\right)$ can be replaced by $\boldsymbol{S}\left(\varphi\right)$ if the support of Z is included in that of C.

The complete proof of this theorem is postponed in Section 5.2. From the Equation (1.5), we could obtain asymptotic confidence intervals if functions $\dfrac{1}{1-G}$, $\lambda_1^{(v)}$ and $\lambda_2^{(v)}$ were known. This can be done by just replacing the distribution functions H, M, M_0 and $M^{(v)}$ in the expression of $\boldsymbol{\Sigma}\left(\varphi\right)$ by their empirical counterparts. However, this calculation may be tedious due to the expression of $\boldsymbol{a}\left(\varphi\right)$. Thus, implementing a non-parametric bootstrap or jackknife procedures is the most appropriate way to obtain asymptotic variance-covariance estimators.

2 Application for transition probabilities estimation

In this subsection, we focus on the asymptotic properties of transition probabilities estimators introduced in Section 3 of the paper. By considering some particular functions φ, Theorems 1 and 2 can be applied to derive asymptotic properties for $p_{a_0 j}\left(s, t, 0, \infty \mid \boldsymbol{\theta}\right)$, $p_{a_0 e}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right)$, $q_{a_0 ed}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right)$, $\bar{p}_{ee}\left(s, t, \Delta_u \mid \boldsymbol{\theta}\right)$ and $p_{ed}\left(s, t, \Delta_u, \infty \mid \boldsymbol{\theta}\right)$[4]. Note that the consistency and weak convergence of $\widehat{\bar{p}}_{a_0 a_0}\left(s, t, 0 \mid \boldsymbol{\theta}\right)$ can be proved easily for all $t \leqslant \tau_Z$ applying the results of Stute (1996) and the delta-method.

Proposition 3. *Under Assumptions 1 and 2, $\widehat{p}_{a_0 j}\left(s, t, 0, \infty \mid \boldsymbol{\theta}\right)$, $\widehat{p}_{a_0 e}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right)$, $\widehat{q}_{a_0 ed}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right)$, $\widehat{\bar{p}}_{ee}\left(s, t, \Delta_u \mid \boldsymbol{\theta}\right)$ and $\widehat{p}_{ed}\left(s, t, \Delta_u, \infty \mid \boldsymbol{\theta}\right)$ converge almost surely to the pertaining transition probabilities of interest, if the support of Z is included in that of C. Under these assumptions, they are also asymptotically normal.*

Proof. To prove these results, transition probabilities are expressed by means of integrals of the form $\int \varphi \ dH_0$ and $\int \varphi \ dF_0^{(v)}$. For brevity's sake, we only report the details of the proof for $p_{a_0 e}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right)$ as the results for other probabilities come using the same reasoning. Hence, its estimator can be rewritten as

$$\widehat{p}_{a_0 e}\left(s, t, 0, \Delta_v \mid \boldsymbol{\theta}\right) = \frac{\widehat{S}_n^{(e, \mathcal{C}(e))}\left(\phi\right)}{\widehat{S}_n\left(\psi\right)},$$

[4] In case where covariates are introduced, a term $\boldsymbol{\theta}$ is added in the transition probabilities expression meaning that we work with the probability measure conditionally on $\{\boldsymbol{\Theta} = \boldsymbol{\theta}\}$.

where $\phi(x,y,z) = \mathbb{1}_{\{s<x\leqslant t<y,\, t-x\in\Delta_v, z=\theta\}}$ and $\psi(x,z) = \mathbb{1}_{\{x>s, z=\theta\}}$. The subset $\mathcal{C}(e)$ is the set of states to which a direct transition from e is possible.

First, the simple function ψ satisfies conditions of Theorem 1.1 in Stute (1996) and therefore $\widehat{S}_n(\psi)$ admits consistent and weak convergence properties as a Kaplan-Meier integral. Second, applying the result of Theorem 1 to the function ϕ which is clearly F_0-integrable, we obtain the consistency results to the limit $S_\infty^{(e,\mathcal{C}(e))}(\phi)$. Since ϕ satisfies Assumptions 3 and 4, the numerator of $\widehat{p}_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ is asymptotically normal. Finally, we have with the delta method

$$\sqrt{n}\left\{\widehat{p}_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta}) - \frac{S_\infty^{(e,\mathcal{C}(e))}(\phi)}{S_\infty(\psi)}\right\} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})\right),$$

where $\sigma_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ is a limit variance function which is complex to calculate[5]. The estimator $\widehat{p}_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ converges to $p_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$, if the support of Z is included is that of C. □

Note that $\widehat{p}_{a_0j}(s,t,0,\infty \mid \boldsymbol{\theta})$ is also consistent if $t \leqslant \tau_Y$. The other estimators are systematically biased if the support of C is strictly included in that of T.

We are now in position to study the alternative estimators' properties by following a similar approach to that of de Uña-Álvarez and Meira-Machado (2015).

Proposition 4. *Under Assumptions 1 and 2, $\widehat{p}^*_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ is consistent w.p.1 if $t \leqslant \tau_Y$, and $\widehat{p}^*_{ed}(s,t,\Delta_u,\infty \mid \boldsymbol{\theta})$ is consistent w.p.1 if $t \leqslant \tau_Z$. In this situation, they are also asymptotically normal.*

Proof. The results are direct applications of Theorems 1, 2 and Stute (1996)'s Theorem for the denominator of $\widehat{p}^*_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$. Let's prove this Proposition for $\widehat{p}^*_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$, as the second estimator's properties are demonstrated similarly. For that, we write

$$\widehat{p}^*_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta}) = \frac{\sum_{i=1}^n W_{in}^{(e)}\phi_1\left(Y_{i:n}, \boldsymbol{\theta}_{[i:n]}\right) - \widehat{S}_n^{(e,\mathcal{C}(e))}(\phi_2)}{\widehat{S}_n(\psi)},$$

where $\phi_1(x,z) = \mathbb{1}_{\{s<x\leqslant t,\, t-x\in\Delta_v, z=\theta\}}$ and $\phi_2(x,y,z) = \mathbb{1}_{\{s<x, y\leqslant t,\, t-x\in\Delta_v, z=\theta\}}$, and then we apply Theorems 1 and 2 on ϕ_1 and ϕ_2. □

Note that this proposition is also verified for $\widehat{p}^*_{ed}(s,t,\Delta_u,\infty \mid \boldsymbol{\theta})$ when $s \leqslant \tau_Y$ and $t \leqslant \tau_Z$. In practice for insurance applications, we always take $s-u+1 \leqslant s$ and $t-v+1 \leqslant t$.

We now consider the second class of alternative estimators involving subsamples that we suppose to be not empty with a positive probability. Introduce ${}_s\tau_Y$ and ${}_s\tau_Z$ the least upper bound of Y and Z pertaining to the subsample $\{i : Y_i > s\}$, and analogously ${}_{s,u}\tau_Y$ and ${}_{s,u}\tau_Z$ for the subsample $\{i : Y_i < s \leqslant Z_i, s - Y_i \in \Delta_u\}$.

Proposition 5. *Under Assumptions 1 and 2, $\breve{p}_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ is consistent w.p.1 if $t \leqslant {}_s\tau_Y$, and $\breve{p}_{ed}(s,t,\Delta_u,\infty \mid \boldsymbol{\theta})$ is consistent w.p.1 if $t \leqslant {}_{s,u}\tau_Z$. In this situation, they are also asymptotically normal.*

Proof. As C is independent of (S,V,T), we find immediately that $({}_sS, {}_sV, {}_sT)$, i.e. the vector of (S,V,T) conditionally on $\{Y > s\}$, is also independent of ${}_sC$, i.e. C conditionally on $\{Y > s\}$. This is also valid when conditioning (S,V,T) on $\{Y < s \leqslant Z, s - Y \in \Delta_u\}$. Thus, we can apply Theorems 1 and 2 on these particular subsamples for $\breve{p}_{a_0e}(s,t,0,\Delta_v \mid \boldsymbol{\theta})$ and $\breve{p}_{ed}(s,t,\Delta_u,\infty \mid \boldsymbol{\theta})$ with the same reasoning that in Proposition 4.

□

[5] By using the function $g(x,y) = \dfrac{x}{y}$, the delta method gives an explicit formula which depends on complex terms. Thus, this is not very useful to write its expression.

3 Estimation with left-truncation and right-censoring

In this section, we additionally allow for left-truncation and discuss the conditions to adapt the asymptotic results of Section 1.2 for competing risks data. It is worth noting that truncation can only occur when the individual is in the healthy state a_0 in our framework. Once we have such results (Theorem 6), the approach followed in Section 2 can be repeated directly without any additional assumption.

For survival data, Sánchez-Sellero et al. (2005) obtain a representation for product-limit integrals for a family of functions $\{\varphi\}$ under censoring and truncation and with independent covariates. This representation implies consistency and weak convergence properties for Kaplan-Meier integrals. Their proofs are demonstrated followed a similar approach than Stute (1995) and Stute (1996) (see Section 5) directly on a family of functions, except that they assume stronger conditions on the moments of φ to avoid imposing hypothesis similar to Assumptions 3 and 4. Although their conditions are more restrictive, they are sufficient in our framework as we work with very simple functions φ. Hence, we simply need to adapt their assumptions to our competing risks data framework.

Assumption 5.

- **i.** the largest lower bound for the support of L is lower than that of T,
- **ii.** (C, L) is independent of (S, T, V) and C is independent of L,
- **iii.** (C, L) is independent of δ (resp. γ) conditionally on (T, V) (resp. (S, V_1)),
- **iv.** $\mathbb{P}(L \leqslant Y) > 0$.

This assumption replaces Assumptions 1 and 2. A less restrictive condition could be considered for assumption i. assuming that φ is nil between the largest lower bound of T and that of L. To be consistent with what we mention above, we do not necessarily assume that the support of T is included in that of C, which is the practical assumption used by Sánchez-Sellero et al. (2005). Similarly to Assumption 2, the independence assumption iii. should only be considered in the situation with covariates, that we introduce in this supplementary materiel for the sake of completeness. Note than Sánchez-Sellero et al. (2005) take a stronger independent condition, which would lead adjusting condition ii. in our case, i.e. (C, L) is independent of $(S, T, V, \boldsymbol{\Theta})$ and C is independent of L.

Some additional notations and conditions are requested for the rest of this section. We note $\Lambda(z) = \mathbb{P}(L \leqslant z \leqslant C)$ and consider a family of functions $\{\varphi\}$ which is a measurable VC-subgraph class of functions admitting an envelope $\boldsymbol{\Phi}$, such that

$$\int \boldsymbol{\Phi}(S, T, \boldsymbol{\Theta}) \Lambda(T)^{-2} (1 - F(T))^{-5} < \infty, \int \boldsymbol{\Phi}(S, T, \boldsymbol{\Theta})^2 \Lambda(T)^{-2} (1 - F(T))^{-3} d\mathbb{P} < \infty.$$

We also assume that $\int \Lambda(T)^{-2} d\mathbb{P} < \infty$. For such a function φ, we re-write our Aalen-Johansen integral estimator (1.3) as

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \varphi(Y_i, Z_i, \boldsymbol{\Theta}_i),$$

with the weights

$$\widetilde{W}_{in}^{(v)} = \frac{\delta_i J_i^{(v)}}{n\widetilde{C}_n(Z_i)} \prod_{\{j: Z_j < Z_i\}} \left(1 - \frac{1}{n\widetilde{C}_n(Z_i)}\right)^{\delta_j},$$

where $C_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{L_i \leqslant x \leqslant Y_i\}}$.

Theorem 6. *Under Assumption 5, $\widehat{S}_n^{(v)}(\varphi)$ converges almost surely and weakly to $\widehat{S}_\infty^{(v)}(\varphi)$, for $v \in \mathcal{V}$. $\boldsymbol{S}_\infty(\varphi)$ can be replaced by $\boldsymbol{S}(\varphi)$ if the support of Z is included in that of C.*

Proof. The proof relies on similar arguments as those of Theorem 1 of Sánchez-Sellero et al. (2005). Formally, we can write $\widehat{S}_n^{(v)}(\varphi)$ as a sum of a dominant part and a negligible term, i.e.

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^{n} \varphi(Y_i, Z_i, \boldsymbol{\Theta}_i) \frac{\delta_i J_i^{(v)}}{n\widetilde{C}_n(Z_i)} \prod_{\{j:Z_j<Z_i\}} \left(1 - \frac{1}{n\widetilde{C}_n(Z_j)+1}\right) + O\Big(n^{-1}(\ln n)^3\Big).$$

Similarly to the proofs of Theorems 1 and 2, Y plays the role of a covariate here and the term $J_i^{(v)}$, $v \in \mathcal{V}$, has no particular effect when expressing the dominant part of $S_n^{(v)}(\varphi)$. Hence, following the same steps as Sánchez-Sellero et al. (2005), we can decompose $S_n^{(v)}(\varphi)$ as follows

$$\widehat{S}_n^{(v)}(\varphi) - S_\infty^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \boldsymbol{\Theta}_i) J_i^{(v)} \delta_i \gamma_0(Z_i) - \gamma_1^{(v)}(Z_i)\delta_i + \gamma_2^{(v)}(L_i, Z_i) - \gamma_3^{(v)}(L_i, Z_i) + R_n^{(v)},$$

where $|R_n^{(v)}| = O\Big(n^{-1}(\ln n)^3\Big)$ w.p.1. For this formulation, we consider

$$\gamma_0(x) = \frac{1}{N(x)} \exp\left\{-\int_{-\infty}^{x} \frac{1}{N(t)} N_1(dt)\right\},$$

$$\gamma_1^{(v)}(x) = \frac{1}{N(x)} \int \varphi(s,t,\theta)\, \mathbb{1}_{\{x<t<\tau_Z\}} \gamma_0(t)\, N^{(v)}(ds, dt, d\boldsymbol{\theta}),$$

$$\gamma_2^{(v)}(x,y) = \int \frac{\gamma_1^{(v)}(t)}{N(t)} \mathbb{1}_{\{x<t<y\}} N_1(dt),$$

$$\gamma_3^{(v)}(x,y) = \int \frac{\varphi(s,t,\theta)\,\gamma_0(t)}{N(t)} \mathbb{1}_{\{x<\tau<y\}} N^{(v)}(ds, dt, d\boldsymbol{\theta}),$$

where

$$N(z) = \mathbb{P}(L \leqslant z \leqslant Z \mid L \leqslant Z),\ N_1(z) = \mathbb{P}(Z \leqslant z, \delta = 1 \mid L \leqslant Z),$$
$$N^{(v)}(y, z, \boldsymbol{\theta}) = \mathbb{P}(Y \leqslant y, Z \leqslant z, \boldsymbol{\Theta} \leqslant \boldsymbol{\theta}, \delta = 1, V = v \mid L \leqslant Z).$$

Finally, the convergence results are easily obtained with this formulation. □

Remark that similar asymptotic results can be obtained for $\widehat{S}_n(\varphi)$ following the same methodology.

As functions φ that we use to compute transition probabilities in Section 2 are very simple, we can easily show that each of these functions is in a VC-subgraph of functions and exhibits an envelop which satisfies the moment conditions introduced above. Applying Theorem 6 in each case, it is shown that Propositions 3-5 are also verified when adding independent left-truncation under Assumption 5.

4 Additional simulation results

This Section presents in Tables 1 and 2 the mean bias, variance and the mean square error related to each of the estimators for $p_{a_0e_2}(s, s+4, 0, \Delta_v)$ and $p_{e_2d}(s, s+4, \Delta_u, \infty)$. The same calculation methodology as explained in Section 4 of the paper is used.

Table 1: Performance analysis for estimated transition probabilities from state a_0 to state e_2.

(s,t,Δ_v)	n	Censoring	$p_{a_0e_2}(s,t,0,\Delta_v)$	$\widehat{p}_{a_0e_2}(s,t,0,\Delta_v)$ BIAS	VAR	MSE	$\widehat{p}^{*}_{a_0e_2}(s,t,0,\Delta_v)$ BIAS	VAR	MSE	$\breve{p}_{a_0e_2}(s,t,0,\Delta_v)$ BIAS	VAR	MSE
(28.78,32.78,]0,2])	100	Scenario 1	0.111	0.83	3.01	3.01	0.68	2.02	2.02	0.69	2.02	2.02
	100	Scenario 2	0.111	2.88	2.42	2.42	2.92	2.33	2.34	2.92	2.33	2.34
	200	Scenario 1	0.111	2.95	1.58	1.59	3.67	1.02	1.04	3.67	1.02	1.04
	200	Scenario 2	0.111	3.36	1.22	1.23	3.10	1.18	1.19	3.11	1.18	1.19
	400	Scenario 1	0.111	0.79	0.82	0.82	1.12	0.54	0.54	1.12	0.54	0.54
	400	Scenario 2	0.111	1.98	0.61	0.62	1.81	0.58	0.58	1.81	0.58	0.58
(28.78,32.78,]2,4])	100	Scenario 1	0.794	2.42	1.90	1.90	1.54	1.35	1.35	1.56	1.35	1.35
	100	Scenario 2	0.794	1.11	1.92	1.92	0.83	1.81	1.81	0.81	1.81	1.81
	200	Scenario 1	0.794	1.59	0.90	0.90	1.24	0.70	0.70	1.24	0.69	0.70
	200	Scenario 2	0.794	-0.15	0.96	0.96	-0.25	0.92	0.92	-0.24	0.92	0.92
	400	Scenario 1	0.794	1.00	0.47	0.47	0.95	0.36	0.36	0.96	0.36	0.36
	400	Scenario 2	0.794	0.91	0.45	0.46	0.88	0.44	0.44	0.88	0.44	0.44
(32.35,36.35,]0,2])	100	Scenario 1	0.139	0.99	13.46	13.45	-3.19	6.48	6.48	-3.21	6.48	6.48
	100	Scenario 2	0.139	0.15	5.62	5.62	-0.20	5.45	5.45	-0.21	5.45	5.44
	200	Scenario 1	0.139	0.05	6.69	6.68	-0.71	2.98	2.98	-0.68	2.98	2.98
	200	Scenario 2	0.139	-5.23	2.59	2.61	-5.32	2.53	2.55	-5.32	2.53	2.55
	400	Scenario 1	0.139	-1.28	3.25	3.25	-1.84	1.46	1.46	-1.85	1.46	1.46
	400	Scenario 2	0.139	-1.21	1.35	1.35	-1.27	1.27	1.27	-1.27	1.27	1.27
(32.35,36.35,]2,4])	100	Scenario 1	0.142	4.83	9.24	9.25	1.95	5.49	5.49	1.83	5.49	5.48
	100	Scenario 2	0.142	2.95	5.92	5.92	3.75	5.70	5.70	3.70	5.70	5.70
	200	Scenario 1	0.142	0.34	4.43	4.43	-0.41	2.53	2.53	-0.32	2.55	2.54
	200	Scenario 2	0.142	1.95	2.82	2.82	2.22	2.69	2.70	2.18	2.69	2.69
	400	Scenario 1	0.142	3.38	2.17	2.18	2.34	1.35	1.35	2.42	1.35	1.35
	400	Scenario 2	0.142	1.72	1.41	1.41	1.74	1.34	1.34	1.72	1.34	1.34
(35.49,39.49,]0,2])	100	Scenario 1	0.143	55.69	40.37	43.43	0.28	20.98	20.96	7.85	20.06	20.10
	100	Scenario 2	0.143	-1.21	16.22	16.21	-0.24	15.19	15.18	-0.25	15.20	15.18
	200	Scenario 1	0.143	49.10	29.26	31.64	-0.87	10.54	10.53	-0.60	10.54	10.53
	200	Scenario 2	0.143	1.70	7.86	7.85	2.36	7.46	7.46	2.36	7.46	7.46
	400	Scenario 1	0.143	39.83	18.35	19.92	1.91	5.37	5.37	1.96	5.37	5.37
	400	Scenario 2	0.143	-2.11	3.62	3.63	-1.89	3.48	3.48	-1.90	3.48	3.48
(35.49,39.49,]2,4])	100	Scenario 1	0.217	24.68	59.05	59.60	-8.63	28.44	28.48	-1.72	28.67	28.64
	100	Scenario 2	0.217	1.53	20.94	20.92	2.56	19.40	19.38	2.58	19.46	19.45
	200	Scenario 1	0.217	8.51	35.11	35.14	-1.44	12.89	12.88	-1.39	12.94	12.93
	200	Scenario 2	0.217	-2.64	10.29	10.28	-0.23	9.63	9.62	-0.36	9.64	9.64
	400	Scenario 1	0.217	5.50	21.13	21.14	0.06	6.47	6.46	0.27	6.46	6.45
	400	Scenario 2	0.217	-0.10	5.14	5.13	0.77	4.72	4.71	0.76	4.72	4.72

Note: This table contains the estimates bias (BIAS) $\times 10^3$, variance (VAR) $\times 10^3$ and mean square error (MSE) $\times 10^3$ with our non-parametric estimators. We compare the results at time $s = \tau_{.20}$, $s = \tau_{.40}$ and $s = \tau_{.60}$ for samples with size $n = 100$, $n = 200$ and $n = 400$. The sojourn time is comprised in $]0,2]$ and in $]2,4]$. The results are obtained with $K = 1,000$ Monte Carlo simulations.

Table 2: Performance analysis for estimated transition probabilities from state e_2 to state d.

				$\widehat{p}_{e_2 d}(s,t,0,\Delta_v)$			$\widehat{p}^{*}_{e_2 d}(s,t,0,\Delta_u)$			$\breve{p}_{e_2 d}(s,t,0,\Delta_v)$		
(s,t,Δ_v)	n	Censoring	$p_{e_2 d}(s,t,0,\Delta_v)$	BIAS	VAR	MSE	BIAS	VAR	MSE	BIAS	VAR	MSE
(28.78,32.78,]0,2])	100	Scenario 1	0.718	-28.83	61.98	62.75	-3.69	60.52	60.47	-11.06	62.72	62.78
	100	Scenario 2	0.718	0.31	68.78	68.71	12.61	68.63	68.72	-1.50	69.79	69.72
	200	Scenario 1	0.718	-12.71	30.00	30.13	-3.28	28.01	28.00	-7.95	26.90	26.93
	200	Scenario 2	0.718	-2.51	36.67	36.64	0.76	36.43	36.39	-1.91	36.51	36.48
	400	Scenario 1	0.718	-8.95	13.72	13.78	-10.36	14.27	14.36	-8.13	12.26	12.31
	400	Scenario 2	0.718	-10.54	15.18	15.28	-9.62	15.75	15.83	-10.72	15.08	15.18
(28.78,32.78,]2,4])	100	Scenario 1	0.958	1.72	23.65	23.63	81.20	72.10	78.63	22.84	34.58	35.06
	100	Scenario 2	0.958	-2.75	24.48	24.46	95.58	93.15	102.19	-2.47	25.59	25.57
	200	Scenario 1	0.958	-5.29	10.13	10.15	28.10	20.94	21.71	9.30	16.04	16.11
	200	Scenario 2	0.958	-0.17	13.30	13.28	50.51	31.72	34.24	1.27	13.62	13.61
	400	Scenario 1	0.958	-6.71	4.76	4.80	15.09	6.60	6.82	-2.02	5.77	5.77
	400	Scenario 2	0.958	-0.04	6.25	6.25	25.16	9.84	10.46	0.91	6.31	6.31
(32.35,36.35,]0,2])	100	Scenario 1	0.697	-15.10	73.64	73.79	33.61	75.03	76.09	17.44	73.52	73.75
	100	Scenario 2	0.697	-4.84	64.69	64.65	7.02	65.49	65.47	-4.37	65.80	65.75
	200	Scenario 1	0.697	-8.89	36.13	36.17	5.56	39.09	39.08	1.62	32.22	32.19
	200	Scenario 2	0.697	-9.72	27.94	28.00	-7.57	29.19	29.22	-7.35	28.09	28.11
	400	Scenario 1	0.697	-2.52	16.25	16.24	1.33	20.55	20.53	2.29	14.88	14.87
	400	Scenario 2	0.697	-1.34	12.25	12.24	-0.63	13.53	13.52	-0.74	12.28	12.27
(32.35,36.35,]2,4])	100	Scenario 1	0.957	-0.05	19.37	19.35	133.00	63.43	81.05	50.14	43.53	46.00
	100	Scenario 2	0.957	-9.40	11.50	11.58	58.13	47.71	51.04	-6.19	13.01	13.03
	200	Scenario 1	0.957	-3.83	8.80	8.80	62.93	20.73	24.67	9.21	10.82	10.90
	200	Scenario 2	0.957	1.28	8.09	8.09	31.81	13.21	14.21	3.46	8.62	8.63
	400	Scenario 1	0.957	1.13	4.73	4.73	45.33	12.35	14.39	4.99	4.76	4.78
	400	Scenario 2	0.957	-1.24	3.60	3.59	13.39	4.45	4.63	-0.78	3.60	3.59
(35.49,39.49,]0,2])	100	Scenario 1	0.653	-51.08	123.00	125.49	76.80	127.19	132.96	43.41	133.41	135.16
	100	Scenario 2	0.653	8.08	83.19	83.17	21.92	82.52	82.92	7.24	84.47	84.44
	200	Scenario 1	0.653	-33.34	76.87	77.90	28.55	69.55	70.29	14.12	67.77	67.91
	200	Scenario 2	0.653	15.09	38.55	38.74	19.17	38.81	39.14	15.29	38.45	38.64
	400	Scenario 1	0.653	-5.85	36.59	36.59	15.67	40.25	40.45	15.92	29.41	29.63
	400	Scenario 2	0.653	10.86	17.73	17.83	13.14	18.04	18.19	11.83	17.39	17.51
(35.49,39.49,]2,4])	100	Scenario 1	0.944	-4.50	27.92	27.91	192.76	106.07	143.12	84.08	77.95	84.94
	100	Scenario 2	0.944	-3.26	20.58	20.57	54.37	48.09	51.00	-1.18	21.50	21.48
	200	Scenario 1	0.944	-3.60	17.79	17.79	99.00	44.12	53.88	29.33	26.37	27.20
	200	Scenario 2	0.944	-0.27	10.21	10.20	23.02	13.38	13.90	2.13	11.29	11.28
	400	Scenario 1	0.944	-3.56	8.52	8.53	54.66	18.19	21.16	7.59	8.54	8.58
	400	Scenario 2	0.944	-1.69	4.18	4.18	11.02	5.28	5.39	-0.93	4.17	4.16

Note: This table contains the estimates bias (BIAS) $\times 10^3$, variance (VAR) $\times 10^3$ and mean square error (MSE) $\times 10^3$ with our non-parametric estimators. We compare the results at time $s = \tau_{.20}$, $s = \tau_{.40}$ and $s = \tau_{.60}$ for samples with size $n = 100$, $n = 200$ and $n = 400$. The sojourn time is comprised in $]0,2]$ and in $]2,4]$. The results are obtained with $K = 1,000$ Monte Carlo simulations.

5 Proofs

5.1 Proof of Theorem 1

Let, for $i = 1, \dots, n$ and $v \in \mathcal{V}$, $D_i^{(v)} = \left(Y_i, \delta_i, J_i^{(v)}, \boldsymbol{\Theta}_i\right)$ and for each $n \geqslant 0$, the σ-algebra

$$\mathcal{F}_n^{(v)} = \sigma\left(Z_{i:n}, D_{[i:n]}^{(v)}, 1 \leqslant i \leqslant n, Z_{n+1}, D_{n+1}^{(v)}, \dots\right),$$

where $D_{[i:n]}^{(v)}$ are the value paired with $Z_{i:n}$.

Clearly for $v \in \mathcal{V}$, $\widehat{S}_n^{(v)}(\varphi)$ is adapted to $\mathcal{F}_n^{(v)}$ and $\mathcal{F}_n^{(v)}$ is decreasing and converges towards $\mathcal{F}_\infty^{(v)} = \bigcap_{n \geqslant 1} \mathcal{F}_n^{(v)}$. Our strategy, following Stute and Wang (1993), is to demonstrate that $\left(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geqslant 0\right)$ is a reverse-time supermartingale and then apply convergence result given by Neveu (1975, Proposition V-3-11, p. 116) to obtain consistency. For the following lemma, we consider that φ is a nonnegative fonction. Otherwise, the results remain applicable by decomposing φ into positive and negative parts.

Lemma 1. *For $\varphi \geqslant 0$ and assuming that the distribution function of Z is continuous, $\left(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geqslant 0\right)$ is a reverse-time supermartingale for $v \in \mathcal{V}$.*

Proof. Denote by $\widehat{F}_n^{(v)}(z) = \sum_{i=1}^n \widetilde{W}_{in}^{(v)} \mathbb{1}_{\{Z_{i:n} \leqslant z\}}$ and let $\widehat{F}_n^{(v)}\{z\} = \widehat{F}_n^{(v)}(z) - \widehat{F}_n^{(v)}(z-)$, we can remark that

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^n \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{\Theta}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\}.$$

If Z_{n+1} has rank k with $1 \leqslant k \leqslant n+1$, then $Z_{i:n} = Z_{i:n+1}$ for all $i < k$. Thus, we have

$$\sum_{i=1}^{k-1} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{\Theta}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\} = \sum_{i=1}^{k-1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{\Theta}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\},$$

$$\sum_{i=k}^{n} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{\Theta}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\} = \sum_{i=k+1}^{n+1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{\Theta}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\},$$

and

$$\varphi\left(Y_{[k:n+1]}, Z_{k:n+1}, \boldsymbol{\Theta}_{[k:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{k:n+1}\} = 0.$$

Hence, we obtain that

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^{n+1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{\Theta}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\}. \tag{5.1}$$

Following the same lines of the proof of Lemma 2.2 in Stute and Wang (1993) (see also Stute (1993, Lemma 2.2)), we show with Lemma 2.1 of Stute and Wang (1993) applied to $D_{[i:n]}$ that

$$\mathbb{E}\left[\widehat{F}_n^{(v)}\{Z_{i:n+1}\} \mid \mathcal{F}_{n+1}^{(v)}\right] = \widetilde{W}_{i,n+1}^{(v)}, \ 1 \leqslant i \leqslant n$$

and

$$\mathbb{E}\left[\widehat{F}_n^{(v)}\{Z_{n+1:n+1}\} \mid \mathcal{F}_{n+1}^{(v)}\right] \leqslant \widetilde{W}_{n+1,n+1}^{(v)}.$$

Since $\varphi \geqslant 0$, the result follows immediately by writing the conditionnal expectation of (5.1). □

From Lemma 1, we have by applying the Proposition V-3-11 of Neveu (1975) that $\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi) \mid \mathcal{F}_\infty^{(v)}\right]$ admits limit $\mathbb{P}$-almost surely. Due to the Hewitt-Savage zero-one law, $\mathcal{F}_\infty^{(v)}$ is trivial and then

$$\lim_{n\to\infty} \mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi) \mid \mathcal{F}_\infty^{(v)}\right] = \lim_{n\to\infty} \mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] = S_\infty^{(v)}(\varphi).$$

Now, we aim to determine the value of $S_\infty^{(v)}(\varphi)$. To do this, we write

$$m(z) = \mathbb{P}(\delta = 1 \mid Z = z),$$

$$\Psi_n(z) = \prod_{i=1}^{n}\left(1 + \frac{1 - m(Z_{i:n})}{n-i+1}\right)^{\mathbb{1}\{Z_{i:n} < z\}},$$

and for $v \in \mathcal{V}$

$$\widetilde{\varphi}^{(v)}(z) = \mathbb{E}\left[\varphi(Y, Z, \mathbf{\Theta})\, \delta J^{(v)} \mid Z = z\right].$$

Lemma 2. *Under the assumptions of Lemma 1, we have for $v \in \mathcal{V}$*

$$\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] = \mathbb{E}\left[\widetilde{\varphi}^{(v)}(Z)\, \mathbb{E}\left[\Psi_{n-1}(Z)\right]\right].$$

Proof. Let R_{jn} denote the rank of Z_j among $Z_1, \ldots, Z_n$, we can write

$$\begin{aligned}
\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \varphi\left(Y_{[i:n]}, Z_{i:n}, \mathbf{\Theta}_{[i:n]}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{n-i+1} \mathbb{E}\left[\begin{array}{l} \varphi\left(Y_{[i:n]}, Z_{i:n}, \mathbf{\Theta}_{[i:n]}\right) \delta_{[i:n]} J_{[i:n]}^{(v)} \\ \times \displaystyle\prod_{j=1}^{i-1}\left(\frac{n-j}{n-j+1}\right)^{\delta_{[j:n]}} \mid Z_{1:n}, \ldots, Z_{n:n} \end{array}\right]\right].
\end{aligned}$$

From Lemma 2.1 of Stute and Wang (1993) applied to $D_i^{(v)}$ for $i = 1, \ldots, n$, we know that, conditionally on $Z_{1:n} < \ldots < Z_{n:n}$, the concomitants among the D's are independent. Hence,

$$\begin{aligned}
\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n-i+1} \prod_{j=1}^{i-1} \mathbb{E}\left[\left(\frac{n-j}{n-j+1}\right)^{\delta_{[j:n]}} \mid Z_{j:n}\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n-i+1} \prod_{j=1}^{i-1}\left(1 - \frac{m(Z_{j:n})}{n-j+1}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n} \prod_{j=1}^{i-1}\left(1 + \frac{1 - m(Z_{j:n})}{n-j}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\widetilde{\varphi}^{(v)}(Z_{i})}{n} \prod_{j=1}^{n}\left(1 + \frac{1 - m(Z_{j})}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j < Z_i\}}\right] \\
&= \mathbb{E}\left[\widetilde{\varphi}^{(v)}(Z_1) \prod_{j=1}^{n}\left(1 + \frac{1 - m(Z_{j})}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j < Z_1\}}\right].
\end{aligned} \tag{5.2}$$

If $Z_j < Z_1$ then $R_{jn} = R_{j,n-1}$. Conditioning on Z_1, the result follows easily. $\square$

A similar proof is established in Stute and Wang (1993)[Lemma 2.4] and reused in Stute (1994). Now, we are in position to prove Theorem 1 by studying the process $\Psi_n(z)$.

Proof of Theorem 1. From Stute and Wang (1993)[Lemma 2.5 and Lemma 2.6] and assuming that G and the distribution function of F are continuous, for each $z < \tau_Z$, we have

$$\mathbb{E}\left[\Psi_n(z)\right] \uparrow \frac{1}{1 - G(z)}. \tag{5.3}$$

Hence, under the Assumption 2 and $\varphi \geqslant 0$, we obtain by applying Lemma 2, Equation (5.3) and the monotone convergence theorem that

$$\begin{aligned}
S_\infty^{(v)}(\varphi) &= \int \mathbb{1}_{\{Z<\tau_Z\}} \frac{\widetilde{\varphi}^{(v)}(Z)}{1-G(Z)}\, d\mathbb{P} \\
&= \int \mathbb{1}_{\{Z<\tau_Z\}} \mathbb{E}\left[\varphi(Y, Z, \boldsymbol{\Theta})\, \delta J^{(v)} \mid Z\right] \frac{1}{1-G(Z)}\, d\mathbb{P} \\
&= \int \varphi(S, T, \boldsymbol{\Theta}) \frac{\mathbb{1}_{\{T<\tau_Z\}} \delta J^{(v)}}{1-G(T)}\, d\mathbb{P} \\
&= \int \varphi(S, T, \boldsymbol{\Theta}) \frac{\mathbb{1}_{\{T<\tau_Z\}} J^{(v)}}{1-G(T)} \mathbb{P}(T \leqslant C \mid S, V, T, \boldsymbol{\Theta})\, d\mathbb{P} \\
&= \int \varphi(S, T, \boldsymbol{\Theta}) \frac{\mathbb{1}_{\{T<\tau_Z\}} J^{(v)}}{1-G(T)} \mathbb{P}(T \leqslant C \mid V, T)\, d\mathbb{P}.
\end{aligned}$$

Since C and (V, T) are independent (see Assumption 1), we remark that $\mathbb{P}(T \leqslant C \mid V, T) = 1 - G(T)$. Hence, we obtain

$$S_\infty^{(v)}(\varphi) = \int \mathbb{1}_{\{t<\tau_Z\}} \varphi(s, t, \boldsymbol{\theta})\; F_0^{(v)}(ds, dt, d\boldsymbol{\theta}). \tag{5.4}$$

As indicated earlier for a continuous $F^{(v)}$, the desired proof follows from Lemma 1, Equation (5.4) and proposition V-3-11 of Neveu (1975). □

5.2 Proof of Theorem 2

Here, we denote

$$\begin{aligned}
\widehat{M}_n(z) &= \sum_{i=1}^n \mathbb{1}_{\{Z_i \leqslant z\}}, \\
\widehat{M}_{0n}(z) &= \sum_{i=1}^n \mathbb{1}_{\{Z_i \leqslant z, \delta_i = 0\}}, \\
\widehat{M}_n^{(v)}(y, z, \boldsymbol{\theta}) &= \sum_{i=1}^n \mathbb{1}_{\{Y_i \leqslant y, Z_i \leqslant z, \boldsymbol{\Theta}_i \leqslant \boldsymbol{\theta}, \delta_i = 1, V_i = v\}},
\end{aligned}$$

the empirical distribution functions of M, M_0 and $M_0^{(v)}$. Directly based on Stute (1995)'s proof, our strategy is in 2 steps: prove CLT when φ vanishes to the right of some $\nu < \tau_Z$ and then extend it on $[0, \tau_Z]$. Note that Suzukawa (2002) also follows the same strategy.

Lemma 3. *We have for $v \in \mathcal{V}$*

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \mathbf{\Theta}_i)\,\delta_i J_i^{(v)} \exp\left\{ n\int_0^{Z_i-} \ln\left\{1 + \frac{1}{n\left(1-\widehat{M}_n(\tau)\right)}\right\}\widehat{M}_{0n}(d\tau)\right\} \tag{5.5}$$

Proof. From the same rationale used to obtain (5.2), we find

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \mathbf{\Theta}_i)\,\delta_i J_i^{(v)} \prod_{j=1}^{n}\left(1+\frac{1-\delta_j}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j<Z_i\}}. \tag{5.6}$$

The result follows immediately by definition of $\widehat{M}_n(z)$ and $\widehat{M}_{0n}(z)$, see proof of Lemma 2.1 in Stute (1995). □

The exponential term in (5.5) is expanded in Stute (1995) as follows

$$\exp\{\ldots\} = \frac{1}{1-G(Z_i)}(1+B_{in}+C_{in}) + \frac{1}{2}\exp\{\Delta_i\}(B_{in}+C_{in})^2, \tag{5.7}$$

where

$$B_{in} = n\int_0^{Z_i-} \ln\left\{1+\frac{1}{n\left(1-\widehat{M}_n(\tau)\right)}\right\}\widehat{M}_{0n}(d\tau) - \int_0^{Z_i-}\frac{\widehat{M}_{0n}(d\tau)}{1-\widehat{M}_n(\tau)},$$

$$C_{in} = \int_0^{Z_i-}\frac{\widehat{M}_{0n}(d\tau)}{1-\widehat{M}_n(\tau)} - \int_0^{Z_i-}\frac{M_0(d\tau)}{1-M(\tau)},$$

and Δ is between the two terms

$$n\int_0^{Z_i-} \ln\left\{1+\frac{1}{n\left(1-\widehat{M}_n(\tau)\right)}\right\}\widehat{M}_{0n}(d\tau) \quad\text{and}\quad \int_0^{Z_i-}\frac{M_0(d\tau)}{1-M(\tau)}.$$

Considering (5.5) and (5.7), we write

$$\begin{aligned}\widehat{S}_n^{(v)}(\varphi) &= \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \mathbf{\Theta}_i)\,\delta_i J_i^{(v)} \frac{1+B_{in}+C_{in}}{1-G(Z_i)} \\ &+ \frac{1}{2n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \mathbf{\Theta}_i)\,\delta_i J_i^{(v)} \exp\{\Delta_i\}(B_{in}+C_{in})^2.\end{aligned} \tag{5.8}$$

Now, we decompose the last equation and study approximations for each component. To do this, we make for φ the following assumption

Assumption 6. φ is an F_0-integrable function such that $\int \varphi^2 dF_0 < \infty$ and $\varphi(s,t,\boldsymbol{\theta}) = 0$ for $\nu < t$ where $\nu < \tau_Z$.

This assumption aims to bound the denominators of the terms obtained in the following lemmas.

Lemma 4. *Under Assumption 6, we have*

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, Z_i, \mathbf{\Theta}_i)\,\delta_i J_i^{(v)} \frac{C_{in}}{1-G(Z_i)} \\ &= -\iiint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{(1-G(t))(1-M(\tau))^2}\widehat{M}_n(d\omega)\,M_0(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) \\ &+ \iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))(1-M(\tau))}\widehat{M}_{0n}(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) + R_{n1}^{(v)},\end{aligned} \tag{5.9}$$

where $|R_{n1}^{(v)}| = O\left(n^{-1}\ln n\right)$ *w.p.1.*

Proof. Using the following decomposition for $z < Z_{n:n}$ in C_{in},

$$\frac{1}{1-\widehat{M}_n(z)} = -\frac{1-\widehat{M}_n(z)}{(1-M(z))^2} + \frac{2}{1-M(z)} + \frac{\left(\widehat{M}_n(z)-M(z)\right)^2}{(1-M(z))^2\left(1-\widehat{M}_n(z)\right)},$$

we can write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\varphi(Y_i,Z_i,\mathbf{\Theta}_i)\,\delta_i J_i^{(v)}\frac{C_{in}}{1-G(Z_i)} \\
&= -\iiint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{(1-G(t))(1-M(\tau))^2}\widehat{M}_n(d\omega)\,\widehat{M}_{0n}(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}) \\
&+2\iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))(1-M(\tau))}\widehat{M}_{0n}(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}) \\
&-\iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))(1-M(\tau))}M_0(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}) + R_{n2}^{(v)},
\end{aligned} \tag{5.10}$$

where

$$R_{n2}^{(v)} = \iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))}\frac{\left(\widehat{M}_n(t)-M(t)\right)^2}{(1-M(t))^2\left(1-\widehat{M}_n(t)\right)}\widehat{M}_{0n}(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}).$$

Under Assumption 6 and with same argument as Stute (1995, Lemma 2.5), i.e. using iterated logarithm for empirical measures and strong law of large numbers (SLLN), we obtain $|R_{n2}^{(v)}| = O\big(n^{-1}\ln n\big)$ w.p.1. For the rest of the proof, we shall decompose the other terms in the previous equation (5.10) as a U-statistic plus a negligible remainder. Formally, we have

$$\begin{aligned}
&\iiint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{(1-G(t))(1-M(\tau))^2}\widehat{M}_n(d\omega)\,\widehat{M}_{0n}(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}) \\
&= \iiint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{(1-G(t))(1-M(\tau))^2} \\
&\quad\times\left[\begin{array}{l}\widehat{M}_n(d\omega)\,M_0(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) + M(d\omega)\,\widehat{M}_{0n}(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) \\ -2M(d\omega)\,M_0(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) + M(d\omega)\,M_0(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta})\end{array}\right] \\
&+ R_{n3}^{(v)},
\end{aligned} \tag{5.11}$$

and

$$\begin{aligned}
&\iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))(1-M(\tau))}\widehat{M}_{0n}(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta}) \\
&= \iint \frac{\varphi(s,t,\boldsymbol{\theta})\,\mathbb{1}_{\{\tau<t\}}}{(1-G(t))(1-M(\tau))} \\
&\quad\times\left[\widehat{M}_{0n}(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) - M_0(d\tau)\,M^{(v)}(ds,dt,d\boldsymbol{\theta}) + M_0(d\tau)\,\widehat{M}_n^{(v)}(ds,dt,d\boldsymbol{\theta})\right] \\
&+ R_{n4}^{(v)},
\end{aligned} \tag{5.12}$$

where $|R_{n3}^{(v)}| = O\big(n^{-1}\ln n\big)$ and $|R_{n4}^{(v)}| = O\big(n^{-1}\ln n\big)$ w.p.1. We refer to similar arguments as for Lemmas 2.3 and 2.4 of Stute (1995) to obtain the two representations based on the Hajek projection of a V-statistic of the multivariate data $\left(Y_i, Z_i, \mathbf{\Theta}_i, \delta_i, J_i^{(v)}\right)$, $1 \leqslant i \leqslant n$. Finally, the proof of (5.9) follows by substituting (5.11) and (5.12) into (5.10). □

Now, we study the other terms in (5.8) in the following Lemma.

Lemma 5. *Under Assumption 6, we have with w.p.1*

$$\frac{1}{n}\left|\sum_{i=1}^{n} \varphi\left(Y_i, Z_i, \mathbf{\Theta}_i\right) \delta_i J_i^{(v)} \frac{B_{in}}{1-G\left(Z_i\right)}\right| = O\big(n^{-1}\big), \tag{5.13}$$

and

$$\frac{1}{2n}\left|\sum_{i=1}^{n} \varphi\left(Y_i, Z_i, \mathbf{\Theta}_i\right) \delta_i J_i^{(v)} \exp\left\{\Delta_i\right\}\left(B_{in}+C_{in}\right)^2\right| = O\big(n^{-1}\ln n\big). \tag{5.14}$$

Proof. The proof follows immediately from the proofs of Lemmas 2.6 and 2.7 of Stute (1995). □

Proof of Theorem 2. With Lemmas 4 and 5, Equation (5.8) yields

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \frac{\varphi\left(Y_i, Z_i, \mathbf{\Theta}_i\right) \delta_i J_i^{(v)}}{1-G\left(Z_i\right)} + \frac{1}{n}\sum_{i=1}^{n}\left[\lambda_1^{(v)}\left(Z_i\right)\left(1-\delta_i\right) - \lambda_2^{(v)}\left(Z_i\right)\right] + R_{n5}^{(v)} \tag{5.15}$$

where $|R_{n5}^{(v)}| = O\big(n^{-1}\ln n\big)$ w.p.1. As a consequence, we have CLT results for $\widehat{S}_n^{(v)}(\varphi)$, $v \in \mathcal{V}$, and Theorem 2 follows under Assumption 6.

Finally, the results of Theorem 2 can be extended on $]\nu, \tau_Z]$ by an argument similar to that of the proof of Theorem 1.1 in Stute (1995) under Assumptions 3 and 4. Thus, we have

$$\widehat{S}_n^{(v)}(\varphi) - S_\infty^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \frac{\varphi\left(Y_i, Z_i, \mathbf{\Theta}_i\right) \delta_i J_i^{(v)}}{1-G\left(Z_i\right)} + \frac{1}{n}\sum_{i=1}^{n}\left[\lambda_1^{(v)}\left(Z_i\right)\left(1-\delta_i\right) - \lambda_2^{(v)}\left(Z_i\right)\right] + R_{n5}^{(v)}.$$

□

References

De Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics*. DOI: 10.1111/biom.12288.

Neveu, J. (1975). Discrete Parameter Martingales. Amsterdam and New-York: North-Holland / American Elsevier.

Sánchez-Sellero, C., Manteiga, W. G., and Van Keilegom, I. (2005). Uniform Representation of Product-Limit Integrals with Applications. *Scandinavian Journal of Statistics* 32.4, pp. 563–581. DOI: 10.1111/j.1467-9469.2005.00453.x.

Stute, W. and Wang, J.-L. (1993). The Strong Law under Random Censorship. *The Annals of Statistics* 21.3, pp. 1591–1607. DOI: 10.1214/aos/1176349273.

Stute, W. (1993). Consistent Estimation Under Random Censorship When Covariables Are Present. *Journal of Multivariate Analysis* 45.1, pp. 89–103. DOI: 10.1006/jmva.1993.1028.

— (1994). The Bias of Kaplan-Meier Integrals. *Scandinavian Journal of Statistics* 21.4, pp. 475–484.

Stute, W. (1995). The Central Limit Theorem Under Random Censorship. *The Annals of Statistics* 23.2, pp. 422–439. DOI: 10.1214/aos/1176324528.

— (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian journal of statistics* 23.4, pp. 461–471.

Suzukawa, A. (2002). Asymptotic properties of Aalen-Johansen integrals for competing risks data. *Journal of the Japan Statistical Society* 32, pp. 77–93. DOI: 10.14490/jjss.32.77.