Exact MLE and asymptotic properties for nonparametric semi-Markov models

Samis Trevezas, Nikolaos Limnios

▶ To cite this version:

RESEARCH ARTICLE

Exact MLE and asymptotic properties for nonparametric Semi-Markov models

Samis Trevezas[a,b*]   and Nikolaos Limnios[c]

[a]*Laboratoire Mathématiques Appliquées aux Systèmes, Ecole Centrale Paris, France*; [b]*EPI Digiplante, INRIA Saclay - Île-de-France, France*; [c]*Laboratoire de Mathématiques Appliquées de Compiègne, Université de Technologie de Compiègne, France*

This article concerns maximum likelihood estimation for discrete time homogeneous nonparametric semi-Markov models with finite state space. In particular, we present the exact maximum likelihood estimator of the semi-Markov kernel which governs the evolution of the semi-Markov chain. We study its asymptotic properties in the following cases: (i) for one observed trajectory, when the length of the observation tends to infinity, and (ii) for parallel observations of independent copies of a semi-Markov chain censored at a fixed time, when the number of copies tends to infinity. In both cases, we obtain strong consistency, asymptotic normality, and asymptotic efficiency for every finite dimensional vector of this estimator. Finally, we obtain explicit forms for the covariance matrices of the asymptotic distributions.

Keywords: Asymptotic efficiency; Asymptotic normality; Backward recurrence times; Consistency; Maximum likelihood estimation; Nonparametric estimation; Semi-Markov chains

1. Introduction

As opposed to Semi-Markov Processes (SMP) in continuous time, the probabilistic modeling in discrete time has not received analogous attention. One of the earliest references

---

*Corresponding author. Email: samis.trevezas@ecp.fr

in discrete time is Anselone (1960), where the author develops some aspects of the ergodic theory of discrete semi-Markov chains (SMCs). When the SMC is not ergodic, but transient with one absorbing state, Gerontidis (1994) studies the long-term behaviour of semi-Markov replacement chains, where independent copies are connected every time an absorption occurs. Applications of semi-Markov models with emphasis on decision processes are given by Howard (1971). Other applications, such as queuing theory and epidemiology, as well as algorithms for implementing SMCs on a computer, are presented by Mode and Pickens (1998) and Mode and Sleeman (2000). For recent references on theory and applications (e.g. reliability) of discrete SMCs see Barbu and Limnios (2008) and Barbu et al. (2004).

As far as statistical estimation in discrete time is concerned, Barbu and Limnios (2006) considered empirical estimation for nonparametric semi-Markov models (SMMs) in the case of one observed trajectory. The authors constructed an empirical estimator for the semi-Markov kernel (the kernel that determines the evolution of a SMC) at a finite observation interval, and studied some of its asymptotic properties when the observation length tends to infinity. In particular, they obtained strong consistency and asymptotic normality for each component of the empirical estimator. This estimator can be considered as an approximate maximum likelihood estimator (MLE), since for its construction the sojourn time in the last visited state is systematically neglected. The asymptotic results hold also true for each component of the exact MLE.

In this study, we consider a unified approach in order to tackle the problem of estimation for a single observed trajectory, as well as for multiple trajectories that are generated by independent copies of a semi-Markov chain censored at a fixed time. Both scenarios are interesting for applications. For general references on statistical estimation based on both scenarios, see Kalbfleisch and Prentice (2002), with emphasis on models and methods for the analysis of failure time data, or Anderson et al. (1993). Let us take an example in reliability. In the case of one observed trajectory, one is mostly concerned with the evolution of an ergodic semi-Markov chain, describing the temporal evolution of one reparable system or one component of the system. The case of L independent copies of a semi-Markov chain arises as well in this context, when one sets into operation L identical components or systems. In this case, there is a useful fixed life time for each system. At the end of this time, the system is rejected independently of its current state. For further applications that concern multiple observed trajectories, see Section 4.

In this article, we present the exact MLE that corresponds to the nonparametric SMM, in order to refine the corresponding estimation of the approximate MLE at a finite interval. We generalise the results of Barbu and Limnios (2006) by obtaining asymptotic normality for every finite vector of components of the MLE, proving at the same time that the MLE is asymptotically efficient. Additionally, we study the case of parallel observations of independent copies of a SMC and we obtain asymptotic properties, such as strong consistency, asymptotic normality, and asymptotic efficiency, for the part of the semi-Markov kernel that can be estimated until the censoring time. This unified approach reveals that even if the sojourn time in the last visited state can be neglected for obtaining the asymptotic properties of the MLE in the case of one observed trajectory, this is not the case for parallel observations censored at a fixed time. When the last sojourn time is systematically neglected, then the SMCs under study are forced to have a finite support (bounded by the censoring time), and the information for the survival probabilities at the censoring time gets lost.

Our proposed method for recovering this information and obtaining the exact MLE is based on a reparameterisation of the semi-Markov kernel by the parameters that

correspond to the natural parametric space of an associated Markov chain (MC), which is formed by the SMC and the sequence of backward recurrence times. The associated MC has an infinite state space under the condition that there exists a state that has conditional sojourn time distribution with an infinite support. Consequently, in this case, the reparameterisation of the semi-Markov kernel by the natural parametric space of the coupled process implies an infinite number of parameters. This justifies the nonparametric framework. The associated MC can play an important role in the understanding of the SMC. Actually, it can be useful, on one hand, to study the probabilistic behavior and limit theorems for SMCs (see, for example, Stenflo (1996) and Chryssaphinou et al. (2008)), and on the other hand, it can be used to develop statistical inference for the SMMs. The goal of this paper is to develop the latter idea and study MLE for this class of models.

The rest of this paper is organised as follows: In Section 2, we introduce the necessary mathematical notation. In Section 3, we give the exact MLE that corresponds to the semi-Markov kernel of the SMC and we study some of its asymptotic properties. Finally, in Section 4, we study the problem of MLE for parallel observations of independent copies of a SMC censored at a fixed time.

2.   Preliminaries and Notation

Let E be a finite set with cardinality $s \in \mathbb{N}^*$ and $\mathcal{M}_E$ be the set of real matrices on $E \times E$. We denote by $\mathcal{M}_E(\mathbb{N})$ the set of matrix-valued functions defined on $\mathbb{N}$, with values in $\mathcal{M}_E$. The following definitions introduce the semi-Markov kernel, the Markov renewal chain (MRC) and the SMC.

Definition 2.1:   A matrix valued function $q := \big(q(k),\, k \in \mathbb{N}\big) \in \mathcal{M}_E(\mathbb{N})$ is said to be a discrete time semi-Markov kernel if it satisfies the following two properties:

(1) $q_{ij}(k) \geq 0,\ i,j \in E,\ k \in \mathbb{N},$
(2) $\sum_{k=0}^{\infty} \sum_{j \in E} q_{ij}(k) = 1,\ i \in E.$

Definition 2.2:   Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathbf{J}, \mathbf{S}) := (J_l, S_l)_{l \in \mathbb{N}}$ a bivariate chain with values in $E \times \mathbb{N}$. The chain $(\mathbf{J}, \mathbf{S})$ is said to be a (homogeneous) Markov renewal chain (MRC) associated to the semi-Markov kernel q if for all $k, l \in \mathbb{N}$, $i, j \in E$, it satisfies $\mathbb{P}$-a.s.

$$\mathbb{P}(J_{l+1} = j, S_{l+1} - S_l = k \mid J_l = i, J_{l-1}, \dots, J_0, S_l, \dots, S_0)$$
$$= \mathbb{P}(J_{l+1} = j, S_{l+1} - S_l = k \mid J_l = i) = q_{ij}(k).$$

For the rest we assume that $q(0)$ is the null matrix, and this implies that $\mathbf{S}$ is a strictly increasing sequence (therefore excluding instantaneous transitions).

Definition 2.3:   The chain $\mathbf{Z} := (Z_n)_{n \in \mathbb{N}}$ with values in E is said to be a (homogeneous) semi-Markov chain (SMC) associated to $(\mathbf{J}, \mathbf{S})$ if

$$Z_n = J_{N(n)}, \quad n \in \mathbb{N},$$

where

$$N(n) = \max\{l \in \mathbb{N} : S_l - S_0 \leq n\}, \quad n \in \mathbb{N}. \tag{1}$$

The process $\mathbf{S}$ keeps track of the successive time points that changes of states in $\mathbf{Z}$ occur (jump times), and $\mathbf{J}$ records the visited states at these time points. We also make the assumption that $S_0 = 0$, and this means that the first observation of the SMC is considered to be on a jump time. The sequence defined by (1) is the counting process of the number of jumps (changes of states) until time n. Let also $\mathbf{X} := (X_l)_{l \in \mathbb{N}^*}$, where $X_{l+1} = S_{l+1} - S_l$. The process $\mathbf{X}$ corresponds to the sequence of total sojourn times in the successively visited states.

Remark 1: We want a jump time to indicate a change of state, so we will assume that $q_{ii}(k) = 0$, for all $i \in E, \ k \in \mathbb{N}$. This assumption excludes absorbing states from our analysis, but in any case, we can modify easily our setting by partitioning the state space in a subset where this condition holds true and in the complementary set, where only absorbing states are included and satisfy $\sum_k q_{ii}(k) = 1$. In this case, each absorbing state forms a simple renewal chain.

For the statistical analysis, we are interested in the nonparametric semi-Markov model, where the SMC $\mathbf{Z}$ is defined on the probability space $(\Omega, \mathscr{A}, \mathbb{P}_{\boldsymbol{q}})$, and the semi-Markov kernel q is identified with the infinite dimensional vector $\boldsymbol{q} := \big(q_{ij}(k)\big)_{i,j \in E, k \in \mathbb{N}}$. We also assume that its components keep the lexicographical order of the triplet (i, k, j).

Let $E_{-i} := E \setminus \{i\}$. We denote by

$$\mathcal{Q}_i := \left\{ \boldsymbol{q}_i = \big(q_{ij}(k)\big) \in \mathbb{R}^{(\mathbb{N}^*)^{s-1}}; \ q_{ij}(k) \geq 0, \ \sum_{j \in E_{-i}} \sum_{k \in \mathbb{N}^*} q_{ij}(k) = 1 \right\},$$

for all $i \in E$, the parametric space of all the possible discrete distributions on the set $(\mathbb{N}^*)^{s-1}$, where the components of the vector $\boldsymbol{q}_i$ are considered to be ordered with respect to the lexicographical order of $(k, j) \in \mathbb{N}^* \times E$. Then, by taking s independent copies of the space $\mathcal{Q}_i$, we will represent the semi-Markov kernel $\boldsymbol{q}$ as a point in the parametric space $\mathcal{Q} := \prod_{i \in E} \mathcal{Q}_i$.

Let $\mathbf{U} := (U_n)_{n \in \mathbb{N}}$ be the sequence of backward recurrence times (BRT), where

$$U_n = n - S_{N(n)}.$$

Note that $U_0 = 0$, since $S_0 = 0$ by assumption. Each BRT U_n keeps track of the elapsed time after the last jump of the process $\mathbf{Z}$ until time n. The pair $(\mathbf{Z}, \mathbf{U})$ can be shown to be a Markov chain (e.g., Limnios and Oprişan (2001), Theorem 3.12). See also Chryssaphinou et al. (2008), where the authors study some properties of $(\mathbf{Z}, \mathbf{U})$. In this paper, the identically non-zero transition probabilities of the MC $(\mathbf{Z}, \mathbf{U})$ will serve as a suitable reparameterisation of the semi-Markov kernel $\boldsymbol{q}$, in order to obtain its exact MLE.

Now, we denote by $\bar{H}_i(\cdot)$ the survival function in state i, which is given as a function of the semi-Markov kernel as follows:

$$\bar{H}_i(u) := \mathbb{P}_{\boldsymbol{q}}(X_{l+1} > u \mid J_l = i) = 1 - \sum_{j \in E} \sum_{k=0}^{u} q_{ij}(k), \ u, l \in \mathbb{N}. \tag{2}$$

If $(\mathbf{Z}, \mathbf{U})$ is irreducible and positive recurrent, then the corresponding stationary proba-

bilities are given by

$$\pi_{i,u} = \frac{\bar{H}_i(u)}{\mu_{ii}}, \quad (i, u) \in E \times \mathbb{N},$$

where μ_{ii} corresponds to the mean recurrence time between two successive visits of the MC $(\mathbf{Z}, \mathbf{U})$ to the state $(i, 0)$. The mean recurrence time has the following representation

$$\mu_{ii} = \mathbb{E}(G_i \mid Z_0 = i, U_0 = 0), \tag{3}$$

where G_i is the first hitting time of state $i \in E$, after leaving the initial state, that is,

$$G_i = \inf\{n \geq S_1 : Z_n = i\}.$$

For the remainder let $T_m := \{0, 1, \ldots, m-1\}$ and T_m^+ be the set T_m translated one unit to the right. Let also $E_*^2 := E^2 \setminus \{(i, i), \ i \in E\}$. Now, we define some counting processes that will be used in the sequel, appropriate for the pair $(\mathbf{J}, \mathbf{X})$:

$$N_{ij}(k, m) := \sum_{l=0}^{N(m)-1} \mathbb{1}_{\{J_l=i, J_{l+1}=j, X_{l+1}=k\}}, \quad (i, j) \in E_*^2, \ k \in T_m^+, \tag{4}$$

$$N_i(m) := \sum_{l=0}^{N(m)-1} \mathbb{1}_{\{J_l=i\}}. \qquad\qquad i \in E, \tag{5}$$

The process $N_{ij}(k, m)$ counts the number of (one-step) transitions of $\mathbf{J}$, from i to j, that took place in k time units, until time m. The process $N_i(m)$ counts the number of visits of $\mathbf{J}$ to state i, until time m, not counting the last visit $J_{N(m)}$. Note that the time step is given by the process $\mathbf{Z}$.

Similarly, we define some counting processes, appropriate for the pair $(\mathbf{Z}, \mathbf{U})$:

$$N_{i,u}^B(j; m) := \sum_{n=0}^{m-1} \mathbb{1}_{\{Z_n=i, U_n=u, Z_{n+1}=j\}}, \ (i, j) \in E^2, \ u \in T_m, \tag{6}$$

$$N_{i,u}^B(m) := \sum_{n=0}^{m-1} \mathbb{1}_{\{Z_n=i, U_n=u\}}, \qquad\qquad (i, u) \in E \times T_m. \tag{7}$$

The process $N_{i,u}^B(j; m)$ counts the number of (one-step) transitions of $(\mathbf{Z}, \mathbf{U})$, from (i, u) to (j, k), until time m. Since k is uniquely determined when the next j is known, we refer only to j for notational convenience. The process $N_{i,u}^B(m)$ counts the number of visits of $(\mathbf{Z}, \mathbf{U})$ to state (i, u), until time $m-1$. We note that the superindex B (from Backward) is only used in order to distinguish easier the counting processes that refer to $(\mathbf{Z}, \mathbf{U})$.

If we denote by

$$M_{(s)} := \inf\{m \in \mathbb{N}^* : N_i(m) > 0, \forall i \in E\},$$

then for $m \geq M_{(s)}$, every state and at least one transition from every state of the MC $\mathbf{J}$ to some other state have been recorded.

3.   Exact MLE based on a single trajectory

In this Section, we construct the exact MLE of the semi-markov kernel, when a fixed length observation $\mathbf{Z}_0^m := (Z_0, Z_1, \ldots, Z_m)$ of a SMC $\mathbf{Z}$ is available, and we study its asymptotic properties as m tends to infinity.

The likelihood function $\mathcal{L}_m(\boldsymbol{q}) := \mathbb{P}_{\boldsymbol{q}}(\mathbf{Z}_0^m)$ that corresponds to an observation $\mathbf{Z}_0^m$ can be written as:

$$\mathcal{L}_m(\boldsymbol{q}) = a(Z_0) \left(\prod_{k=1}^{N(m)} q_{J_{k-1}J_k}(X_k) \right) \bar{H}_{J_{N(m)}}(U_m), \tag{8}$$

where $a(\cdot)$ is the initial distribution and $\bar{H}_i(\cdot)$ defined as in Relation (2).

Since the factor $\bar{H}_{J_{N(m)}}(U_m)$ that appears in equation (8) corresponds to the sojourn time in the last visited state of the system until time m, the information that we get for the semi-Markov kernel from this part can be neglected for large m for positive recurrent SMCs (see Barbu and Limnios (2008)). This leads to an approximate MLE (quasi-MLE), which is asymptotically equivalent with the exact one in the case of a single trajectory. Barbu and Limnios (2006) used this fact and considered the approximate MLE in order to develop statistical inference for nonparametric SMMs. Under the assumption that the initial distribution does not depend on $\boldsymbol{q}$, and by (4), the authors reduced the maximisation problem to the maximisation of an approximate log-likelihood function:

$$\tilde{l}_m(\boldsymbol{q}) = \sum_{i,j \in E_*^2} \sum_{k \in T_m^+} N_{ij}(k, m) \log q_{ij}(k). \tag{9}$$

In this way, they obtained the following M-estimator (see Van Der Vaart (1998) for a general treatment of M-estimators):

$$\tilde{q}_{ij}(k, m) = \frac{N_{ij}(k, m)}{N_i(m)}. \tag{10}$$

If $N_i(m) = 0$, we take by convention $0/0 = 0$.

Remark 1 :   This convention allows us to extend the estimation to the case that we have not observed all the states ($m < M_{(s)}$). This is valid if we assume that the considered s-state SMCs that we deal with, are inside a larger parametric space that includes all the SMCs with cardinality bounded by s, or equivalently, the cardinality of the state space of the considered SMCs is a parameter bounded by s. This remark extends and proves to be fruitful even in the case of the chain $(\mathbf{Z}, \mathbf{U})$. This pair forms an infinite state space Markov chain on condition that there exists at least one state of the SMC $\mathbf{Z}$ with unbounded sojourn time. In this case, the state space of the considered MCs should be considered as a nonnegative integer parameter (not bounded).

3.1.   *Construction of the MLE*

In this subsection, we infer the exact MLE of this model and its relationship with the M-estimator (empirical estimator or approximate MLE) given by (9).

Let

$$p_{i,u;j} := \begin{cases} \mathbb{P}_{\mathbf{q}}(Z_{n+1} = j, U_{n+1} = 0 \mid Z_n = i, U_n = u) & \text{if } j \neq i, \\ \mathbb{P}_{\mathbf{q}}(Z_{n+1} = i, U_{n+1} = u + 1 \mid Z_n = i, U_n = u) & \text{if } j = i. \end{cases} \tag{11}$$

We will reparameterise the model by the natural parameter that corresponds to the MC $(\mathbf{Z}, \mathbf{U})$, that is, $\boldsymbol{\eta} := (p_{i,u;j})_{i,j \in E, u \in \mathbb{N}}$, where all the identically zero probabilities of the transition matrix of $(\mathbf{Z}, \mathbf{U})$ have been excluded. Note also that $\boldsymbol{q} \equiv \Phi(\boldsymbol{\eta})$, where the component functions of Φ for $i \neq j$ are given by

$$\Phi_{ijk}(\boldsymbol{\eta}) = \begin{cases} p_{i,0;j} & \text{if } k = 1, \\ p_{i,k-1;j} \prod_{u=0}^{k-2} p_{i,u;i} & \text{if } k \geq 2. \end{cases} \tag{12}$$

Each one of the functions Φ_{ijk} depends actually on k variables (finite number). Let $\boldsymbol{\eta}_m := (p_{i,u;j})_{i,j \in E, u \in T_m}$, where $\boldsymbol{\eta}_m$ parameterise the first sm distributions with finite support (see Relation (11)), for which we can have information until time m.

In the following proposition, we will obtain the MLE for the above distributions and we will transfer the results to the MLE for the parameters $\boldsymbol{q}_m := (q_{ij}(k))_{(i,j) \in E_*^2, k \in T_m^+}$ of the semi-Markov kernel $\boldsymbol{q}$ that can be estimated until time m. Additionally, we connect the empirical estimators $\tilde{q}_{ij}(k, m)$ given by (10) with the MLE $\widehat{q}_{ij}(k, m)$ obtained here.

Proposition 3.1:   *i) The MLE that corresponds to a fixed length observation $\mathbf{Z}_0^m$ of a nonparametric SMM, driven by the semi-Markov kernel $\boldsymbol{q}$, is given for $i \neq j$ by*

$$\widehat{q}_{ij}(k, m) = \Phi_{ijk}(\widehat{\boldsymbol{\eta}}_m(m)) = \widehat{p}_{i,k-1;j}(m)\widehat{H}_i(k-1, m), k \in T_m^+, \tag{13}$$

where

$$\widehat{H}_i(k-1, m) = \prod_{u=0}^{k-2} \widehat{p}_{i,u;i}(m).$$

ii) The exact relationship between the MLE $\widehat{q}_{ij}(k, m)$, given by (13), and the approximate MLE $\tilde{q}_{ij}(k, m)$, given by (10), reads for (i, j, k) such that $\tilde{q}_{ij}(k, m) > 0$ as follows

$$\widehat{q}_{ij}(k, m) = \tilde{q}_{ij}(k, m) \left(1 - \frac{\mathbb{1}_{\{Z_m = i, U_m > 0\}}}{N_i(m) + 1} \right) \left(1 + \frac{\mathbb{1}_{\{Z_m = i, k > U_m > 0\}}}{N_{i, U_m}^B(m)} \right). \tag{14}$$

Proof: By using the counting processes (6), the likelihood function $L_m(\boldsymbol{q})$ given by (8) can be written as a function of $\boldsymbol{\eta}_m$ in the form

$$L_m(\boldsymbol{\eta}_m) = a(Z_0) \prod_{i,j \in E} \prod_{u \in T_m} p_{i,u;j}^{N_{i,u}^B(j;m)}.$$

Since the initial distribution does not vary with $\boldsymbol{\eta}_m$, the log-likelihood function that has to be maximized is given by

$$l_m(\boldsymbol{\eta}_m) = \sum_{i,j \in E} \sum_{u \in T_m} N_{i,u}^B(j; m) \log p_{i,u;j}. \tag{15}$$

The dependence relations that we have to take into consideration in a nonparametric framework are imposed from the stochastic nature of the transition probability matrix P of the MC $(\mathbf{Z}, \mathbf{U})$. For each $i \in E$, $u \in \mathbb{N}$,

$$\sum_{j \in E} p_{i,u;j} = 1. \tag{16}$$

By using the expressions (6) and (7), then by maximizing with respect to $\boldsymbol{\eta}_m$, under the constraints (16), the likelihood function $l_m(\boldsymbol{\eta}_m)$ given by (15), we get easily

$$\widehat{\boldsymbol{\eta}}_m(m) = \left(\widehat{p}_{i,u;j}(m)\right) = \left(\frac{N^B_{i,u}(j;m)}{N^B_{i,u}(m)}\right).$$

From the invariance principle of MLE and Relation (12) we conclude (13).
ii) By using (6) and (7), Relation (13) can be written in the form

$$\widehat{q}_{ij}(k,m) = \frac{N^B_{i,k-1}(j;m)}{N^B_{i,k-1}(m)} \frac{N^B_{i,k-2}(i;m)}{N^B_{i,k-2}(m)} \cdots \frac{N^B_{i,1}(i;m)}{N^B_{i,1}(m)} \frac{N^B_{i,0}(i;m)}{N^B_{i,0}(m)}, \ k \geq 2. \tag{17}$$

Notice that

$$N^B_{i,k-1}(j;m) = N_{ij}(k,m), \tag{18}$$

since in order to have a transition from i to j in k time units for the process $\mathbf{Z}$, a transition from $(i, k-1)$ to $(j, 0)$ has to take place for the process $(\mathbf{Z}, \mathbf{U})$. Moreover, for $1 \leq l \leq k-1$ we have

$$N^B_{i,l-1}(i;m) = N^B_{i,l}(m) + \mathbb{1}_{\{Z_m=i,U_m=l\}}, \tag{19}$$

and finally

$$N^B_{i,0}(m) = N_i(m) + \mathbb{1}_{\{Z_m=i,U_m>0\}}. \tag{20}$$

By inserting (18)-(20) in Relation (17) we get

$$\widehat{q}_{ij}(k,m) = \frac{N_{ij}(k,m)\left(N^B_{i,U_m}(m) + \mathbb{1}_{\{Z_m=i,k>U_m>0\}}\right)}{\left(N_i(m) + \mathbb{1}_{\{Z_m=i,U_m>0\}}\right) N^B_{i,U_m}(m)}. \tag{21}$$

By equation (21) we conclude easily (14). □

Remark 2: Note that $N^B_{i,U_m}(m) = 0$ iff at time m the system stayed for the first time at state i for U_m time units. Consequently, the subsequence of m that satisfies this equation defines the time points that a new state of $(\mathbf{Z}, \mathbf{U})$ has been recorded. Also, at these time points we have that $\mathbb{1}_{\{Z_m=i,k>U_m>0\}} = 0$ for k such that $\tilde{q}_{ij}(k,m) > 0$, since at time m no transition has occurred from state i to state j in more than U_m time units. Therefore, the right-hand member of equation (14) is well defined by the convention $0/0 = 0$.

3.2.   *Asymptotic properties of the MLE*

In this subsection, we study some asymptotic properties of the nonparametric MLE of the semi-Markov kernel. Theorem 3.4 summarises the asymptotic results in the framework of convergence of stochastic processes, since we have an infinite dimensional parametric space. Under a suitable condition, the MLE is strongly consistent and every finite-dimensional subvector of the MLE is asymptotically normally distributed and asymptotically efficient. In order to prove asymptotic normality, we use the reparameterisation (12) of the semi-Markov kernel. For every fixed $\tilde{n} \in \mathbb{N}$, we obtain in Lemma 3.2 the asymptotic distribution of the MLE of the vector parameter $\boldsymbol{\eta}_{\tilde{n}}$ and then we transfer this result to the MLE of the vector parameter $\boldsymbol{q}_{\tilde{n}}$ (Proposition 3.3).

Now, we state the following condition concerning the subclass of SMCs to be considered:

Condition (A1): The SMC $\mathbf{Z}$ under $\mathbb{P}_{\boldsymbol{q}}$ is irreducible and positive recurrent.

The above condition is necessary in the case that we search the long term behavior of the MLE for one trajectory of the SMC $\mathbf{Z}$. This condition will be relaxed in the next Section, where we study the asymptotics of the MLE for an increasing number of independent realisations of a SMC $\mathbf{Z}$, censored at a fixed time. Also, for notational simplicity, we assume in this Section that all the elements of the semi-Markov kernel are strictly positive under the true value. This assumption implies that condition (A1) holds. Otherwise, the following Lemma holds true for the MLE that corresponds to the probabilities that are strictly positive under the true value. The proof of this Lemma is deferred to Appendix A.

Lemma 3.2:   *Let $(\mathbf{Z}, \mathbf{U})$ be the MC associated with the SMC $\mathbf{Z}$, where $\mathbf{U}$ is the sequence of BRT. Under condition (A1), the sequence of random vectors $\boldsymbol{\xi}_{\tilde{n}}(m) := \sqrt{m}(\widehat{\boldsymbol{\eta}}_{\tilde{n}}(m) - \boldsymbol{\eta}_{\tilde{n}})$, converges $\mathbb{P}_{\boldsymbol{\eta}}-$weakly, as m tends to infinity, to a random vector $\boldsymbol{\xi}_{\tilde{n}}$, where $\boldsymbol{\xi}_{\tilde{n}} \sim \mathcal{N}(\mathbf{0}, \Gamma_{\tilde{n}})$ and $\Gamma_{\tilde{n}}$ is a covariance matrix of dimension $(s^2\tilde{n}) \times (s^2\tilde{n})$, defined in a block diagonal form*

$$\Gamma_{\tilde{n}} = \mathsf{diag}\{\frac{1}{\pi_{i,u}}\Lambda_{iu}; (i,u) \in E \times T_{\tilde{n}}\}, \tag{22}$$

where the matrix Λ_{iu}, for each $(i,u) \in E \times T_{\tilde{n}}$ is given by

$$\Lambda_{iu} = \big(p_{i,u;j}[\delta_{jl} - p_{i,u;l}]\big)_{j,l \in E}. \tag{23}$$

The proof of the following Proposition is based on Lemma 3.2 and it is deferred to Appendix A.

Proposition 3.3:   *Under condition (A1), the sequence of random vectors $\boldsymbol{\xi}_{\tilde{n}}(m) := \sqrt{m}(\widehat{\boldsymbol{q}}_{\tilde{n}}(m) - \boldsymbol{q}_{\tilde{n}})$, converges $\mathbb{P}_{\boldsymbol{q}}-$weakly, as m tends to infinity, to a random vector $\boldsymbol{\xi}_{\tilde{n}}$, where $\boldsymbol{\xi}_{\tilde{n}} \sim \mathcal{N}(\mathbf{0}, \Delta_{\tilde{n}})$ and $\Delta_{\tilde{n}}$ is a covariance matrix of dimension $(s^2 - s)\tilde{n} \times (s^2 - s)\tilde{n}$, defined in a block diagonal form*

$$\Delta_{\tilde{n}} = \mathsf{diag}\{\mu_{ii}K_i; i \in E\}, \tag{24}$$

where the matrix K_i, for each $i \in E$ is given by

$$K_i = \big(q_{ij}(k)[\delta_{(j,k)(r,l)} - q_{ir}(l)]\big)_{(j,k),(r,l)\in E_{-i}\times T_{\tilde{n}}^+}. \tag{25}$$

and μ_{ii} is given by (3).

Remark 3: A direct proof of Proposition 3.3 is also possible by means of the technique that we use in order to prove Lemma 3.2. In this case, the martingale CLT can be applied by considering an appropriate martingale difference with respect to the natural filtration generated by the Markov chain $(\mathbf{J}, \mathbf{X})$. A possible choice of the martingale difference (compare with (A8), in the case of the MC $(\mathbf{Z}, \mathbf{U})$) is given as follows:

$$Y_n := \mathbb{1}_{\{X_n \leq \tilde{n}\}}[t_{J_{n-1}J_nX_n} - \mathbb{E}_q(\mathbb{1}_{\{X_n \leq \tilde{n}\}}t_{J_{n-1}J_nX_n} \mid J_{n-1})]. \tag{26}$$

Nevertheless, our method gives a unified treatment of the study of the MLE, when we have a single or multiple trajectories.

In order to prove asymptotic efficiency, we will use the same reparameterisation (12). Since this property is parameterisation independent, it is enough to show that for an arbitrary $\tilde{n} \in \mathbb{N}$ the MLE of the vector parameter $\boldsymbol{\eta}_{\tilde{n}}$ is asymptotically efficient. This will be true, if the asymptotic covariance matrix of the MLE of the vector parameter $\boldsymbol{\eta}_{\tilde{n}}$ is the inverse of the asymptotic FIM for this model. Then, the MLE of the vector parameter $\boldsymbol{q}_{\tilde{n}}$ will be asymptotically efficient as well.

Let $\boldsymbol{\eta}_m^*$ be a minimal representation of $\boldsymbol{\eta}_m$, by excluding for each (i,u) the parameter $p_{i,u;i}$ (see Relation (16), for the dependence relations). In order to compute the asymptotic Fisher Information Matrix (FIM), we define first the following matrix:

$$I(\boldsymbol{\eta}_m^*) := \left(\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}\left(\frac{-\partial^2\ell_m(\boldsymbol{\eta}_m)}{\partial p_{i_1,u_1;j_1}\partial p_{i_2,u_2;j_2}}\right)\right)_{(i_1,u_1,j_1),(i_2,u_2,j_2)}, \tag{27}$$

where the admissible indices (i_1, u_1, j_1) and (i_2, u_2, j_2) are these that correspond to $\boldsymbol{\eta}_m^*$, and $\ell_m(\boldsymbol{\eta}_m)$ is given by (15). The matrix $I(\boldsymbol{\eta}_m^*)$ is an unnormalised FIM that corresponds to the trajectory $\mathbf{Z}_0^m$, when the scenario of a single trajectory of a SMC is considered. When multiple trajectories (i.i.d.) of a SMC censored at time m are considered (see Section 4), then it is the exact FIM that corresponds to this scenario. The matrix

$$I_m(\boldsymbol{\eta}_{\tilde{n}}^*) := m^{-1}I(\boldsymbol{\eta}_{\tilde{n}}^*) \tag{28}$$

is a normalised version (information per unit time) of the matrix defined by (27), where now the admissible indices (i_1, u_1, j_1) and (i_2, u_2, j_2) are these that correspond to $\boldsymbol{\eta}_{\tilde{n}}^*$ (recall that $\boldsymbol{\eta}_{\tilde{n}}^*$ parameterises the first $s\tilde{n}$ transition probability distributions of $(\mathbf{Z}, \mathbf{U})$).

Now, let for $m \in \mathbb{N}^*$,

$$\widehat{\boldsymbol{q}}(m) = \big(\widehat{\boldsymbol{q}}_m(m), 0, 0, \dots\big),$$

that is, the sequence of estimators that can be decomposed into two parts, a finite part that corresponds to the MLE of the vector $\boldsymbol{q}_m$, and the second part that completes with infinite zeros the parameters that we have no information up to time m.

Theorem 3.4:   *Under condition (A1),*
i) the sequence of MLE $\widehat{\boldsymbol{q}}(m)$ is a strongly consistent estimator of $\boldsymbol{q}$, that is, as $m \to \infty$,

$$\widehat{\boldsymbol{q}}(m) \xrightarrow{\mathbb{P}_q\text{-}a.s.} \boldsymbol{q},$$

ii) the sequence of random processes $\boldsymbol{\xi}(m) := \big(\boldsymbol{\xi}_m(m), 0, 0, \dots\big)$, $m \in \mathbb{N}^*$, where $\boldsymbol{\xi}_m(m) := \sqrt{m}(\widehat{\boldsymbol{q}}_m(m) - \boldsymbol{q}_m)$, converges $\mathbb{P}_{\boldsymbol{q}}-$weakly, as m tends to infinity, to a random process $\boldsymbol{\xi}$, where for any $\tilde{n} \in \mathbb{N}^*$, the corresponding random vector $\boldsymbol{\xi}_{\tilde{n}} \sim \mathcal{N}(\boldsymbol{0}, \Delta_{\tilde{n}})$, and $\Delta_{\tilde{n}}$ is given by (24),
iii) every finite-dimensional subvector of the MLE $\widehat{\boldsymbol{q}}(m)$ is asymptotically efficient.

Proof: i) It is straightforward to see by Relation (14) that the MLE $\widehat{q}_{ij}(k, m)$ and the empirical estimator $\tilde{q}_{ij}(k, m)$ of the parameter $q_{ij}(k, m)$ converge $\mathbb{P}_{\boldsymbol{q}}$-a.s. to the same limit, if that limit exists. Barbu and Limnios (2008) have proved that $\tilde{q}_{ij}(k, m)$ is strongly consistent (see Theorem 4.1.). Therefore, the strong consistency of $\widehat{\boldsymbol{q}}(m)$ follows from the coordinatewise consistency.

ii) The sequence of random processes $\boldsymbol{\xi}(m)$ converges $\mathbb{P}_{\boldsymbol{q}}-$weakly to the process $\boldsymbol{\xi}$ iff for any $\tilde{n} \in \mathbb{N}^*$ the sequence of random vectors $\boldsymbol{\xi}_{\tilde{n}}(m)$ converges $\mathbb{P}_{\boldsymbol{q}}-$weakly, as m tends to infinity, to the corresponding finite dimensional vector $\boldsymbol{\xi}_{\tilde{n}}$. This result is a consequence of Proposition 3.3.

iii) Since the asymptotic efficiency of the MLE is parameterization independent, it suffices to show that the MLE $\widehat{\boldsymbol{\eta}}_{\tilde{n}}(m)$ of the vector parameter $\boldsymbol{\eta}_{\tilde{n}}$ (for an arbitrary $\tilde{n}$) is asymptotically efficient. This is true if the limiting matrix $I_\infty(\boldsymbol{\eta}_{\tilde{n}}^*) := \lim_{m \to \infty} I_m(\boldsymbol{\eta}_{\tilde{n}}^*)$, where $I_m(\boldsymbol{\eta}_{\tilde{n}}^*)$ is given by (28), exists and its inverse equals the asymptotic covariance matrix $\Gamma_{\tilde{n}}$ given by (22). The proof of this assertion is deferred to Appendix A. $\square$

4.   Independent observations of a SMC censored at a fixed time

In this Section, we consider L trajectories of independent copies of a SMC **Z** censored at a fixed time m. The setting of multiple independent trajectories arises in many situations, and statistical inference should be developed in this framework, when each trajectory is assumed to follow semi-Markov dynamics. We give here some relative examples: i) in reliability, when L identical but independent systems are observed for a fixed time period (see Introduction), ii) in epidemiology (biomedical interest), when L patients, victims of a certain disease, are observed during disease progression (see, e.g., Commenges (1999), for the so-called multistate models). Nevertheless, in this case, the SMCs are absorbing, including a fatal state, iii) In DNA analysis, when L large sets of trusted examples (called training sets and correspond here to fixed-length trajectories) are typically used in order to estimate the parameters of a probabilistic model that underlies the generation of a sequence of DNA bases. Hidden Markov models and hidden semi-Markov models have been used extensively in this context (see Durbin et al. (1998) and Burge (1997)). In the case of hidden semi-Markov models, there is a hidden semi-Markov chain that models the succession of different regimes (evolutionary states), where the emission probability distribution of the DNA bases changes. MLE based methods can be used in order to combine the different training sets and obtain the parameter estimates, iv) in Monte-Carlo EM (Expectation-Maximisation) type algorithms for hidden semi-Markov models (intended for inference with a large data set), where independent trajectories of a SMC are generated (simulated) at each iteration of the EM algorithm (at each E-step, and

under the current value of the parameter), in order to approximate the so-called Q-function of the EM-algorithm (equivalent to the incomplete likelihood function, as far as maximisation is concerned) that corresponds to this model (see, e.g., Malefaki et al. (2010)).

In the sequel, we study the MLE that corresponds to this model and we are interested in its asymptotic properties. We denote by $(\mathbf{Z}_0^m)_l := (Z_0^{(l)}, Z_1^{(l)}, \ldots, Z_m^{(l)})$ the l-th realisation of independent copies of a SMC $\mathbf{Z}$ censored at the fixed time m. If the problem was to make inferences for a class of non positive recurrent SMCs, then the estimation based on a single trajectory of a SMC is not satisfactory. On the other hand, when many trajectories corresponding to independent copies of a CSM are available, the structure of the model can be better explored. For this reason, we make no assumption here for the type of communication among the different states of the SMC $\mathbf{Z}$. For state classification of a finite-state SMC and different types of communication, see, e.g., Definition 3.11 in Barbu and Limnios (2008).

Now, we give a suitable condition for this type of estimation.

Condition (A2): The initial distribution $\alpha(\cdot)$ under $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{q}}$

i) has a support that gives strictly positive probability to every communication class of $\mathbf{Z}$,
ii) does not depend on $\boldsymbol{q}$.

Interpretation. Since in each trajectory, it is only the first observed state that determines the set of observed states, it suffices to assume that at least one state for each communication class of $\mathbf{Z}$ is accessible through the initial observation. This is done in part i) of the above assumption, which guarantees that all classes will be observed (with probability one) after a sufficiently large L .
Besides, due to part ii) we can decompose the maximisation problem into two independent parts. The first part corresponds to the initial distribution and the other part to the semi-Markov kernel.

Despite the advantage of the arbitrary structure of the system, nothing guarantees that the complete structure of the system will be revealed even after a great number of observed trajectories. The censoring time m plays an important role on the transitions that can actually be observed. Naturally, we cannot observe transitions from a state i to a state j when the minimum time of doing so, exceeds m, even if j is accessible from i under the true value of the parameter. Consequently, our target is to estimate the part of the semi-Markov kernel that is accessible until time m. By assuming that $S_0 = 0$, this part corresponds to $\boldsymbol{q}_m$. We prove in the sequel (see Theorem 4.4) that the MLE of the parameter $\boldsymbol{q}_m$ is strongly consistent, asymptotically normal and asymptotically efficient.

Additionally, this type of estimation allows to measure the concentration of the sojourn time distribution for each $i \in E$ on the finite set $\{1, 2, \ldots, m\}$. This can be achieved by the MLE of the distribution function of the sojourn time in each state at time m. The subsequent asymptotic normality of the MLE (see Corollary 4.5) enables us to construct asymptotic confidence intervals in order to measure the error of this estimation.

In this Section, we indicate how we can treat generally the case where under the true value of the semi-Markov kernel not necessarily all the elements are strictly positive. To this end, we denote by

$$\mathbb{D}_{i|m}^* := \{(j, k) \in E \times T_m^+ : q_{ij}(k) > 0\} \quad \text{and} \quad d_{i|m}^* := |\mathbb{D}_{i|m}^*|.$$

Moreover, the part of $\boldsymbol{q}_m$ that corresponds to the strictly positive parameters $q_{ij}(k)$ is denoted by $\boldsymbol{q}_m^+$. With the same reasoning the vector $\boldsymbol{\eta}_m^+$ denotes the probabilities of the MC $(\mathbf{Z}, \mathbf{U})$ that are strictly positive. Additionally, $\boldsymbol{\eta}_m^*$ designates $\boldsymbol{\eta}_m^+$ in a minimal representation by excluding for each (i, u) the parameter $p_{i,u;i}$.

4.1.   *Construction of the MLE*

By following the lines of the previous Section we use the reparameterisation of the SMM by the natural parametric space of the MC $(\mathbf{Z}, \mathbf{U})$ censored at time m. Let $\boldsymbol{N}_{i,u}^{B,l}$ and $\boldsymbol{N}_{i,u;j}^{B,l}$ be the l-th realisations of the counting processes $\boldsymbol{N}_{i,u}^B$ and $\boldsymbol{N}_{i,u;j}^B$ respectively, defined in the previous Section by (7) and (6). Moreover, we define the counting processes

$$\boldsymbol{N}_{i,u}^B(L) := \sum_{l=1}^{L} \boldsymbol{N}_{i,u}^{B,l}, \quad \text{where } N_{i,u}^B(m; L) = \sum_{l=1}^{L} N_{i,u}^{B,l}(m), \tag{29}$$

and

$$\boldsymbol{N}_{i,u;j}^B(L) := \sum_{l=1}^{L} \boldsymbol{N}_{i,u;j}^{B,l}, \quad \text{where } N_{i,u}^B(j, m; L) = \sum_{l=1}^{L} N_{i,u}^{B,l}(j, m). \tag{30}$$

The r.v. $N_{i,u}^B(m; L)$ indicates the number of visits at state $(i, u) \in E \times \mathbb{N}$, until time $m \in \mathbb{N}^*$ for the L trajectories, by neglecting the last visited state (Z_m, U_m). Besides, the r.v. $\boldsymbol{N}_{i,u}^B(j, m; L)$ indicates the number of possible transitions of the type (i, u), (j, k) (that is, $(j, k) = (j, 0)$, $j \neq i$, or $(j, k) = (i, u + 1)$), until time m for the L trajectories.

This type of estimation enables us to estimate the initial distribution $\alpha(\cdot)$ of the SMC $\mathbf{Z}$. To this end, we define for each $i \in E$ the counting process $\boldsymbol{N}_i^\alpha := \{N_i^\alpha(L)\}_{L \in \mathbb{N}^*}$ of the times that i appeared as initial state in the L trajectories, where

$$N_i^\alpha(L) = \sum_{l=1}^{L} \mathbb{1}_{\{Z_0^{(l)} = i\}}. \tag{31}$$

The following proposition extends the results of Proposition 3.1 for MLE from 1 to L trajectories of a SMC $\mathbf{Z}$ censored at time m. Moreover, it is not necessary to assume that $\mathbf{Z}$ is positive recurrent since by the independence of the copies of $\mathbf{Z}$ the initial states can belong to different communication classes.

Proposition 4.1: *Let $\mathbf{Z}$ be a SMC defined on $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{\alpha}, \boldsymbol{q}})$ (nonparametric model), with state space $E = \{1, 2, \ldots, s\}$. Under condition (A2), the MLE $(\widehat{\boldsymbol{\alpha}}(L), \widehat{\boldsymbol{q}}_m(m; L))$ that corresponds to L trajectories $(\mathbf{Z}_0^m)_i$ for $i \in T_L^+$ from L independent copies of a SMC $\mathbf{Z}$ censored at time m, is given by its components, for $(i, j) \in E_*^2$, $k \in T_m^+$, as follows*

$$\widehat{q}_{ij}(k, m; L) = \Phi_{ijk}(\widehat{\boldsymbol{\eta}}_m(m; L)) = \widehat{p}_{i,k-1;j}(m; L) \widehat{H}_i(k - 1; L), \tag{32}$$

$$\widehat{\alpha}_i(L) = \frac{N_i^\alpha(L)}{L}, \tag{33}$$

where

$$\widehat{H}_i(k-1; L) = \prod_{u=0}^{k-2} \widehat{p}_{i,u;i}(m; L)$$

and the MLE $\widehat{\boldsymbol{\eta}}_m(m; L)$ *of the parameter* $\boldsymbol{\eta}_m$ *is given by*

$$\widehat{\boldsymbol{\eta}}_m(m; L) = \left(\widehat{p}_{i,u;j}(m; L) \right) = \left(\frac{N_{i,u}^B(j, m; L)}{N_{i,u}^B(m; L)} \right). \tag{34}$$

If $N_{i,u}^B(m; L) = 0$, *then by convention* $0/0 = 0$.

Proof: The likelihood function $\mathcal{L}_L(\boldsymbol{\alpha}, \boldsymbol{q})$ for L independent observations $(\mathbf{Z}_0^m)_i$, $i \in T_L^+$, is given by

$$\mathcal{L}_L(\boldsymbol{\alpha}, \boldsymbol{q}) = \prod_{l=1}^{L} \alpha_{J_0^{(l)}} \left(\prod_{k=1}^{N^{(l)}(m)} q_{J_{k-1}^{(l)} J_k^{(l)}}(X_k^{(l)}) \right) \bar{H}_{J_{N^{(l)}(m)}}(U_m^{(l)}).$$

By (31) and (30), the likelihood function $\mathcal{L}_L(\boldsymbol{\alpha}, \boldsymbol{q})$ is given as a function of $(\boldsymbol{\alpha}, \boldsymbol{\eta}_m)$ by

$$\mathcal{L}_L(\boldsymbol{\alpha}, \boldsymbol{\eta}_m) = \left(\prod_{i \in E} \alpha_i^{N_i^\alpha(L)} \right) \prod_{i,j \in E} \prod_{l \in T_L^+} \prod_{u \in T_m} p_{i,u;j}^{N_{i,u}^{B,l}(j,m)}.$$

Consequently, we obtain the following log-likelihood function

$$\ell_L(\boldsymbol{\alpha}, \boldsymbol{\eta}_m) = \sum_{i \in E} N_i^\alpha(L) \log \alpha_i + \sum_{i,j \in E} \sum_{l \in T_m^+} \sum_{u \in T_m} N_{i,u}^{B,l}(j, m) \log p_{i,u;j}. \tag{35}$$

The dependence relations that we have to take into consideration in a nonparametric framework are given by (16) for $\boldsymbol{\eta}_m$ and similarly for the initial vector $\boldsymbol{\alpha}$. From these dependence relations, condition (A2), (29) and by maximising the log-likelihood function (35) with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}_m$, we obtain (33) and (34). From the invariance principle of the MLE and Relation (12) we infer (32).

$$\square$$

4.2. *Asymptotic properties of the MLE*

In this subsection, we study some asymptotic properties of the MLE of the parameter $(\boldsymbol{\alpha}, \boldsymbol{q}_m)$. In particular, we prove strong consistency, asymptotic normality and asymptotic efficiency of the MLE. These properties will be deduced from the corresponding properties of the MLE of the parameter $(\boldsymbol{\alpha}, \boldsymbol{\eta}_m)$, which we present in the following lemma.

Remark 1:   Under condition (A2), the MLE $\widehat{\boldsymbol{\alpha}}(L)$ is strongly consistent, asymptotically normal and asymptotically efficient. Strong consistency is a direct consequence of the strong law of large numbers, applied to the sequence of i.i.d. random vectors $\{(\mathbb{1}_{\{Z_0^{(l)}=i\}} : i \in E)\}_{l \in \mathbb{N}^*}$. Asymptotic normality follows from the CLT for i.i.d. random vectors, that

is, the vector $\boldsymbol{\zeta}(L) := \sqrt{L}(\widehat{\boldsymbol{\alpha}}(L) - \boldsymbol{\alpha})$ converges $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}$-weakly, when $L \to \infty$, to $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \Gamma^{(L)})$, where $\Gamma_{ij}^{(L)} := \mathsf{Cov}_{\boldsymbol{\alpha},\boldsymbol{\eta}}(\zeta_i, \zeta_j) = \alpha_i(\delta_{ij} - \alpha_j)$. Asymptotic efficiency also follows from standard arguments for the MLE, in the i.i.d. framework.

From the above remark, and since the MLE $\widehat{\boldsymbol{\alpha}}(L)$ and the MLE $\widehat{\boldsymbol{\eta}}_m^+(m; L)$ are asymptotically independent, we will ignore in the rest the parameter $\boldsymbol{\alpha}$ from the estimation, and we will focus our attention on the parameter $\boldsymbol{\eta}_m$ (or $\boldsymbol{\eta}_m^+$).

The part ii) of the lemma is an adaptation of the theorem of Anderson and Goodman (1957) to our case. For the rest, we denote by $d_{i,u}$ the number of different j such that $p_{i,u;j}$ is strictly positive.

Lemma 4.2:   *Under condition (A2),*
i) the MLE $\widehat{\boldsymbol{\eta}}_m(m; L)$ of the parameter $\boldsymbol{\eta}_m$, given by (34), is strongly consistent, that is, as $L \to \infty$, we have $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}$-a.s., that $\widehat{\boldsymbol{\eta}}_m(m; L) \to \boldsymbol{\eta}_m$,
ii) the sequence of random vectors $\boldsymbol{\zeta}_m(L) := \sqrt{L}(\widehat{\boldsymbol{\eta}}_m^+(m; L) - \boldsymbol{\eta}_m^+)$, converges $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}$-weekly, as $L \to \infty$, to a random vector $\boldsymbol{\zeta}_m$, where $\boldsymbol{\zeta}_m \sim \mathcal{N}(\boldsymbol{0}, \Gamma_m^{(L)})$ and $\Gamma_m^{(L)}$ is a covariance matrix given in a block diagonal form as follows

$$\Gamma_m^{(L)} = \mathsf{diag}\{\frac{1}{\pi_{i,u|m}}\Lambda_{iu} : (i, u) \in E \times T_m\}, \tag{36}$$

where for each $(i, u) \in E \times T_m$, the matrix Λ_{iu} of dimension $d_{i,u} \times d_{i,u}$ is given by (23) and $\pi_{i,u|m}$ is given by

$$\pi_{i,u|m} := \mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}N_{i,u}^B(m) = \sum_{r \in E}\sum_{n \in T_m} \alpha_r p_{r,0;i,u}^{(n)}, \tag{37}$$

where

$$p_{r,0;i,u}^{(n)} = \mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}\big((Z_n, U_n) = (i, u)|(Z_0, U_0) = (r, 0)\big),$$

iii) the MLE $\widehat{\boldsymbol{\eta}}_m^*(m; L)$ of the parameter $\boldsymbol{\eta}_m^*$ is asymptotically efficient.

Proof: i) Notice by the form of $\widehat{\boldsymbol{\eta}}_m(m; L)$, that if we divide each term of the fraction with L, and apply the strong law of large numbers, we have as $L \to \infty$,

$$\widehat{p}_{i,u;j}(m; L) \xrightarrow{\mathbb{P}_{\boldsymbol{\alpha},\eta}\text{-a.s.}} \frac{\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}N_{i,u}^{B,l}(j, m)}{\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}N_{i,u}^{B,l}(m)}, \tag{38}$$

since both counting processes are sequences of i.i.d. bounded (by m) r.v. .
By (6), we have also that

$$\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}N_{i,u}^{B,l}(j, m) = p_{i,u;j} \sum_{n=0}^{m-1} \mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}\left((Z_n^{(l)}, U_n^{(l)}) = (i, u)\right) = p_{i,u;j}\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}}N_{i,u}^{B,l}(m). \tag{39}$$

By inserting (39) in (38), we obtain the result.
ii) Anderson and Goodman (1957) proved the asymptotic normality of the MLE of a finite MC even in the inhomogeneous case. At first, they considered known the vector of observed frequencies for the initial states in L trajectories. Next, they proved the

asymptotic results when this vector is multinomially distributed. The steps of their proof can be applied to the case of L trajectories corresponding to L independent copies of the MC $(\mathbf{Z}, \mathbf{U})$ observed until time m, since they are based on the multidimensional CLT for i.i.d. random vectors and not on the structure of $(\mathbf{Z}, \mathbf{U})$. Nevertheless, in order to apply them in our case, it only suffices to note that the support of the initial distribution reduces to the states $\{(i, 0)\}_{i \in E}$.

iii) Since we are in an iid setting, in order to compute the FIM, it is enough to take the log-likelihood contribution of a single term, corresponding here to the vector $\mathbf{Z}_0^m$. Therefore, the FIM is the matrix $I(\boldsymbol{\eta}_m^*)$, defined by (27). We have directly that

$$\frac{\partial \ell_m(\boldsymbol{\eta}_m)}{\partial p_{i_1,u_1;j_1}} = \frac{N_{i_1,u_1}^B(j_1, m)}{p_{i_1,u_1;j_1}} - \frac{N_{i_1,u_1}^B(i_1, m)}{p_{i_1,u_1;i_1}},$$

and also that

$$\frac{\partial^2 \ell_m(\boldsymbol{\eta}_m)}{\partial p_{i_1,u_1;j_1} \partial p_{i_2,u_2;j_2}} = \begin{cases} -\dfrac{N_{i_1,u_1}^B(j_1, m)}{p_{i_1,u_1;j_1}^2} - \dfrac{N_{i_1,u_1}^B(i_1, m)}{p_{i_1,u_1;i_1}^2}, & \text{if } (i_1, u_1, j_1) = (i_2, u_2, j_2), \\[4mm] -\dfrac{N_{i_1,u_1}^B(i_1, m)}{p_{i_1,u_1;i_1}^2}, & \text{if } (i_1, u_1) = (i_2, u_2), \ j_1 \neq j_2, \\[4mm] 0, & \text{if } (i_1, u_1) \neq (i_2, u_2). \end{cases} \tag{40}$$

By taking the negative expectations of the terms (40) and by (39) and (37), we deduce that

$$I(\boldsymbol{\eta}_m^*) = \mathsf{diag}\{I_{i,u}(\boldsymbol{\eta}_m^*); \ (i, u) \in E \times T_m\}, \tag{41}$$

where

$$I_{i,u}(\boldsymbol{\eta}_m^*) := \pi_{i,u|m} \big(p_{i,u;i}^{-1} + \delta_{j_1 j_2} p_{i,u;j_1}^{-1} \big)_{j_1,j_2}. \tag{42}$$

By (36), the asymptotic covariance matrix $\Gamma_m^{(L)}$ of the MLE, considered in a minimal representation (by excluding the lines and columns that correspond to $p_{i,u;i}$), is the inverse of the FIM $I(\boldsymbol{\eta}_m^*)$ given by (41) and (42), iff the matrix Λ_{iu} given by (23), is the inverse of $\pi_{i,u|m}^{-1} I_{i,u}(\boldsymbol{\eta}_m^*)$ given by (42). This can be verified with elementary calculus. $\square$

The following definition introduces the Markov renewal function that plays a very important role in the theory of Markov renewal chains. This definition was given by Barbu and Limnios (2008) in a different context, in the study of one trajectory.

Definition 4.3: The matrix function $\Psi = \big(\Psi(m), \ m \in \mathbb{N}\big) \in \mathcal{M}_E(\mathbb{N})$, where

$$\Psi_{ij}(m) := \mathbb{E}(N_{j,0}^B(m + 1) \mid Z_0 = i)$$

is called the discrete time Markov renewal function. We also denote by

$$\Psi_i(m) := \mathbb{E} N_{i,0}^B(m + 1). \tag{43}$$

With the following theorem we adapt the asymptotic results of the MLE for the MCs to the case of SMCs by making also explicit the covariance matrix form of the asymptotically

normal distribution of the MLE. The trick consists in showing that this theorem can be proved by Lemma 4.2 in a similar way that Proposition 3.3 can be proved by Lemma 3.2.

Theorem 4.4: *Under condition (A2),*
i) the MLE $\widehat{\boldsymbol{q}}_m(m; L)$ of the parameter $\boldsymbol{q}_m$, given by (32), is strongly consistent,
ii) the sequence of random vectors $\boldsymbol{\xi}_m(L) := \sqrt{L}(\widehat{\boldsymbol{q}}_m^+(m; L) - \boldsymbol{q}_m^+)$, converges $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}$-weakly,
when $L \to \infty$, to a random vector $\boldsymbol{\xi}_m$, where $\boldsymbol{\xi}_m \sim \mathcal{N}(\boldsymbol{0}, \Delta_m^{(L)})$ and $\Delta_m^{(L)}$ is a covariance matrix given in a block diagonal form as follows

$$\Delta_m^{(L)} = \mathsf{diag}\left\{\frac{1}{\Psi_i(m-1)}K_i : i \in E\right\}, \tag{44}$$

where for each $i \in E$ the matrix K_i of dimension $d_{i|m}^* \times d_{i|m}^*$ is given by (25) and $\Psi_i(m-1)$ is given by (43) and assumed to be strictly positive.
iii) the MLE $\widehat{\boldsymbol{q}}_m^+(m; L)$ of the parameter $\boldsymbol{q}_m^+$, is asymptotically efficient.

Proof: i) By Lemma 4.2, we have that the MLE $\widehat{\boldsymbol{\eta}}_m(m; L)$ of the parameter $\boldsymbol{\eta}_m$, given by (34), is strongly consistent. Let Φ_m be the restriction of Φ, given by (12), to the domain that corresponds to the vector parameter $\boldsymbol{\eta}_m$ in a minimal representation. By applying the continuous mapping theorem for the function Φ_m we obtain the desired result.
ii) Notice by (37) and (43) that

$$\pi_{i,0|m} = \Psi_i(m-1). \tag{45}$$

Moreover,

$$\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}} N_{i,u}^B(m) = \sum_{n=0}^{m-1} \mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}(Z_n = i, U_n = u) = \sum_{n=u}^{m-1} \mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}(Z_{n-u} = i, U_{n-u} = 0) \times$$

$$\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}(X_{N(n-u)+1} > u | Z_{n-u} = i, U_{n-u} = 0)$$

$$= \bar{H}_i(u) \sum_{n=0}^{m-1} \mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{\eta}}(Z_n = i, U_n = 0) = \bar{H}_i(u)\, \mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\eta}} N_{i,0}^B(m).$$

This implies by (37) and (45) that

$$\pi_{i,u|m} = \bar{H}_i(u)\Psi_i(m-1). \tag{46}$$

We deduce that the ratio of $\bar{H}_i(u)$ with $\pi_{i,u|m}$ does not depend on u. By inserting $(m, \pi_{i,u|m}, (\Psi_i(m-1))^{-1}, L, \Gamma_m^{(L)})$ in the place of $(\tilde{n}, \pi_{i,u}, \mu_{ii}, m, \Gamma_{\tilde{n}})$ in the statement of Lemma 3.2, the steps of the proof needed for the part ii) of this theorem can be deduced by Lemma 4.2, in the same way that Proposition 3.3 can be deduced by Lemma 3.2. iii) The proof is a direct consequence of the part iii) of Lemma 4.2, since the asymptotic efficiency of the MLE is parameterisation independent.   □

Remark 2: i) The term $\Psi_i(m-1)$ that appears in the expression (44) and which is defined by (43), can be expressed as a function of the semi-Markov kernel q. If $q^{(n)}$

represents the n-th convolution power of q, (see Barbu and Limnios (2008), pages 48-49), then the following relation holds

$$\Psi_i(m-1) = \sum_{r \in E} \sum_{n,l \in T_m} \alpha_r q_{ri}^{(n)}(l). \tag{47}$$

Consequently, by (47) and (46) we obtain an alternative form of (37),

$$\pi_{i,u|m} = \bar{H}_i(u) \sum_{r \in E} \sum_{n,l \in T_m} \alpha_r q_{ri}^{(n)}(l).$$

ii) The term $\Psi_i(m-1)$ by (43) is strictly positive iff there exists a state j where $\alpha_j > 0$ and the state $i \in E$ is accessible from j until time m. If the support of the initial distribution is E, this is always true. Otherwise, in Theorem 4.4 we have to consider only the states that satisfy the above property.

In order to measure the concentration of the sojourn time distribution at each state $i \in E$ inside the set $\{1, 2, \ldots, m\}$, we consider the MLE of the distribution function $H_i(m)$ at time m. Since

$$H_i(m) = \sum_{(j,k) \in \mathbb{D}_{i|m}^*} q_{ij}(k), \tag{48}$$

the MLE is given by

$$\widehat{H}_i(m; L) = \sum_{(j,k) \in \mathbb{D}_{i|m}^*} \widehat{q}_{ij}(k, m; L).$$

This additive form is convenient in order to deduce easily the asymptotic results of strong consistency and asymptotic normality of the MLE. The proof is deferred to Appendix A.

Corollary 4.5:   *Under condition (A2) and the probability measure $\mathbb{P}_{\boldsymbol{\alpha},\boldsymbol{q}}$,*
i) the MLE $\widehat{\boldsymbol{H}}(m; L) := \big(\widehat{H}_i(m; L)\big)_{i \in E}$ of the vector $\boldsymbol{H}(m) := \big(H_i(m) : i \in E\big)$ is strongly consistent, and
ii) the sequence of random vectors $\boldsymbol{\zeta}_m(L) := \sqrt{L}(\widehat{\boldsymbol{H}}(m; L) - \boldsymbol{H}(m))$ converges weakly, as $L \to \infty$, to a random vector $\boldsymbol{\zeta}_m$, where $\boldsymbol{\zeta}_m \sim \mathcal{N}(\boldsymbol{0}, Z_m^{(L)})$ and $Z_m^{(L)}$ is a diagonal covariance matrix of dimension $s \times s$ given by

$$Z_m^{(L)} = \mathsf{diag}\left\{ \frac{H_i(m)\bar{H}_i(m)}{\Psi_i(m-1)} : i \in E \right\}. \tag{49}$$

Acknowledgements

References

Anderson, P.K., Borgan, O., Gill, R.D., and Keiding, N., *Statistical models based on counting processes*, Springer, New York (1993).

Anderson, T., and Goodman, L. (1957), "Statistical inference about Markov Chains," *Ann. Math. Stat.*, 28, 89–110.

Anselone, P. (1960), "Ergodic theory for discrete semi-Markov chains," *Duke Math. J.*, 27, 33–40.

Barbu, V., and Limnios, N. (2006), "Empirical estimation for discrete time semi-Markov processes with applications in reliability," *Journal of Nonparametric Statistics*, 18, 483–498.

Barbu, V., and Limnios, N., *Semi-Markov chains and hidden semi-Markov models toward applications: Their use in reliability and DNA analysis (Lecture notes in statistics)*, Vol. 191, Springer (2008).

Barbu, V., Boussemart, M., and Limnios, N. (2004), "Discrete time semi-Markov model for reliability and survival analysis," *Communications in Statistics-Theory and Methods*, 33, 2833–2868.

Burge, C. (1997), "Identification of Genes in Human Genomic DNA," Technical report.

Chryssaphinou, O., Karaliopoulou, M., and Limnios, N. (2008), "On Discrete Time Semi-Markov Chains and Applications in Words occurrences," *Communications in Statistics-Theory and Methods*, 37, 1306–1322.

Commenges, D. (1999), "Multi-state Models in Epidemiology," *Lifetime Data Analysis*, 5, 315–327.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G., *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University press, Cambridge (1998).

Gerontidis, I. (1994), "Semi-Markov replacement chains," *Adv. in Appl. Probab.*, 26, 728–755.

Howard, R., *Dynamic Probabilistic Systems*, Vol. 2, Wiley, New York (1971).

Kalbfleisch, J., and Prentice, R., *The statistical analysis of failure time data*, 2 ed., Wiley, New Jersey (2002).

Limnios, N., and Oprişan, G., *Semi-Markov Processes and Reliability*, Birkhäuser, Boston (2001).

Malefaki, S., Trevezas, S., and Limnios, N. (2010), "An EM and a Stochastic Version of the EM Algorithm for Nonparametric Hidden Semi-Markov Models," *Communications in Statistics - Simulation and Computation*, 39, 240–261.

Mode, C., and Pickens, G. (1998), "Computational Methods for Renewal Theory and Semi-Markov Processes with Illustrative Examples," *The American Statistician*, 42, 143–152.

Mode, C., and Sleeman, C., *Stochastic processes in epidemiology*, World Scientific, New Jersey (2000).

Stenflo, O. (1996), "Iterated Function Systems Controlled by a Semi-Markov Chain," *Theory Stoch. Process.*, 18, 305–313.

Van Der Vaart, A., *Asymptotic Statistics*, Cambridge University Press, New York (1998).

Appendix A. Proof of Results

Proof of Lemma 3.2:

We define the random vector

$$\boldsymbol{\zeta}(m) = (\zeta_{iuj}(m);\ i,j \in E,\ u \in T_{\tilde{n}}),$$

where

$$\zeta_{iuj}(m) = N_{i,u}^B(j; m) - N_{i,u}^B(m) p_{i,u;j}. \tag{A1}$$

Notice that if we denote by

$$R(m) := \mathsf{diag}\{r_{iuj}(m);\ i,j \in E,\ u \in T_{\tilde{n}}\},\ \text{where}\ r_{iuj}(m) = \frac{m}{N_{i,u}^B(m)}, \tag{A2}$$

then it is easy to see that

$$\boldsymbol{\xi}_{\tilde{n}}(m) = m^{-1/2} R(m) \boldsymbol{\zeta}(m). \tag{A3}$$

Let also

$$R := \mathsf{diag}\{r_{iuj};\ i,j \in E,\ u \in T_{\tilde{n}}\},\ \text{where}\ r_{iuj} = \frac{1}{\pi_{i,u}}. \tag{A4}$$

Recall that the process $N_{i,u}^B(m)$ is a counting process of the number of visits to the state (i, u) until time m of the irreducible and positive recurrent Markov chain $(\mathbf{Z}, \mathbf{U})$ (as a consequence of condition (A1)). From the ergodic theorem for positive recurrent MCs we have

$$\frac{N_{i,u}^B(m)}{m} \xrightarrow{\mathbb{P}_\eta\text{-a.s.}} \pi_{i,u} > 0. \tag{A5}$$

Therefore by Relation (A5) and the expressions (A2) and (A4) we have

$$R(m) \xrightarrow{\mathbb{P}_\eta\text{-a.s.}} R. \tag{A6}$$

If we establish that $m^{-1/2}\boldsymbol{\zeta}(m)$ converges in distribution to a random vector $\boldsymbol{\zeta}$, then in virtue of the representation of $\boldsymbol{\xi}_{\tilde{n}}(m)$ as in (A3) and the limiting result (A6), $\boldsymbol{\xi}_{\tilde{n}}(m)$ will converge in distribution to the random vector $R\boldsymbol{\zeta}$. Note that by Relation (A1)

$$\sum_{j \in E} \zeta_{iuj}(m) = \sum_{j \in E} N_{i,u}^B(j; m) - N_{i,u}^B(m) \sum_{j \in E} p_{i,u;j}$$

$$= N_{i,u}^B(m) - N_{i,u}^B(m) = 0. \tag{A7}$$

This implies that there is one dependence relation for the components of $\boldsymbol{\zeta}(m)$ for each $(i, u) \in E \times T_{\tilde{n}}$. Consequently, by excluding the the r.v. $\zeta_{iui}(m)$ for each $(i, u) \in E \times T_{\tilde{n}}$, we obtain a new vector $\boldsymbol{\zeta}^*(m)$ that represents $\boldsymbol{\zeta}(m)$ in a minimal representation. This

allows us to apply the Cramér-Wold device, where we need to show that for an arbitrary $\mathbf{t} \in \mathbb{R}^{(s^2-s)\tilde{n}}$, the r.v. $m^{-1/2}\mathbf{t}\boldsymbol{\zeta}^*(m)$ converges in distribution to the r.v. $\mathbf{t}\boldsymbol{\zeta}^*$. For this purpose, we will represent the sequence of r.v. $\{\mathbf{t}\boldsymbol{\zeta}^*(m)\}_{m\geq 1}$ as the sequence of partial sums of another sequence $\{Y_n\}_{n\geq 1}$ that has the property to be a martingale difference adapted to the natural filtration of the underlying probability space $(\Omega, \mathscr{A}, \mathbb{P}_{\boldsymbol{\eta}})$ of the Markov chain $(\mathbf{Z}, \mathbf{U})$. Let $\mathscr{F}_n := \sigma\{(Z_k, U_k)_{0\leq k\leq n}\}$, that is, the σ-algebra generated by the random vector $(Z_k, U_k)_{0\leq k\leq n}$. For each $n \in \mathbb{N}^*$ we define the random variable

$$Y_n := \mathbb{1}_{\{U_{n-1}\leq \tilde{n}-1, U_n=0\}}[t_{Z_{n-1}U_{n-1}Z_n} - \mathbb{E}_{\boldsymbol{\eta}}(\mathbb{1}_{\{U_n=0\}}t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}U_{n-1})]. \qquad (A8)$$

It is straightforward that the sequence of r.v. $(Y_n)_{n\in\mathbb{N}^*}$ is adapted to the filtration $(\mathscr{F}_n)_{n\in\mathbb{N}^*}$ of the underlying probability space $(\Omega, \mathscr{A}, \mathbb{P}_{\boldsymbol{\eta}})$. Notice also that by construction $\mathbb{E}_{\boldsymbol{\eta}}|Y_n| < \infty$ and $\mathbb{E}_{\boldsymbol{\eta}}(Y_n \mid \mathscr{F}_{n-1}) = 0$. Therefore, $(Y_n)_{n\in\mathbb{N}^*}$ is an $(\mathscr{F}_n)$-martingale difference. Since

$$\mathbb{E}_{\boldsymbol{\eta}}\big(\mathbb{1}_{\{U_n=0\}}t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}U_{n-1}\big) = \sum_{j\in E_{-i}} p_{Z_{n-1},U_{n-1};j}\, t_{Z_{n-1}U_{n-1}j}, \qquad (A9)$$

by combining (A8) and (A9) we get

$$\sum_{n=1}^m Y_n = \sum_{(i,j)\in E_*^2} \sum_{u\in T_{\tilde{n}}} \sum_{n=1}^m \big[t_{iuj}\mathbb{1}_{\{Z_{n-1}=i,U_{n-1}=u\}}(\mathbb{1}_{\{Z_n=j\}} - p_{i,u;j})\big]$$
$$= \sum_{i,j\in E_*^2} \sum_{u\in T_{\tilde{n}}} t_{iuj}\big(N_{i,u}(j;m) - N_{i,u}(m)p_{i,u;j}\big) = \mathbf{t}\boldsymbol{\zeta}^*(m). \qquad (A10)$$

In order to apply the CLT for Martingales note that

$$\mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mid \mathscr{F}_{n-1}) = \mathbb{1}_{\{U_{n-1}\leq \tilde{n}-1\}}\mathsf{Var}_{\boldsymbol{\eta}}\big(\mathbb{1}_{\{U_n=0\}}t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}U_{n-1}\big),$$

and therefore

$$\sum_{n=1}^m \mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mid \mathscr{F}_{n-1}) = \sum_{n=1}^m \mathbb{1}_{\{U_{n-1}\leq \tilde{n}-1\}}\mathsf{Var}_{\boldsymbol{\eta}}\big(\mathbb{1}_{\{U_n=0\}}t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1},U_{n-1}\big)$$
$$= \sum_{i\in E} \sum_{u\in T_{\tilde{n}}} N_{i,u}(m)\mathsf{Var}_{\boldsymbol{\eta}}\big(\mathbb{1}_{\{U_n=0\}}t_{iuZ_n} \mid Z_{n-1}=i,U_{n-1}=u\big).$$
$$\qquad (A11)$$

Since

$$\mathbb{E}_{\boldsymbol{\eta}}\big(\mathbb{1}_{\{U_n=0\}}t_{iuZ_n}^2 \mid Z_{n-1}=i, U_{n-1}=u\big) = \sum_{j\in E_{-i}} t_{iuj}^2 p_{i,u;j}, \qquad (A12)$$

and

$$\mathbb{E}_{\boldsymbol{\eta}}^2\big(\mathbb{1}_{\{U_n=0\}}t_{iuZ_n} \mid Z_{n-1}=i, U_{n-1}=u\big) = \Big(\sum_{j\neq i}t_{iuj}p_{i,u;j}\Big)\Big(\sum_{l\neq i}t_{iul}p_{i,u;l}\Big),$$

$$= \sum_{j\neq i}t_{iuj}p_{i,u;j}\Big(\sum_{l\neq i}t_{iul}p_{i,u;l}\Big), \qquad (A13)$$

by taking the difference of (A12) and (A13) to form the variance that appears in Relation (A11) we get

$$\sum_{n=1}^{m}\mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mid \mathscr{F}_{n-1}) = \sum_{i\in E}\sum_{u\in T_{\tilde{n}}}N_{i,u}(m)\sum_{j\neq i}t_{iuj}p_{i,u;j}\Big(t_{iuj}-\sum_{l\neq i}t_{iul}p_{i,u;l}\Big)$$

$$= \sum_{i\in E}\sum_{u\in T_{\tilde{n}}}N_{i,u}(m)\sum_{\substack{j,l\in E\\j,l\neq i}}t_{iuj}p_{i,u;j}(\delta_{jl}t_{iuj}-t_{iul}p_{i,u;l}). \qquad (A14)$$

We denote by

$$\lambda_{i,u}(\mathbf{t},\boldsymbol{\eta}) := \sum_{\substack{j,l\in E\\j,l\neq i}}t_{iuj}p_{i,u;j}(\delta_{jl}t_{iuj}-t_{iul}p_{i,u;l}). \qquad (A15)$$

Let also $\mathbf{t}_{iu} := \big(t_{iuj}; j\in E_{-i}\big)$, and $\Lambda_{iu}^* \in \mathbb{R}^{(s-1)\times(s-1)}$ the matrix that results from Λ_{iu}, which appears in Relation (22) if we extract the i-row and the i-column. Then, the coefficients $\lambda_{i,u}(\mathbf{t},\boldsymbol{\eta})$ given by (A15) can be expressed as follows

$$\lambda_{i,u}(\mathbf{t},\boldsymbol{\eta}) = \mathbf{t}_{iu}\Lambda_{iu}^*\mathbf{t}_{iu}^\top.$$

It is easy to see that

$$\det(\Lambda_{iu}^*) = \prod_{j\in E_{-i}}p_{i,u;j}\big(1-\sum_{j\in E_{-i}}p_{i,u;j}\big) = \prod_{j\in E}p_{i,u;j}.$$

Therefore, since Λ_{iu}^* is a symmetric matrix and $p_{i,u;j} > 0$ we conclude that for each $(i,u)\in E\times T_{\tilde{n}}$,

$$\lambda_{i,u}(\mathbf{t},\boldsymbol{\eta}) > 0, \quad \text{iff} \quad \mathbf{t}_{iu}\neq\mathbf{0}. \qquad (A16)$$

By relations (A14) and (A15) we have that

$$\sum_{n=1}^{m}\mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mid \mathscr{F}_{n-1}) = \sum_{i\in E}\sum_{u\in T_{\tilde{n}}}\lambda_{i,u}(\mathbf{t},\boldsymbol{\eta})N_{i,u}(m).$$

By (A5) and the fact that the coefficients $\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta})$ are constant for fixed $\mathbf{t}$, we get that $\mathbb{P}_{\boldsymbol{\eta}}$-a.s.,

$$\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mid \mathscr{F}_{n-1}) = \sum_{i \in E} \sum_{u \in T_{\tilde{n}}} \pi_{i,u} \lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) := \sigma^2(\mathbf{t}, \boldsymbol{\eta}). \tag{A17}$$

Since for $(i, u) \in E \times T_{\tilde{n}}$, we have $\pi_{i,u} > 0$, then by (A16) we conclude that

$$\sigma^2(\mathbf{t}, \boldsymbol{\eta}) > 0, \quad \text{iff} \quad \mathbf{t} \neq \mathbf{0}. \tag{A18}$$

Additionally, for arbitrary $\epsilon > 0$

$$\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mathbb{1}_{\{|Y_n| > \epsilon \sqrt{m}\}}) = 0, \tag{A19}$$

since Y_m is bounded for any $m \in \mathbb{N}^*$. Since the limit expressed by equation (A17) is strictly positive by (A18) and also (A19) holds true, we can apply the CLT for martingales to conclude by (A10) that $\mathbb{P}_{\boldsymbol{\eta}}$-weakly,

$$m^{-1/2} \mathbf{t} \boldsymbol{\zeta}^*(m) \to \mathcal{N}(0, \sigma^2(\mathbf{t}, \boldsymbol{\eta})). \tag{A20}$$

Now, let $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, R^{-1}\Lambda)$ and the minimal representation $\boldsymbol{\zeta}^* \sim N(\mathbf{0}, (R^*)^{-1}\Lambda^*)$. Then, the r.v. $\mathbf{t}\boldsymbol{\zeta}^* \sim \mathcal{N}(0, \mathbf{t}(R^*)^{-1}\Lambda^* \mathbf{t}^\top)$. It is easy to check by (A17) that

$$\mathbf{t}(R^*)^{-1}\Lambda^* \mathbf{t}^\top = \sigma^2(\mathbf{t}, \boldsymbol{\eta}). \tag{A21}$$

By (A20), (A21) and the definition of $\boldsymbol{\zeta}^*$ we get that $\mathbb{P}_{\boldsymbol{\eta}}$-weakly,

$$m^{-1/2} \mathbf{t} \boldsymbol{\zeta}^*(m) \to \mathbf{t} \boldsymbol{\zeta}^*. \tag{A22}$$

Relation (A22) is valid for any $\mathbf{t} \in \mathbb{R}^{(s^2-s)\tilde{n}}$, and consequently by (A3), (A6) and Slutsky theorem we get that $\mathbb{P}_{\boldsymbol{\eta}}$-weakly,

$$\boldsymbol{\xi}_{\tilde{n}}^*(m) \to \mathcal{N}(\mathbf{0}, \Lambda R^\top).$$

The proof is complete for the minimal representation $\boldsymbol{\xi}_{\tilde{n}}^*(m)$, since the above covariance matrix coincides with $\Gamma_{\tilde{n}}^*$ that results from $\Gamma_{\tilde{n}}$ which is given by Relation (22). In order to obtain $\boldsymbol{\xi}_{\tilde{n}}(m)$ from the corresponding vector $\boldsymbol{\xi}_{\tilde{n}}^*(m)$ it is enough to add the components that correspond to the parameters $p_{i,u;i}$, for all the couples (i, u). Since every such component is a linear combination of the others with the different j, identical to the case of the components of a multinomial distribution we conclude that the result can be extended easily in order to obtain the form of $\Gamma_{\tilde{n}}$ given by (22).

$\square$

Proof of Proposition 3.3:
We denote by Φ_m the restriction of Φ, given by (12), to the domain that corresponds to the vector parameter $\boldsymbol{\eta}_m$ in a minimal representation. Notice that

$$\sqrt{m}(\widehat{\boldsymbol{q}}_{\tilde{n}}(m) - \boldsymbol{q}_{\tilde{n}}) = \sqrt{m} \left[\Phi_{\tilde{n}}(\widehat{\boldsymbol{\eta}}_{\tilde{n}}(m)) - \Phi_{\tilde{n}}(\boldsymbol{\eta}_{\tilde{n}}) \right]. \tag{A23}$$

Also, by Lemma 3.2 we have that $\mathbb{P}_{\boldsymbol{\eta}}-$weakly

$$\sqrt{m}\big(\widehat{\boldsymbol{\eta}}_{\tilde{n}}(m) - \boldsymbol{\eta}_{\tilde{n}}\big) \to \mathcal{N}(\mathbf{0}, \Gamma_{\tilde{n}}^*).$$

By Relation (A23) and delta method, since $\Phi_{\tilde{n}}$ is an analytic function and therefore the total derivative $\Phi_{\tilde{n}}'$ exists, we conclude that $\mathbb{P}_{\boldsymbol{\eta}}-$weakly

$$\sqrt{m}\big(\widehat{\boldsymbol{q}}_{\tilde{n}}(m) - \boldsymbol{q}_{\tilde{n}}\big) \to \mathcal{N}\big(\mathbf{0}, \Phi_{\tilde{n}}'\Gamma_{\tilde{n}}^*(\Phi_{\tilde{n}}')^{\top}\big).$$

The next step is to identify that the covariance matrix $\Phi_{\tilde{n}}'\Gamma_{\tilde{n}}^*(\Phi_{\tilde{n}}')^{\top}$ coincides with $\Delta_{\tilde{n}}$, which is given by (24). Firstly, it is easy to see that

$$\Phi_{\tilde{n}}' = \mathsf{diag}\big\{\Phi_{1;\tilde{n}}', \Phi_{2;\tilde{n}}', \cdots, \Phi_{s;\tilde{n}}'\big\},$$

where for $i \in E$,

$$\Phi_{i;\tilde{n}}' = \left(\frac{\partial \Phi_{ijk}}{\partial p_{i,l;r}}\right)_{(k,j)\in T_{\tilde{n}}^+ \times E_{-i},(l,r)\in T_{\tilde{n}} \times E_{-i}}. \tag{A24}$$

Notice here that we consider the functions Φ_{ijk} ordered with respect to the lexicographical order of (i, k, j). By Relation (12) we get

$$\frac{\partial \Phi_{ijk}}{\partial p_{i,l;r}} = \begin{cases} -\dfrac{\bar{H}_i(k-1)}{p_{i,l;i}} p_{i,k-1;j}, & \text{if } 0 \le l < k-1, \\ \bar{H}_i(k-1), & \text{if } l = k-1, \ r = j, \\ 0, & \text{if } k-1 < l < \tilde{n}-1. \end{cases} \tag{A25}$$

Now, let for $k \in T_{\tilde{n}}^+, l \in T_{\tilde{n}}$,

$$P_{kl}^i := \left(\frac{p_{i,k-1;j}}{p_{i,l;i}}\right)_{j,r \in E_{-i}}. \tag{A26}$$

By combining (A24), (A25) and (A26) we conclude that

$$(\Phi_{i;\tilde{n}}')_{kl} = \begin{cases} -\bar{H}_i(k-1)P_{kl}^i, & \text{if } 0 \le l < k-1, \\ \bar{H}_i(k-1)I_{s-1}, & \text{if } l = k-1, \\ O_{s-1}, & \text{if } k-1 < l < \tilde{n}-1. \end{cases} \tag{A27}$$

We have that

$$\Phi_{\tilde{n}}'\Gamma_{\tilde{n}}^*(\Phi_{\tilde{n}}')^{\top} = \Delta_{\tilde{n}}, \text{ iff } \Phi_{i;\tilde{n}}'\Lambda_i(\Phi_{i;\tilde{n}}')^{\top} = K_i, \text{ for each } i \in E.$$

Notice that

$$
\Phi_{i;\tilde{n}}^{'}\Lambda_i =
\begin{pmatrix}
\Lambda_{i0} & O_{s-1} & O_{s-1} & \dots & O_{s-1} \\
-\bar{H}_i(1)P_{20}^i\Lambda_{i0} & \Lambda_{i1} & O_{s-1} & \dots & O_{s-1} \\
-\bar{H}_i(2)P_{30}^i\Lambda_{i0} & -\frac{\bar{H}_i(2)}{\bar{H}_i(1)}P_{31}^i\Lambda_{i1} & \Lambda_{i2} & \dots & O_{s-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\bar{H}_i(s-1)P_{s0}^i\Lambda_{i0} & -\frac{\bar{H}_i(s-1)}{\bar{H}_i(1)}P_{s1}^i\Lambda_{i1} & -\frac{\bar{H}_i(s-1)}{\bar{H}_i(2)}P_{s2}^i\Lambda_{i2} & \dots & \Lambda_{i,s-1}
\end{pmatrix}
\tag{A28}
$$

We have

$$
P_{ku}^i\Lambda_{iu} = \big(p_{i,k-1;j}p_{i,u;r}\big)_{j,r\in E_{-i}},
$$

and therefore

$$
P_{ku}^i\Lambda_{iu}(P_{lu}^i)^\top = \frac{1-p_{i,u;i}}{p_{i,u;i}}\big(p_{i,k-1;j}p_{i,l;r}\big)_{j,r\in E_{-i}}.
$$

By equation (A27) and (A28) we get for $k < l+1$,

$$
\begin{aligned}
\big(\Phi_{i;\tilde{n}}^{'}\Lambda_i(\Phi_{i;\tilde{n}}^{'})^\top\big)_{kl} &= \bar{H}_i(k-1)\bar{H}_i(l)\left(\sum_{u=0}^{k-2}\frac{1-p_{i,u;i}}{p_{i,u;i}\bar{H}_i(u)}\right)\big(p_{i,k-1;j}p_{i,l;r}\big)_{j,r} \\
&\quad -\bar{H}_i(l)\big(p_{i,k-1;j}p_{i,l;r}\big)_{j,r} \\
&= \left\{\sum_{u=0}^{k-2}\left(\frac{1-p_{i,u;i}}{\bar{H}_i(u+1)}\right) - \frac{1}{\bar{H}_i(k-1)}\right\}\big(q_{ij}(k)q_{ir}(l+1)\big)_{j,r} \\
&= \left\{\sum_{u=0}^{k-2}\left(\frac{1}{\bar{H}_i(u+1)} - \frac{1}{\bar{H}_i(u)}\right) - \frac{1}{\bar{H}_i(k-1)}\right\}\big(-K_i\big)_{kl} \\
&= \big(K_i\big)_{kl}.
\end{aligned}
\tag{A29}
$$

Since the matrices $\Phi_{i;\tilde{n}}^{'}\Lambda_i(\Phi_{i;\tilde{n}}^{'})^\top$ and K_i are symmetric, their blocks coincide also for the case that $k > l+1$. With similar arguments as in (A29), the diagonal blocks ($k = l+1$) for the above matrices coincide.

$\square$

Proof of the part iii) of Theorem 3.4:
By (42), and since $\lim_{m\to\infty} m^{-1}\pi_{i,u|m} = \pi_{i,u}$ (ergodic theorem applied to the positive recurrent MC $(\mathbf{Z},\mathbf{U})$), we deduce that

$$
I_\infty(\boldsymbol{\eta}_{\tilde{n}}^*) = \mathsf{diag}\left\{\pi_{i,u}\big(p_{i,u;i}^{-1} + \delta_{j_1 j_2}p_{i,u;j_1}^{-1}\big)_{j_1,j_2}; \ (i,u)\in E\times T_{\tilde{n}}\right\},
$$

and its inverse $I_\infty^{-1}(\boldsymbol{\eta}_{\tilde{n}}^*)$ equals $\Gamma_{\tilde{n}}$ in a minimal representation.

Remark *A1*: The limiting FIM can be represented as

$$I_\infty(\boldsymbol{\eta}_{\tilde{n}}^*) := \left(\mathbb{E}_{\boldsymbol{\pi}} \left(\frac{-\partial^2 \log \mathbb{P}_{\boldsymbol{\eta}}(Z_1, U_1 \mid Z_0, U_0)}{\partial p_{i_1, u_1; j_1} \partial p_{i_2, u_2; j_2}} \right) \right)_{(i_1, u_1, j_1), (i_2, u_2, j_2)},$$

where the expectation is taken wrt to the stationary measure $\boldsymbol{\pi}$ of the MC $(\mathbf{Z}, \mathbf{U})$.

$\square$

Proof of Corollary 4.5: Let $R_m : \boldsymbol{q}_m^+ \mapsto R_m(\boldsymbol{q}_m^+) := \boldsymbol{H}(m)$, where each component $R_{i|m} := H_i(m)$ is given by (48). We conclude directly from the additive form of each component of R_m that it is continuous and the total derivative exists. Therefore, since by the part i) of Theorem 4.4, the MLE $\widehat{\boldsymbol{q}}_m(m; L)$ is strongly consistent, we can apply the continuous mapping theorem, and obtain the strong consistency of $\widehat{\boldsymbol{H}}(m; L)$. Likewise, by the part ii) of Theorem 4.4 and an application of delta method, the MLE $\widehat{\boldsymbol{H}}(m; L)$ is asymptotically normal with zero mean and diagonal covariance matrix of dimension $s \times s$, given by

$$Z_m^{(L)} = \mathsf{diag} \left\{ \frac{1}{\Psi_i(m-1)} R'_{i|m} K_i (R'_{i|m})^\top : i \in E \right\}. \tag{A30}$$

By (48) notice that $R'_{i|m} = \mathbf{1}_{d^*_{i|m}}$, and consequently, by the form of K_i given by (25),

$$R'_{i|m} K_i (R'_{i|m})^\top = \mathbf{1}_{d^*_{i|m}} K_i \mathbf{1}_{d^*_{i|m}}^\top = \sum_{(j,k),(l,u) \in \mathbb{D}^*_{i|m}} K_i((j,k),(l,u))$$

$$= \sum_{(j,k) \in \mathbb{D}^*_{i|m}} q_{ij}(k)(1 - q_{ij}(k)) - \sum_{\substack{(j,k),(l,u) \in \mathbb{D}^*_{i|m} \\ (j,k) \neq (l,u)}} q_{ij}(k) q_{il}(u)$$

$$= \sum_{(j,k) \in \mathbb{D}^*_{i|m}} q_{ij}(k) - \sum_{(j,k),(l,u) \in \mathbb{D}^*_{i|m}} q_{ij}(k) q_{il}(u)$$

$$= H_i(m) - H_i^2(m) = H_i(m) \bar{H}_i(m). \tag{A31}$$

By (A30) and (A31) we obtain the explicit form of $Z_m^{(L)}$ given by (49).