

# Maximum likelihood estimation for general hidden semi-Markov processes with backward recurrence time dependence

Samis Trevezas, Nikolaos Limnios

► **To cite this version:**

Samis Trevezas, Nikolaos Limnios. Maximum likelihood estimation for general hidden semi-Markov processes with backward recurrence time dependence. *Journal of Mathematical Sciences*, Springer Verlag (Germany), 2009, 163 (3), pp.262-274. <10.1007/s10958-009-9675-9>. <inria-00576524>

**HAL Id: inria-00576524**

**<https://hal.inria.fr/inria-00576524>**

Submitted on 14 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Maximum likelihood estimation for general hidden semi-Markov processes with backward recurrence time dependence

Samis TREVEZAS\* Nikolaos LIMNIOS

*Laboratoire de Mathématiques Appliquées de Compiègne*

*Université de Technologie de Compiègne*

---

## Abstract

This article concerns the study of the asymptotic properties of the maximum likelihood estimator (MLE) for the general hidden semi-Markov model (HSMM) with backward recurrence time dependence. By transforming the general HSMM into a general hidden Markov model, we prove that under some regularity conditions, the MLE is strongly consistent and asymptotically normal. We also provide useful expressions for the asymptotic covariance matrices, involving the MLE of the conditional sojourn times and the embedded Markov chain of the hidden semi-Markov chain.

*Key words:* Hidden semi-Markov models, Maximum likelihood estimation, Asymptotic properties

---

**AMS Subject Classification:** 62F12, 60K15, 62M09

\* Corresponding author

*Email address:* [streveza@dma.utc.fr](mailto:streveza@dma.utc.fr) (Samis TREVEZAS).

## 1 Introduction

Hidden Markov models (HMMs) were first introduced by Baum and Petrie (1966), where it is proved the consistency and asymptotic normality of the maximum likelihood estimator (MLE) for this model. In their study, Baum and Petrie consider both the observable and the hidden process with a finite state space. The hidden process forms a Markov chain (MC), and the observable process conditioned on the MC forms a sequence of conditionally independent random variables. This class of HMMs is often referred to, as probabilistic functions of Markov chains. The conditions for consistency are weakened in Petrie (1969). Leroux (1992), Bickel, Ritov and Ryden (1998), proved the consistency and the asymptotic normality of the MLE respectively, when the observable process has a general state space.

The HMMs have a wide range of applications, including speech recognition (see Rabiner (1989), and Rabiner and Juang (1993)), computational biology (see Krogh et al. (1994)), signal processing (see Elliott and Moore (1995)). The reader is also referred to Ephraim and Merhav (2002) for an overview of statistical and information-theoretic aspects of hidden Markov processes (HMPs). Ferguson (1980) introduced the hidden semi-Markov models (HSMMs), where the hidden process actually forms a semi-Markov chain (SMC). This setting allows arbitrary distributions for the sojourn times in the states of the SMC, rather than geometric distributions in the case of the HMM. Recent papers that concentrate on computational techniques for the HSMMs are that of Guédon (2003) and Sansom and Thomson (2001).

To the best of our knowledge, Barbu and Limnios (2006) were the first to study asymptotic properties of the MLE for a HSMM. In this paper we present a different approach which can be summarized as follows:

- i) we generalize the results for the HSMM found therein to the general HSMM, where the state space of the observable process is assumed to be a subset of a Euclidean space. For this purpose, we follow the lines of Leroux (1992) and Bickel et al. (1998),
- ii) we allow the values of the observable process  $Y_n$ , conditioned on the SMC, to depend probabilistically not only on the state  $Z_n$  but also on the time duration that the system has stayed in this current state (backward recurrence time dependence),
- iii) we use minimal representations for the parametric spaces, which are involved in our analysis, taking into consideration the dependence relations among the parameters. We also use for each  $i$  and  $j$  the general constants  $\tilde{n}_{ij}$  to specify the support for the conditional sojourn times, rather than extending the parametric space with identically zero parameters,
- iv) we do a different decomposition of the elements of the semi-Markov kernel, from the one found in Barbu and Limnios (2006).

Together iii) and iv), open the way for explicit expressions for the asymptotic covariance matrices (as functions of the semi-Markov kernel), that appear in the central limit theorems for the MLE of the basic characteristics of the semi-Markov chain.

This paper is organized as follows: In Section 2, we introduce the mathematical notation and we state a first set of conditions. In Section 3, we give a representation of the HSMMs as a subclass of HMMs. In Section 4, we prove the strong consistency of the MLE of the HSMM, and also of the basic characteristics of the SMC, that is, the conditional sojourn times and the embedded Markov chain. In Section 5, we prove the asymptotic normality of the MLE of the HSMM and of the previously mentioned characteristics.

## 2 Preliminaries and assumptions

Let  $(Z_n, Y_n)_{n \in \mathbb{N}}$  be a hidden semi-Markov chain defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ , where  $\theta \in \Theta$ , and  $\Theta$  is a euclidean subset which parametrizes our model and will be specified later in the sequel. We assume that the SMC  $(Z_n)_{n \in \mathbb{N}}$  has finite state space  $E = \{1, 2, \dots, s\}$  and semi-Markov kernel  $(q_{ij}^\theta(k))_{i,j \in E, k \in \mathbb{N}}$ . If we denote  $(J_n, S_n)_{n \in \mathbb{N}^*}$  the associated Markov renewal process to  $Z$ , then  $q_{ij}^\theta(k) = \mathbb{P}_\theta(J_{n+1} = j, S_{n+1} - S_n = k \mid J_n = i)$ ,  $n \geq 1$ . The process  $(S_n)_{n \in \mathbb{N}^*}$  keeps track of the successive time points that changes of states in  $(Z_n)_{n \in \mathbb{N}}$  occur (jump times), and  $(J_n)_{n \in \mathbb{N}^*}$  records the visited states at these time points. Under this consideration,  $q_{ii}^\theta(k) = 0$  for all  $i \in E$ ,  $k \in \mathbb{N}$ . We will use the notation  $\mathbf{Z}_{k_1}^{k_2}$  to denote the vector  $(Z_{k_1}, Z_{k_1+1}, \dots, Z_{k_2})$ ,  $k_1 \leq k_2$ , and  $\mathbf{i}_d$  for a d-dimensional vector with every component equal to the element  $i \in E$ . The distribution of  $\mathbf{Z}_0^{S_1}$  is selected to be  $\mathbb{P}_\theta(\mathbf{Z}_0^{k-1} = \mathbf{i}_k, Z_k = j, S_1 = k) = p_{ij}^\theta \bar{H}_i^\theta(k-1) / \mu_{ii}^\theta$ , where  $p_{ij}^\theta$  refers to the  $(i, j)$  element of the transition matrix of the embedded Markov chain  $(J_n)_{n \in \mathbb{N}^*}$ ,  $\bar{H}_i^\theta(\cdot)$  to the survival function in state  $i$ , and  $\mu_{ii}^\theta$  to the mean recurrence time in the  $i$ -renewal process associated to the semi-Markov chain  $(Z_n)_{n \in \mathbb{N}}$ . We will define later the above quantities as functions of the semi-Markov kernel. The selection of the distribution of  $\mathbf{Z}_0^{S_1}$  is naturally justified from the fact that it corresponds to the distribution of the same vector in a semi-Markov system that has worked for an infinite time period and is censored at an arbitrary time point, that can be considered as the beginning of our observation. In order to be well defined, it is enough  $\mu_{ii} < \infty$ , for all  $i \in E$ .

We state the following conditions concerning the subclass of SMCs to be considered:

**(A1)** There exists a minimum  $\tilde{n} \in \mathbb{N}$  such that  $q_{ij}^\theta(k) = 0$ , for all  $k > \tilde{n}$ ,  $i, j \in E$ , and  $\theta \in \Theta$ .

**(A2)** The MC  $(J_n)_{n \in \mathbb{N}}$  is irreducible.

Under conditions (A1) and (A2), indeed,  $\mu_{ii}^\theta < \infty$  for all  $i \in E$ . It can be shown easily that the previously defined distribution of  $\mathbf{Z}_0^{S_1}$  implies that the SMC  $(Z_n)_{n \in \mathbb{N}}$  is stationary. Because of the stationarity, we can allow  $(Z_n)_{n \in \mathbb{N}}$  to be indexed by  $n \in \mathbb{Z}$ . In this case, we denote  $S_0 = -\inf\{k \in \mathbb{N} : Z_{-k-1} \neq Z_{-k}\}$ . For the observable process, we assume that  $(Y_n)_{n \in \mathbb{N}}$  takes values on the measured space  $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \nu)$ , where usually  $\mathcal{Y} \subset \mathbb{R}^q$  for some  $q \in \mathbb{N}^*$ ,  $\mathcal{B}(\mathcal{Y})$  denotes the Borel subsets on  $\mathcal{Y}$ , and  $\nu$  is a  $\sigma$ -finite measure defined on  $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ . Also, let the conditional probability densities  $g_\theta(y \mid i, k)$  denote the densities that correspond to the conditional distribution functions  $\mathbb{P}_\theta(Y_n \leq y \mid \mathbf{Z}_{n-k}^n = \mathbf{i}_{k+1}, Z_{n-k-1} \neq i)$ ,  $i \in E$ ,  $n, k \in \mathbb{N}$ .

Under condition (A1) there exist constants  $\tilde{n}_{ij}, \tilde{n}_i < \infty$ , such as  $\tilde{n}_{ij} = \max\{k \in \mathbb{N} : q_{ij}^\theta(k) > 0\}$  and  $\tilde{n}_i = \max_{j \in E} \tilde{n}_{ij}$ . The quantities  $\tilde{n}_{ij}$  express the maximum time period that the SMC can stay in state  $i$  before having a direct transition in state  $j$ . These time bounds, for practical purposes, are supposed to be known from the characteristics of the system to which this model can be applied or they can be imposed by the experimenter as an approximation to a more complicated system. The existence of these time bounds is all we need for the theoretical results that will follow. For some  $i, j \in E$ ,  $\tilde{n}_{ij}$  may be equal to zero and this means that no direct transitions from  $i$  to  $j$  are allowed. Under condition (A1), the possible values of  $k$ , referring to the conditional densities  $g_\theta(y \mid i, k)$ , are those for  $0 \leq k \leq \tilde{n}_i - 1$ . In order to simplify the notation we denote by  $D_{ij} = \{1, 2, \dots, \tilde{n}_{ij}\}$  for  $i, j \in E$  that  $\tilde{n}_{ij} > 0$ , and by  $D_i = \{1, 2, \dots, \tilde{n}_i\}$ .

Let  $T$  be a finite index set. Different parametric spaces will be used in the sequel. For the moment we specify the natural parametric space for the HSMM, that is,

$$\Theta := \{q_{ij}(k), \theta_t : k \in D_{ij}, q_{ij}(k) \geq 0, \sum_{j,k} q_{ij}(k) = 1, t \in T\}, \quad (1)$$

and in order to distinguish between the two different kinds of parameters we denote

$$\Theta_1 := \{q_{ij}(k) : k \in D_{ij}, q_{ij}(k) \geq 0, \sum_{j,k} q_{ij}(k) = 1\}, \quad (2)$$

$$\Theta_2 := \{\theta_t : t \in T\}. \quad (3)$$

The space  $\Theta_1$  parametrizes the elements of the semi-Markov kernel, and since in the natural parametrization we have  $q_{ij}^\theta(k) = pr_{ijk}(\theta) = q_{ij}(k)$ , we can then suppress the superindex  $\theta$  from  $q_{ij}^\theta(k)$ . The space  $\Theta_2$  refers to a set of parameters that characterize the conditional densities  $g_\theta(y \mid i, k)$ . It can be the case that they distinguish densities from a specific parametric family, from different parametric families or represent transition probabilities when  $\mathcal{Y}$  is a finite state space. In the most simple case of a single parametric family we have  $g_\theta(y \mid i, k) := g(y \mid \theta(i, k))$ ,  $\theta(i, k) \in A$ , where  $A \subset \mathbb{R}^m$  for some  $m \in \mathbb{N}$ . In this case, the index set  $T$  that appears in  $\Theta_2$  consists of all the possible couples  $(i, k)$ .

From now on, we suppose for simplicity that the cardinality of  $T$ , denoted by  $d_2$ , is equal to  $\sum_i \tilde{n}_i$ , that is, one one-dimensional parameter for each conditional density ( $m = 1$ ). Also, we denote  $d_1 = \sum_{i,j} \tilde{n}_{ij}$ , and  $d = d_1 + d_2$ . Then,  $\Theta_1 \subset \mathbb{R}^{d_1}$ ,  $\Theta_2 \subset \mathbb{R}^{d_2}$ ,  $\Theta = \Theta_1 \times \Theta_2 \subset \mathbb{R}^d$ . Since for all  $i \in E$ ,  $\sum_{j,k} q_{ij}(k) = 1$ , there are  $s$  linear dependence relations among the elements of the semi-Markov kernel. In order to have a minimal representation of  $\Theta$ , we have to express  $s$  elements of the kernel as functions of the others. For this purpose, let

$J_i = \{j \in E : \tilde{n}_{ij} = \tilde{n}_i\}$ . We can choose one element  $j_i \in J_i$ , for all  $i \in E$ , and express the  $s$  elements as follows

$$q_{ij_i}(\tilde{n}_i) = 1 - \sum_{j \in E - \{i, j_i\}} \sum_{1 \leq k \leq \tilde{n}_{ij}} q_{ij}(k) - \sum_{1 \leq k \leq \tilde{n}_i - 1} q_{ij_i}(k). \quad (4)$$

Now, we are in the position to have a minimal representation by using as a parametric space  $\Theta^* := \Theta_1^* \times \Theta_2$ , where  $\Theta_1^*$  results from  $\Theta_1$  after extracting the parameters described as above. Then,  $\Theta_1^* \subset \mathbb{R}^{d_3}$  and  $\Theta^* \subset \mathbb{R}^{d_4}$ , where  $d_3 = d_1 - s$  and  $d_4 = d_1 + d_2 - s = d - s$ .

### 3 Representation of the HSMs as a subclass of HMMs

We will show that the general HSMs with backward recurrence time dependence can be represented as a subclass of HMMs. For this purpose, it is enough to represent the SMCs that satisfy condition (A1) as a special class of MCs. Let  $U = (U_n)_{n \in \mathbb{N}}$  be the sequence of backward recurrence times of the SMC  $(Z_n)_{n \in \mathbb{Z}}$  defined as follows:

$$U_n = n - S_{N(n)}, \quad (5)$$

where  $N(n) = \max\{k \in \mathbb{N} : S_k \leq n\}$ .

Let also  $\bar{H}_i(\cdot)$  be the survival function in state  $i$  defined by

$$\bar{H}_i(n) := \mathbb{P}(S_{l+1} - S_l > n \mid J_l = i) = 1 - \sum_{j \in \mathbf{E}} \sum_{k=0}^n q_{ij}(k), \quad n \in \mathbb{N}, l \in \mathbb{N}^*. \quad (6)$$

It can be shown that the stochastic process  $(Z, U) := (Z_n, U_n)_{n \in \mathbb{N}}$  is a Markov chain (see Limnios and Oprisan (2001), Theorem 3.12). In a recent paper, Chryssaphinou et al. (2008) study properties of the process  $(Z, U)$ . This process plays an important role on the understanding of the semi-Markov structure. On one hand, it can be used to study the probabilistic behavior and



limit theorems for semi-Markov chains and on the other hand to make statistical inference for semi-Markov chains. This role will be extended here in the framework of the HSMMS.

Condition (A1) implies that for all  $i \in E$ , the maximum time period that  $(Z_n)_{n \in \mathbb{N}}$  can stay in this state is  $\tilde{n}_i$ . Therefore, the backward recurrence time  $U_n \in \{0, 1, \dots, \tilde{n}_i - 1\}$  and direct transitions from  $i$  to  $j$  are restricted to maximum backward recurrence time  $\tilde{n}_{ij} - 1$ . Also, it can be easily verified that conditions (A1) and (A2) and the selection of the distribution of  $\mathbf{Z}_0^{S_1}$  as previously mentioned, renders the process  $(Z, U)$  a stationary MC with initial distribution given by  $\mathbb{P}_\theta((Z_0, U_0) = (i, k)) = \overline{H}_i(k)/\mu_{ii}$ ,  $i \in E$ ,  $0 \leq k \leq \tilde{n}_i - 1$ . If we denote by  $P = (p_{(i,k_1)(j,k_2)})$  the  $d_2 \times d_2$  transition probability matrix of the MC  $(Z, U)$ , then the following proposition specifies the transition probabilities of the above MC as a function of the semi-Markov kernel (see also Barbu and Limnios (to appear)). The proof is easy and it is omitted here.

**Proposition 1** *Under condition (A1), the transition probabilities of the Markov chain  $(Z, U)$  can be written as:*

$$P_{(i,k_1)(j,k_2)} = \begin{cases} q_{ij}(k_1 + 1)/\overline{H}_i(k_1), & \text{if } i \neq j \text{ and } k_2 = 0, 0 \leq k_1 \leq \tilde{n}_{ij} - 1, \\ \overline{H}_i(k_1 + 1)/\overline{H}_i(k_1), & \text{if } i = j \text{ and } k_2 - k_1 = 1, 0 \leq k_1 \leq \tilde{n}_i - 2, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where  $\overline{H}_i(\cdot)$  is given by relation (6).

We present here the matrix  $P$  in a block form  $P = (P_{ij})_{i,j \in E}$ , where  $P_{ij}$  is an

$\tilde{n}_i \times \tilde{n}_j$  matrix, and for  $i = j$ ,

$$P_{ii} = \begin{pmatrix} 0 & p_{(i,0)(i,1)} & 0 & \dots & 0 \\ 0 & 0 & p_{(i,1)(i,2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & p_{(i,\tilde{n}_i-2)(i,\tilde{n}_i-1)} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad (8)$$

and for  $i \neq j$ ,

$$P_{ij} = \begin{pmatrix} p_{(i,0)(j,0)} & 0 & \dots & 0 \\ p_{(i,1)(j,0)} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{(i,\tilde{n}_{ij}-1)(j,0)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \quad (9)$$

**Remarks:**

- 1) From relation (7), we conclude that with every semi-Markov kernel that satisfies condition (A1) we can associate a Markov transition matrix with the corresponding transition probabilities.
- 2) If we assume additionally (A2), then  $p_{(i,k)(i,k+1)} > 0$ ,  $i \in E$ ,  $0 \leq k \leq \tilde{n}_i - 2$ .
- 3) When transitions from  $i$  to  $j$  are not allowed ( $\tilde{n}_{ij} = 0$ ), then  $P_{ij}$  is a null matrix, while if  $\tilde{n}_{ij} = \tilde{n}_i$  the first column of  $P_{ij}$  has no fixed zero elements.

In Proposition 1, we regarded the probabilities  $p_{(i,k_1)(j,k_2)}$  as functions of the semi-Markov kernel, which is identified in the natural parametrization with  $\Theta_1$ . These probabilities will be denoted by  $p_{(i,k_1)(j,k_2)}^\theta$  whenever we refer to this parametrization. Additionally, we consider a setting where the parametrization, fits from the beginning, the class of Markov chains described in Proposition 1. Let  $\tilde{\Theta}_1 = \{p_{(i,k_1)(j,k_2)}\} \subset \mathbb{R}^{d_4}$ , where all the identically zero elements that appear in  $P$  have been excluded and the restrictions imposed on the parameters follow from the stochastic nature of the matrix  $P$ . Notice that  $\tilde{\Theta}_1$  can be regarded as the natural parametric space of a subclass of  $d_2$ -state Markov chains with transition matrices that are given in block form by (8) and (9). The number of parameters that appear in  $\tilde{\Theta}_1$  equals  $d_4$ . Since  $P$  is a stochastic matrix, there are exactly  $d_2$  linear relations among the elements of  $P$ . If we exclude one parameter for each row of  $P$ , then the remaining number of parameters equals the dimension of  $\Theta_1^*$ , that is,  $d_3$ .

We denote by  $\tilde{\Theta}_1^* \subset \mathbb{R}^{d_3}$  a minimal representation of  $\tilde{\Theta}_1$ . Similarly, we have  $\tilde{\Theta} = \tilde{\Theta}_1 \times \Theta_2 \subset \mathbb{R}^{d_2+d_4}$ , and  $\tilde{\Theta}^* = \tilde{\Theta}_1^* \times \Theta_2 \subset \mathbb{R}^{d_4}$ . Let  $P_{\tilde{\theta}}$  the generic element of this subclass of  $d_2 \times d_2$  stochastic matrices. We will show the existence of the inverse transformation that represents every MC with  $d_2$  states ( $d_2 = \sum_{i=1}^s \tilde{n}_i$ ) and transition matrix  $P_{\tilde{\theta}}$ , as an  $s$ -state SMC with a kernel that satisfies condition (A1).

**Proposition 2** *There exists a continuous function  $\Psi_1$  from  $\tilde{\Theta}_1^*$  into  $\Theta_1^*$  that reparametrizes every  $d_2$ -state Markov chain with transition probability matrix given by  $P_{\tilde{\theta}}$  by an  $s$ -state semi-Markov chain with a kernel satisfying condition (A1), where the states of the SMC correspond to the blocks that the decomposition of  $P$  indicates from relations (8) and (9).*

**PROOF.** From Theorem 6.7 in Barbu and Limnios (to appear), modified by taking into consideration the constants  $\tilde{n}_{ij}$ , we have for  $i, j$  such that  $\tilde{n}_{ij} > 0$

$$q_{ij}(k) = \begin{cases} p_{(i,0)(j,0)} & \text{if } k = 1, \\ p_{(i,k-1)(j,0)} \prod_{r=0}^{k-2} p_{(i,r)(i,r+1)} & \text{if } 2 \leq k \leq \tilde{n}_{ij}. \end{cases} \quad (10)$$

The proof is complete by letting all the other elements  $q_{ij}(k) = 0$ , for  $\tilde{n}_{ij} = 0$ . For our statistical purposes we will need a specific minimal representation  $\tilde{\Theta}_1^*$ , so as to consider this transformation as a continuous function from the domain  $\tilde{\Theta}_1^*$  to  $\Theta_1^*$ . For this purpose, we find convenient to express  $p_{(i,k_1)(j_i,0)}$  as a function of the other parameters in the same row of  $P$ , where  $j_i$  is defined before relation (4). Therefore, for all  $i \in E$ ,  $0 \leq k_1 \leq \tilde{n}_i - 1$ ,

$$p_{(i,k_1)(j_i,0)} = \begin{cases} 1 - \sum_{\substack{j: \tilde{n}_{ij} \geq k_1+1 \\ j \neq j_i}} p_{(i,k_1)(j,0)} - p_{(i,k_1)(i,k_1+1)} & \text{if } 0 \leq k_1 \leq \tilde{n}_i - 2, \\ 1 - \sum_{j \in G_i} p_{(i,k_1)(j,0)} & \text{if } k_1 = \tilde{n}_i - 1, \end{cases} \quad (11)$$

where  $G_i = \{j : j \neq j_i, \tilde{n}_{ij} = \tilde{n}_i\}$ .

We define  $\Psi_1 : \tilde{\Theta}_1^* \rightarrow \Theta_1^*$ , the desired transformation

$$\Psi_1(p_{(i,k_1)(j,k_2)}) = (q_{ij}(k)), \quad (12)$$

where the component functions of  $\Psi_1$ , for  $i, j \in E$  such that  $\tilde{n}_{ij} > 0$ , are given as follows:

$$q_{ij}(1) = \begin{cases} p_{(i,0)(j,0)} & \text{if } j \neq j_i, \\ 1 - \sum_{j \in G_i} p_{(i,0)(j,0)} - p_{(i,0)(i,1)} & \text{if } j = j_i, \end{cases} \quad (13)$$

$$q_{ij}(k) = \begin{cases} p_{(i,k-1)(j,0)} \prod_{r=0}^{k-2} p_{(i,r)(i,r+1)} & \text{if } j \neq j_i, 2 \leq k \leq \tilde{n}_{ij}, \\ \left(1 - \sum_{j \in G_i} p_{(i,k_1)(j,0)} - p_{(i,k_1)(i,k_1+1)}\right) \prod_{r=0}^{k-2} p_{(i,r)(i,r+1)} & \text{if } j = j_i, 2 \leq k < \tilde{n}_i. \end{cases} \quad (14)$$

By (13) and (14), we conclude that  $\Psi_1$  is continuous.  $\square$

**Remark.** 1) The  $s$  parameters of  $\Theta_1$  that have been excluded in order to obtain  $\Theta_1^*$  can be written as follows:

$$q_{ij_i}(\tilde{n}_i) = \left(1 - \sum_{j \in G} p_{(i,\tilde{n}_i-1)(j,0)}\right) \prod_{r=0}^{\tilde{n}_i-1} p_{(i,r)(i,r+1)}. \quad (15)$$

#### 4 Consistency results

By following the representation of the previous section, the initial HSMM can now be described by this special kind of HMM  $((Z, U), Y)$ . The stationarity of  $(Z, U)$  implies the stationarity of  $((Z, U), Y)$ . We make the assumption in the sequel that the natural parametric space  $\Theta^*$  is a compact subset of  $\mathbb{R}^{d_4}$ . Since  $\Theta_1^*$  is a compact subset of  $\mathbb{R}^{d_3}$ , it is enough  $\Theta_2$  to be compact. If this is not the case, we can use a standard compactification technique (see Leroux (1992), and Kiefer and Wolfowitz (1956)). In the most simple case of a single parametric family we have  $g_\theta(y \mid i, k) := g(y \mid \theta(i, k))$ ,  $\theta(i, k) \in A$ , where  $A \subset \mathbb{R}$ . Here,  $\Theta_2 = A^{d_2}$ . The likelihood function for an observation  $\{\mathbf{Y}_0^n = \mathbf{y}_0^n\}$  can be written as

$$p_\theta(y_0^n) = \sum_{(i,k)_0^n} \pi_\theta(i_0, k_0) \prod_{j=0}^{n-1} p_{(i_j, k_j)(i_{j+1}, k_{j+1})}^\theta \prod_{j=0}^n g(y_j \mid \theta(i_j, k_j)),$$

where  $\pi_\theta(i, k)$  is the stationary distribution of  $P_\theta$ . We denote the real value of the parameter by  $\theta_0$  and  $\tilde{\theta}_0$  when it refers to  $\Theta^*$  and to  $\tilde{\Theta}^*$  respectively. Since

for the results of asymptotic normality of some characteristics of the system we obtain the asymptotic covariance matrices and we calculate derivatives with respect to  $\theta$ , we keep the minimal representation. The estimation problem is to draw inference about this value from a trajectory of  $(Y_n)_{n \in \mathbb{N}}$ . The MLE denoted by  $\hat{\theta}_n$  maximizes  $p_\theta(y_0^n)$  over  $\Theta^*$ . In the “best” case, it is a class, consisting of the parameters  $\theta$ , induced by permutations of a specific value that maximizes the given likelihood. For this reason, we define an equivalence relation  $\sim$  in  $\Theta^*$ , where  $\theta_1 \sim \theta_2$  if  $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ . Then, the results for the estimators should be understood in the context of  $\Theta^*/\sim$ , that is, in the quotient topology induced by this equivalence (see, e.g., Leroux (1992)).

Now, we state some extra conditions in order to deduce that the MLE is consistent. These conditions are found in Leroux (1992), and they are adapted here to our model.

- (B1)** (*Identifiability condition*) The family of mixtures of at most  $d_2$  elements of  $\{g(y | \theta), \theta \in A\}$  is identifiable.
- (B2)** The density function  $g(y | \cdot)$  is continuous in  $A$ , for any  $y \in \mathbb{R}$ .
- (B3)**  $E_{\theta_0}[|\log g(Y_1 | \theta_0(i, k))|] < \infty$ , for all  $i, k$ .
- (B4)**  $E_{\theta_0}[\sup_{|\theta' - \theta| < \delta} (\log g(Y_1 | \theta'))^+] < \infty$  for any  $\theta \in A$ , for some  $\delta > 0$ , where  $x^+ = \max(x, 0)$ .

In this setting, the identifiability of our model is guaranteed if (A1), (A2) and (B1) hold, and additionally the  $\theta(i, k)$  are distinct. For details see Leroux (1992). We are now at the point where the results of consistency for MLE concerning the general HSMMS can be deduced from the corresponding results of the general HMMs. We denote by  $(\hat{q}_{ij}(k, n), \hat{\theta}_t(n))$  the MLE of  $\theta_0 = (q_{ij}^0(k), \theta_t^0)$  over  $\Theta^*$ .

**Theorem 1** *If conditions (A1)-(A2), (B1)-(B4) hold, then the MLE  $\hat{\theta}_n$  is strongly consistent estimator of  $\theta_0$  in the quotient topology, and consequently  $(\hat{q}_{ij}(k, n))$  is strongly consistent estimator of  $(q_{ij}^0(k))$  in the same sense.*

**PROOF.** From Proposition 1 the general HSMM  $(Z, Y)$  parametrized by  $\Theta^*$ , can be viewed as a type of a general HMM  $((Z, U), Y)$  with the same parametric space  $\Theta^*$ . The result will follow from Theorem 3, Section 6, in Leroux (1992), if the conditions 1-6 of that article hold. Indeed, it is easy to verify that Cond.1 of Leroux is deduced from (A1) and (A2). Conditions 2 and 3 are identical to (B1) and (B2). Cond.4 is deduced from the fact that the transition probabilities given in Proposition 1 are continuous functions of the semi-Markov kernel and Conditions 5 and 6 are identical to (B3) and (B4).  $\square$

Let the matrix  $(p_{ij})$  denote the probability matrix of the embedded Markov chain  $(J_n)_{n \in \mathbb{N}}$ , and  $(f_{ij}(k))$  the conditional sojourn times, that is, for  $i, j \in E$ ,

$$p_{ij} = \begin{cases} \sum_{k=1}^{\tilde{n}_{ij}} q_{ij}(k) & \text{if } \tilde{n}_{ij} > 0, \\ 0 & \text{if } \tilde{n}_{ij} = 0, \end{cases} \quad (16)$$

$$f_{ij}(k) = \begin{cases} \frac{q_{ij}(k)}{p_{ij}} & \text{if } \tilde{n}_{ij} > 0, \ 1 \leq k \leq \tilde{n}_{ij}, \\ 0 & \text{if } \tilde{n}_{ij} = 0. \end{cases} \quad (17)$$

Since these quantities are expressed as functions of the semi-Markov kernel, we refer to them as  $p_{ij}^\theta$  and  $f_{ij}^\theta(k)$  to show that they are parametrized over  $\Theta^*$ . Nevertheless, we will omit the superindex  $\theta$  for the estimators. Therefore, we denote by  $(\hat{p}_{ij}(n))$  and  $(\hat{f}_{ij}(k, n))$  the corresponding MLE for the true values  $(p_{ij}^0)$  and  $(f_{ij}^0(k))$  respectively (regarded as vectors), where we exclude the

identically zero parameters. Also, let  $c_i = \text{card}\{j : \tilde{n}_{ij} > 0\}$ , for all  $i \in E$ , and  $\tilde{c} = \sum_i c_i$ .

Then, the following asymptotic results hold:

**Corollary 3** *Under conditions (A1)-(A2), (B1)-(B4),*

*i) the MLE of the embedded Markov chain  $(\hat{p}_{ij}(n))$  is strongly consistent estimator of  $(p_{ij}^0)$ ,*

*ii) the MLE of the conditional sojourn time  $(\hat{f}_{ij}(k, n))$  is strongly consistent estimator of  $(f_{ij}^0(k))$ .*

**PROOF.** i) We define the function  $\Phi : \Theta^* \rightarrow \mathbb{R}^{\tilde{c}}$ , where from relation (16),

$$\Phi(\theta) = \Phi(q_{ij}(k), \theta_t) = (\sum_{k=1}^{\tilde{n}_{ij}} q_{ij}(k)) = (p_{ij}^\theta) \text{ (for } i, j \in E \text{ such that } \tilde{n}_{ij} > 0).$$

We conclude that  $(\hat{p}_{ij}(n)) = \widehat{\Phi}(\hat{\theta})(n) = \Phi(\hat{\theta}_n) = (\sum_{k=1}^{\tilde{n}_{ij}} \hat{q}_{ij}(k, n))$ ,

where the second equality holds from the property of MLE. Consequently, we get from the continuous mapping theorem, using Theorem 1 together with the continuity of  $\Phi$  that

$$(\hat{p}_{ij}(n)) \xrightarrow[n \rightarrow \infty]{a.s.} (p_{ij}^0).$$

ii) Let  $pr_{ijk}(\theta) = q_{ij}(k)$  denote the projection of  $\theta \in \Theta^*$  into the corresponding element of the semi-Markov kernel, and  $\Phi_{ij}$  the component function of  $\Phi$  which corresponds to  $p_{ij}^\theta$ . Let also  $T : \Theta^* \rightarrow \mathbb{R}^{d_1}$ , where  $T(\theta) = (T_{ijk}(\theta)) = (pr_{ijk}(\theta)/\Phi_{ij}(\theta))$ . Then, for  $i, j \in E$  such that  $\tilde{n}_{ij} > 0$ ,  $1 \leq k \leq \tilde{n}_{ij}$ , we have

$$(f_{ij}^\theta(k)) = \left( \frac{q_{ij}(k)}{p_{ij}^\theta} \right) = \left( \frac{pr_{ijk}(\theta)}{\Phi_{ij}(\theta)} \right) = T(\theta),$$

and since  $T$  is continuous, the result follows along the line of reasoning of theorem 1 i).

□



## 5 Asymptotic normality results

Two very useful notions for statistical inference, closely connected with MLE, are the rate of entropy of a stochastic process and the generalized Kullback-Leibler divergence. Because of the stationarity of  $((Z, U), Y)$ , we can allow  $((Z_n, U_n), Y_n)_{n \in \mathbb{N}}$  to be indexed by  $n \in \mathbb{Z}$ . In this case, the rate of entropy of the stochastic process  $((Z, U), Y)$  is defined as

$$-\mathbb{H}(\theta_0) := -\mathbb{E}_{\theta_0}[\log \mathbb{P}_{\theta_0}(Y_0 | Y_{-1}, Y_{-2}, \dots)].$$

and the generalized Kullback-Leibler divergence is defined as

$$\mathbb{H}_{\theta_0}(\theta) := \mathbb{E}_{\theta_0}[\log \mathbb{P}_{\theta}(Y_0 | Y_{-1}, Y_{-2}, \dots)], \theta \in \Theta^*.$$

More details about their use in the proofs of consistency can be found in Leroux (1992). We denote by  $\sigma(\theta_0)$  the opposite of the Hessian matrix of  $\mathbb{H}_{\theta_0}(\theta)$ , calculated in  $\theta_0$ , i.e.,

$$\sigma(\theta_0) = \left( \sigma_{u,v}(\theta_0) \right)_{u,v} := - \left( \frac{\partial^2 \mathbb{H}_{\theta_0}(\theta)}{\partial \theta_u \partial \theta_v} \Big|_{\theta=\theta_0} \right)_{u,v}$$

A third set of conditions will be established, which is based on the article of Bickel et al. (1998), to ensure asymptotic normality of the MLE. The conditions, adapted to our model, can be stated as follows:

- (C1)** The MC  $(Z_n, U_n)_{n \in \mathbb{N}}$  is aperiodic.
- (C2)** The conditional densities  $g(y | \theta(i, k))$  have two continuous derivatives with respect to  $\theta \in \Theta^*$ , in some neighborhood of  $\theta_0$ , for all the possible values  $i, k, y$ .
- (C3)** There exists a  $\delta > 0$ , for all  $i, k$  such as

$$\text{i) } E_{\theta_0} \left[ \sup_{|\theta - \theta_0(i,k)| < \delta} \left| \frac{d}{d\theta} \log g(Y_1 | \theta) \right|^2 \right] < \infty,$$

- ii)  $E_{\theta_0} \left[ \sup_{|\theta - \theta_0(i,k)| < \delta} \left| \frac{d^2}{d\theta^2} \log g(Y_1 | \theta) \right| \right] < \infty,$
- iii)  $\int \sup_{|\theta - \theta_0(i,k)| < \delta} \left| \frac{d^j}{d\theta^j} g(y | \theta) \right| \nu(dy) < \infty, \quad \text{for } 1 \leq j \leq 2.$

(C4) For  $\theta_0 \in \Theta^*$ , there exists a  $\delta > 0$  such as, if

$$r_{\theta_0}(y) := \sup_{\|\theta - \theta_0\| < \delta} \max_{(i_1, k_1), (i_2, k_2)} \frac{g(y | \theta(i_1, k_1))}{g(y | \theta(i_2, k_2))},$$

then,  $\mathbb{P}_{\theta_0}(r_{\theta_0}(Y_1) = \infty | (Z_1, U_1) = (i, k)) < 1$ , for all  $i, k$ .

(C5) The true value  $\theta_0$  is an interior point of  $\Theta^*$ .

(C6) The matrix  $\sigma(\theta_0)$  is nonsingular.

**Remark.** The conditions (C1)-(C3), which involve the densities  $g(y | \theta(i, k))$ , can be substituted with similar conditions for the more general conditional densities  $g_{\theta}(y | i, k)$ , as they appear in Bickel et al. (1998).

**Theorem 2** *Under conditions (A1)-(A2), (B1)-(B4) and (C1)-(C6), the MLE  $\hat{\theta}_n$  of  $\theta_0$  is asymptotically normal, that is,  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma(\theta_0)^{-1})$ .*

**PROOF.** Since Proposition 1 holds, the result will follow from Theorem 1, Section 3 of Bickel et al. (1998), if the conditions for asymptotic normality that are stated there hold. Indeed, conditions (A1), (A2) and (C1) render the process  $(Z, U)$  an ergodic Markov chain with finite state space and therefore condition (A1) of Bickel et al. (1998) is satisfied. The conditions (B1)-(B4), together with (A1) and (A2) imply (A6) of Bickel et al. (1998). The other conditions are adapted naturally to our model.  $\square$

At this point we will connect the two natural parametric spaces  $\Theta^*$  and  $\tilde{\Theta}^*$  for the general HSMM and the type of the general HMM that we have already considered respectively, by giving a connection between the two asymptotic

covariance matrices of the MLE of the HMM and the MLE of the associated HSMM given from Proposition 1.

As we can see from relation (12),  $\Psi_1$  is differentiable on  $\tilde{\Theta}_1^*$ . By extending the domain of  $\Psi_1$  in order to include the  $d_2$  parameters for the conditional densities, but keeping the same range, we define  $\Psi : \tilde{\Theta}^* \rightarrow \Theta^*$ , where  $\Psi = (\Psi_1, pr_{d_2})$ , and  $pr_{d_2}$  is the projection function on  $\Theta_2$ . This function is differentiable at  $\tilde{\theta} \in \tilde{\Theta}^*$ , and we denote by  $\Psi'$  the total derivative of  $\Psi$  calculated at  $\tilde{\theta}_0$ . Let also  $\sigma(\tilde{\theta}_0)^{-1}$  be the asymptotic covariance matrix of the MLE  $\hat{\theta}_n$  of  $\tilde{\theta}_0$ . Whenever necessary we will use the following decomposition of the matrix  $\sigma(\tilde{\theta}_0)^{-1}$ ,

$$\sigma(\tilde{\theta}_0)^{-1} = \begin{array}{cc} & \begin{array}{c} d_3 \qquad d_2 \\ \underbrace{\hspace{1.5cm}} \qquad \underbrace{\hspace{1.5cm}} \end{array} \\ \left[ \begin{array}{cc} \sigma(\tilde{\theta}_0)_{11}^{-1} & \sigma(\tilde{\theta}_0)_{12}^{-1} \\ \sigma(\tilde{\theta}_0)_{21}^{-1} & \sigma(\tilde{\theta}_0)_{22}^{-1} \end{array} \right] & \left. \begin{array}{l} \} d_3 \\ \\ \} d_2 \end{array} \right\} \end{array} \quad (18)$$

The following theorem expresses the asymptotic covariance matrix of the MLE that corresponds to the HSMM in terms of the natural parametric space  $\tilde{\Theta}^*$  associated to the HMM.

**Theorem 3** *Under conditions (A1)-(A2), (B1)-(B4) and (C1)-(C6), the MLE  $\hat{\theta}_n$  of  $\theta_0$ , that corresponds to the natural parametric space of the general HSMM*

*satisfies:  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Psi' \sigma(\tilde{\theta}_0)^{-1} (\Psi')^\top)$ , as  $n \rightarrow \infty$ .*

*Consequently,  $\sqrt{n}(\hat{q}_{ij}(k, n) - q_{ij}^0(k)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Psi'_1 \sigma(\tilde{\theta}_0)_{11}^{-1} (\Psi'_1)^\top)$ , where the matrix  $\Psi'$  is given analytically from relations (31)-(35), and  $\Psi'_1$  is the submatrix of  $\Psi'$ , taking its first  $d_3$  rows and columns.*

**PROOF.** Let for any  $i \in E$ ,  $\tilde{n}_{i\tau_i(1)}, \tilde{n}_{i\tau_i(2)}, \dots, \tilde{n}_{i\tau_i(c_i)}$ , the ordered sequence of  $\tilde{n}_{ij}$ , for those  $j$  such as  $\tilde{n}_{ij} > 0$ . In the case that some elements are equal, the ordering is considered to be done according to the order of the indexes  $j$  as natural numbers. Note that since  $\tilde{n}_{i\tau_i(c_i)} = \tilde{n}_i$ , then  $\tau_i(c_i) \in J_i$ , and therefore we can choose  $j_i = \tau_i(c_i)$ . Let for all  $i \in E$ ,

$$\underline{q}(i\tau_i(j)) = \begin{cases} \left( q_{i\tau_i(j)}(1), q_{i\tau_i(j)}(2), \dots, q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) \right) & \text{if } 1 \leq j \leq c_i - 1, \\ \left( q_{ij_i}(1), q_{ij_i}(2), \dots, q_{ij_i}(\tilde{n}_{ij_i} - 1) \right) & \text{if } j = c_i. \end{cases} \quad (19)$$

and also

$$\underline{q}(i) = \left( \underline{q}(i\tau_i(1)), \underline{q}(i\tau_i(2)), \dots, \underline{q}(ij_i) \right). \quad (20)$$

Then, if we denote by  $\underline{\theta}^{(2)}$  the parameters that correspond to  $\Theta_2$ , an arrangement of the parameters of  $\Theta^*$ , can be presented as follows:

$$(q_{ij}(k), \theta_t) = \left( \underline{q}(1), \underline{q}(2), \dots, \underline{q}(s), \underline{\theta}^{(2)} \right). \quad (21)$$

We will need a corresponding arrangement of the elements of  $\tilde{\Theta}^*$ . For this purpose, let for all  $i \in E$ ,  $1 \leq j \leq c_i - 1$ ,

$$\underline{p}(ii) = \left( p_{(i,0)(i,1)}, p_{(i,1)(i,2)}, \dots, p_{(i,\tilde{n}_i-2)(i,\tilde{n}_i-1)} \right), \quad (22)$$

$$\underline{p}(i\tau_i(j)) = \left( p_{(i,0)(\tau_i(j),0)}, p_{(i,1)(\tau_i(j),0)}, \dots, p_{(i,\tilde{n}_{i\tau_i(j)}-1)(\tau_i(j),0)} \right). \quad (23)$$

Then, denoting by

$$\underline{p}(i) = \left( \underline{p}(i\tau_i(1)), \underline{p}(i\tau_i(2)), \dots, \underline{p}(i\tau_i(c_i - 1)), \underline{p}(ii) \right), \quad (24)$$

an expression for an arrangement of the parameters of  $\tilde{\Theta}^*$ , is given by

$$(p_{(i,k_1)(j,k_2)}, \theta_t) = \left( \underline{p}(1), \underline{p}(2), \dots, \underline{p}(s), \underline{\theta}^{(2)} \right). \quad (25)$$

Using relations (13),(14),(21),(25), we have a block decomposition for  $\Psi'$  as indicated below.

$$\Psi' = \begin{pmatrix} M^{(1)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M^{(2)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{0} & \mathbf{0} & \dots & M^{(s)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_{d_2} \end{pmatrix}, \quad (26)$$

where for all  $i \in E$ ,  $M^{(i)} = \begin{pmatrix} \partial \underline{q}(i) \\ \partial \underline{p}(i) \end{pmatrix}$ . Using relations (13),(14),(20),(24), we decompose  $M^{(i)}$  into blocks as follows:

$$M^{(i)} = \begin{pmatrix} M_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & M_{1c_i}^{(i)} \\ \mathbf{0} & M_{22}^{(i)} & \dots & \mathbf{0} & M_{2c_i}^{(i)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & M_{c_i-1,c_i-1}^{(i)} & M_{c_i-1,c_i}^{(i)} \\ M_{c_i1}^{(i)} & M_{c_i2}^{(i)} & \dots & M_{c_i,c_i-1}^{(i)} & M_{c_i c_i}^{(i)} \end{pmatrix}, \quad (27)$$

where for  $1 \leq j \leq c_i - 1$ ,

$$M_{jj}^{(i)} = \begin{pmatrix} \partial \underline{q}(i\tau_i(j)) \\ \partial \underline{p}(i\tau_i(j)) \end{pmatrix}, M_{jc_i}^{(i)} = \begin{pmatrix} \partial \underline{q}(i\tau_i(j)) \\ \partial \underline{p}(ii) \end{pmatrix}, M_{c_i j}^{(i)} = \begin{pmatrix} \partial \underline{q}(i\tau_i(c_i)) \\ \partial \underline{p}(i\tau_i(j)) \end{pmatrix},$$

and

$$M_{c_i c_i}^{(i)} = \begin{pmatrix} \partial \underline{q}(i\tau_i(c_i)) \\ \partial \underline{p}(ii) \end{pmatrix}.$$

These four different types of matrices summarize all the information we want in order to have an explicit matrix form for  $\Psi'$ , and we study each one of

them.

For all  $i \in E$ ,  $1 \leq k \leq \tilde{n}_i - 1$ , let

$$a_i(k) = \prod_{r=0}^{k-1} p_{(i,r)(i,r+1)}, \quad (28)$$

$$a_i(k; l) = \frac{a_i(k)}{p_{(i,l-1)(i,l)}}, \quad 1 \leq l \leq k, \quad (29)$$

$$b_{iu}^{(j)}(k; l) = p_{(i,u)(\tau_i(j),0)} a_i(k; l), \quad 1 \leq l \leq k, \quad 1 \leq u \leq \tilde{n}_{i\tau_i(j)} - 1. \quad (30)$$

Recall that  $j_i = \tau_i(c_i)$  and we will also use the abbreviation  $c_{ij} = \tilde{n}_{i\tau_i(j)} - 2$ ,  $c_{ij}^+ = c_{ij} + 1$ . Then,

$$M_{jj}^{(i)} = \text{diag}\{1, a_i(1), a_i(2), \dots, a_i(c_{ij}^+)\}. \quad (31)$$

$$M_{c_{ij}}^{(i)} = \begin{pmatrix} \Delta_{c_{ij}}^{(i)} & \underline{0}^\top \\ \mathbf{0} & \underline{0}^\top \end{pmatrix}, \quad (32)$$

where

$$\Delta_{c_{ij}}^{(i)} = -\text{diag}\{1, a_i(1), a_i(2), \dots, a_i(c_{ij})\}, \quad (33)$$

$$M_{jc_i}^{(i)} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \dots 0 \\ b_{i1}^{(j)}(1; 1) & 0 & \dots & 0 & 0 \dots 0 \\ b_{i2}^{(j)}(2; 1) & b_{i2}^{(j)}(2; 2) & \dots & 0 & 0 \dots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ b_{i,c_{ij}^+}^{(j)}(c_{ij}^+; 1) & b_{i,c_{ij}^+}^{(j)}(c_{ij}^+; 2) & \dots & b_{i,c_{ij}^+}^{(j)}(c_{ij}^+; c_{ij}^+) & 0 \dots 0 \end{pmatrix}, \quad (34)$$

$$M_{c_i c_i}^{(i)} = \begin{pmatrix} -1 & 0 & \dots & 0 & 0 \\ b_{i1}^{(j_i)}(1; 1) & -a_i(1) & \dots & 0 & 0 \\ b_{i2}^{(j_i)}(2; 1) & b_{i2}^{(j_i)}(2; 2) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{i, c_{ij_i}}^{(j_i)}(c_{ij_i}; 1) & b_{i, c_{ij_i}}^{(j_i)}(c_{ij_i}; 2) & \dots & b_{i, c_{ij_i}}^{(j_i)}(c_{ij_i}; c_{ij_i}) & -a_i(c_{ij_i}) \end{pmatrix} \quad (35)$$

Since

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\Psi(\hat{\theta}_n) - \Psi(\tilde{\theta}_0)), \quad (36)$$

where  $\Psi$  is differentiable at  $\tilde{\theta}_0$ , then Theorem 3 follows from Theorem 2, by an application of the delta method.  $\square$

**Remark.** In order to find the asymptotic covariance matrix of  $\sqrt{n}(\hat{q}_{ij}(k, n) - q_{ij}^0(k))$ , regarded in  $\Theta_1$  instead of  $\Theta_1^*$ , we add the parameters  $q_{ij}(\tilde{n}_i)$ , given from (15), and using relation (4) we conclude that

$\sqrt{n}(\hat{q}_{ij}(k, n) - q_{ij}^0(k)) \rightarrow \mathcal{N}(0, C\Psi'_1\sigma(\tilde{\theta}_0)_{11}^{-1}(\Psi'_1)^\top C^\top)$ , where

$$C = \text{diag}\{C_i, i \in E\}, \quad C_i = \begin{pmatrix} \mathbf{I}_{r_i} \\ -\underline{1} \end{pmatrix}, \quad r_i = \sum_{j=1}^{c_i} \tilde{n}_{i\tau_i(j)} - 1.$$

Let  $\Phi_1$  and  $T_1$  be  $\Phi$  and  $T$  respectively, regarded as functions with domain  $\Theta_1^*$ , where  $\Phi$  and  $T$  are defined in Corollary 3. We give in the following two propositions the asymptotic normality results for the MLE of the characteristics of the semi-Markov system, defined by (16) and (17).

**Proposition 4** *Under conditions (A1)-(A2), (B1)-(B4) and (C1)-(C6), the MLE of the embedded Markov chain is asymptotically normal, that is,*

$\sqrt{n}((\hat{p}_{i,j}(n)) - (p_{ij}^0)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Phi_1' \Psi_1' \sigma(\tilde{\theta}_0)^{-1} (\Phi_1' \Psi_1')^\top)$ , where  $\Phi_1' \Psi_1'$  is given by relations (41) and (42).

**PROOF.** Let for all  $i \in E$ ,

$$\underline{p}^e(i) = (p_{i\tau_i(1)}, p_{i\tau_i(2)}, \dots, p_{i\tau_i(c_i)}). \quad (37)$$

Then, an arrangement of the parameters  $(p_{ij})$  of the embedded MC can be presented as follows:

$$(p_{ij}) = (\underline{p}^e(1), \underline{p}^e(2), \dots, \underline{p}^e(s)). \quad (38)$$

If we denote by  $\left(\frac{\partial \underline{p}^e(i_1)}{\partial \underline{q}(i_2)}\right) := \left(\frac{\partial p_{i_1 j_1}}{\partial q_{i_2 j_2}(k)}\right) = \Phi_1'$ , and  $V^{(i)} := \left(\frac{\partial \underline{p}^e(i)}{\partial \underline{q}(i)}\right)$ , then,

$$\Phi_1' = \text{diag}\{V^{(i)}, i \in E\}, \quad (39)$$

where

$$V^{(i)} = \begin{pmatrix} \underline{1}_{11}^{(i)} & \underline{0} & \dots & \underline{0} & \underline{0} \\ \underline{0} & \underline{1}_{22}^{(i)} & \dots & \underline{0} & \underline{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \underline{0} & \underline{0} & \dots & \underline{1}_{c_i-1, c_i-1}^{(i)} & \underline{0} \\ -\underline{1}_{c_i 1}^{(i)} & -\underline{1}_{c_i 2}^{(i)} & \dots & -\underline{1}_{c_i, c_i-1}^{(i)} & \underline{0} \end{pmatrix}, \quad (40)$$

and  $\underline{1}_{jj}^{(i)}$ ,  $\underline{1}_{c_i j}^{(i)}$ , are  $\tilde{n}_{i\tau_i(j)}$ -dimensional row vectors, with entries 1, for all  $j$  such that  $1 \leq j \leq c_i - 1$ .

Since  $\sqrt{n}((\hat{p}_{i,j}(n)) - (p_{ij}^0)) = \sqrt{n}(\Phi_1(\hat{q}_{ij}(k, n)) - \Phi_1(q_{ij}^0(k)))$ , by using Theorem 3 and the differentiability of  $\Phi_1$  on  $\Theta_1^*$ , we conclude from an application of the delta method that  $\sqrt{n}((\hat{p}_{i,j}(n)) - (p_{ij}^0)) \rightarrow \mathcal{N}(0, \Phi_1' \Psi_1' \sigma(\tilde{\theta}_0)^{-1} (\Phi_1' \Psi_1')^\top)$ ,



where

$$\Phi'_1 \Psi'_1 = \text{diag}\{V^{(i)} M^{(i)}, i \in E\}, \quad (41)$$

and  $V^{(i)}, M^{(i)}$  are given by (40) and (27) respectively. The explicit form of their product for all  $i \in E$ , is given as follows:

$$V^{(i)} M^{(i)} = \begin{pmatrix} \underline{d}_1^{(i)} & \underline{0} & \dots & \underline{0} & \underline{0} \\ \underline{0} & \underline{d}_2^{(i)} & \dots & \underline{0} & \underline{0} \\ \vdots & \vdots & \ddots & \vdots & \\ \underline{0} & \underline{0} & \dots & \underline{d}_{c_i-1}^{(i)} & \underline{0} \\ -\underline{d}_1^{(i)} & -\underline{d}_2^{(i)} & \dots & -\underline{d}_{c_i-1}^{(i)} & \underline{0} \end{pmatrix}, \quad (42)$$

where  $\underline{d}_j^{(i)} = (1, a_i(1), a_i(2), \dots, a_i(c_{ij}^+))$ , and  $a_i(k)$  are given by (28).  $\square$

**Proposition 5** *Under conditions (A1)-(A2), (B1)-(B4) and (C1)-(C6), the MLE of the conditional sojourn times is asymptotically normal, that is,*

$$\sqrt{n} \left( (\hat{f}_{ij}(k, n)) - (f_{ij}^0(k, n)) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, T'_1 \Psi'_1 \sigma(\theta_0)^{-1} (T'_1 \Psi'_1)^t).$$

**PROOF.** Let for all  $i \in E$ ,  $1 \leq j \leq c_i$ ,

$$\underline{f}(i\tau_i(j)) = \left( f_{i\tau_i(j)}(1), f_{i\tau_i(j)}(2) \dots, f_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) \right), \quad (43)$$

and for all  $i \in E$ ,

$$\underline{f}(i) = \left( \underline{f}(i\tau_i(1)), \underline{f}(i\tau_i(2)), \dots, \underline{f}(i\tau_i(c_i)) \right). \quad (44)$$

Then, an arrangement of the parameters  $(f_{ij}(k))$  of the conditional sojourn times can be presented as follows:

$$(f_{ij}(k)) = \left( \underline{f}(1), \underline{f}(2), \dots, \underline{f}(s) \right). \quad (45)$$

If we denote by  $\left(\frac{\partial \underline{f}(i_1)}{\partial \underline{q}(i_2)}\right) := \left(\frac{\partial f_{i_1 j_1}(k_1)}{\partial q_{i_2 j_2}(k_2)}\right) = T_1'$ , and  $F^{(i)} := \left(\frac{\partial \underline{f}(i)}{\partial \underline{q}(i)}\right)$ , then,

$$T_1' = \text{diag}\{F^{(i)}, i \in E\}, \quad (46)$$

where

$$F^{(i)} = \begin{pmatrix} F_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & F_{22}^{(i)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & F_{c_i-1, c_i-1}^{(i)} & \mathbf{0} \\ F_{c_i 1}^{(i)} & F_{c_i 2}^{(i)} & \dots & F_{c_i, c_i-1}^{(i)} & F_{c_i c_i}^{(i)} \end{pmatrix}, \quad (47)$$

and the matrices  $F_{j_1 j_2}^{(i)} := \left(\frac{\partial \underline{f}(i\tau_i(j_1))}{\partial \underline{q}(i\tau_i(j_2))}\right)$ , for the different values of  $j_1$  and  $j_2$  that correspond to the non zero matrices in (47), are given by

$$F_{jj}^{(i)} = -\frac{1}{p_{i\tau_i(j)}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{i\tau_i(j)}(k) & q_{i\tau_i(j)}(1) & \dots & q_{i\tau_i(j)}(1) \\ q_{i\tau_i(j)}(2) & -\sum_{k \neq 2} q_{i\tau_i(j)}(k) & \dots & q_{i\tau_i(j)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & \dots & -\sum_{k \neq \tilde{n}_{i\tau_i(j)}} q_{i\tau_i(j)}(k) \end{pmatrix}, \quad (48)$$

$$F_{c_i j}^{(i)} = \frac{1}{p_{ij}^2} \begin{pmatrix} q_{ij}(1) & q_{ij}(1) & \dots & q_{ij}(1) \\ \vdots & \vdots & \ddots & \vdots \\ q_{ij}(\tilde{n}_i - 1) & q_{ij}(\tilde{n}_i - 1) & \dots & q_{ij}(\tilde{n}_i - 1) \\ -\sum_{k \neq \tilde{n}_i} q_{ij}(k) & -\sum_{k \neq \tilde{n}_i} q_{ij}(k) & \dots & -\sum_{k \neq \tilde{n}_i} q_{ij}(k) \end{pmatrix}, \quad (49)$$

$$F_{c_i c_i}^{(i)} = \frac{1}{p_{ij}} \begin{pmatrix} \mathbf{I}_{s_i} \\ -\mathbf{1} \end{pmatrix}, \quad \text{where } s_i = c_{ij}^+. \quad (50)$$

Since  $\sqrt{n} \left( (\hat{f}_{ij}(k)) - (f_{ij}^0(k)) \right) = \sqrt{n} \left( T_1(\hat{q}_{ij}(k, n)) - T_1(q_{ij}^0(k)) \right)$ , by using Theorem 3, and the differentiability of  $T_1$  on  $\Theta_1^*$ , we conclude from an application of delta method that  $\sqrt{n} \left( (\hat{f}_{ij}(k)) - (f_{ij}^0(k)) \right) \rightarrow \mathcal{N}(0, T_1' \Psi_1' \sigma(\theta_0)^{-1} (T_1' \Psi_1')^\top)$ , where

$$T_1' \Psi_1' = \text{diag}\{F^{(i)} M^{(i)}, i \in E\}, \quad (51)$$

and  $F^{(i)}$ ,  $M^{(i)}$ , are given by (47) and (27) respectively. The explicit form of these matrices for all  $i \in E$ , is given as follows:

$$F^{(i)} M^{(i)} = \begin{pmatrix} D_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_{22}^{(i)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & D_{c_i-1, c_i-1}^{(i)} & \mathbf{0} \\ D_{c_i 1}^{(i)} & D_{c_i 2}^{(i)} & \dots & D_{c_i, c_i-1}^{(i)} & D_{c_i, c_i}^{(i)} \end{pmatrix}, \quad (52)$$

where for  $1 \leq j \leq c_i - 1$ ,

$$D_{jj}^{(i)} = -\frac{1}{p_{i\tau_i(j)}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{i\tau_i(j)}(k) & a_i(1)q_{i\tau_i(j)}(1) & \dots & a_i(c_{ij}^+)q_{i\tau_i(j)}(1) \\ q_{i\tau_i(j)}(2) & -a_i(1)\sum_{k \neq 2} q_{i\tau_i(j)}(k) & \dots & a_i(c_{ij}^+)q_{i\tau_i(j)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & a_i(1)q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & \dots & -a_i(c_{ij}^+)\sum_{k \neq \tilde{n}_{i\tau_i(j)}} q_{i\tau_i(j)}(k) \end{pmatrix}, \quad (53)$$

$$D_{c_{ij}}^{(i)} = \frac{1}{p_{ij_i}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{ij_i}(k) & a_i(1)q_{ij_i}(1) & \dots & a_i(c_{ij}^+)q_{ij_i}(1) \\ q_{ij_i}(2) & -a_i(1)\sum_{k \neq 2} q_{ij_i}(k) & \dots & a_i(c_{ij}^+)q_{ij_i}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{ij_i}(\tilde{n}_i) & a_i(1)q_{ij_i}(\tilde{n}_i) & \dots & -a_i(c_{ij}^+)\sum_{k \neq \tilde{n}_i} q_{ij_i}(k) \end{pmatrix}, \quad (54)$$

and

$$D_{c_i, c_i}^{(i)} = \sum_{j=1}^{c_i} F_{c_{ij}}^{(i)} M_{j c_i}^{(i)}, \quad (55)$$

where  $F_{c_{ij}}^{(i)}$ ,  $M_{j c_i}^{(i)}$  are given by (34)-(35) and (49)-(50).  $\square$

## References

- Barbu, V., Limnios, N., 2006. Maximum likelihood estimation for hidden semi-Markov models. C.R. Acad. Sci. Paris 342, 201–205.
- Barbu, V., Limnios, N., to appear. Semi-Markov Chains and Hidden Semi-Markov Models toward Applications. Springer.
- Baum, L. E., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains. Ann. Math. Stat. 37, 1554–1563.

- Bickel, P. J., Ritov, Y., Ryden, T., 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden-markov models. *Ann. Statist.* 26 (4), 1614–1635.
- Chryssaphinou, O., Karaliopoulou, M., Limnios, N., 2008. On discrete time semi-markov chains and applications in words occurrences. *Communications in Statistics-Theory and Methods* 37, 1306–1322.
- Elliott, R., A. L., Moore, J., 1995. *Hidden Markov Models: Estimation and Control*. Springer, New York.
- Ephraim, Y., Merhav, N., 2002. Hidden Markov Processes. *IEEE Trans. Inf. Theory* 48 (6), 1518–1569.
- Ferguson, J., 1980. Variable duration models for speech. In: *Proc. of the symposium on the Application of Hidden Markov Models to Text and Speech*. Princeton, New Jersey, pp. 143–179.
- Guédon, Y., 2003. Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics* 12 (3), 604–639.
- Kiefer, J., Wolfowitz, J., 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Stat.* 27, 887–906.
- Krogh, A., Brown, M., Mian, I. S., Sjolander, K., Haussler, D., 1994. Hidden Markov models in computational biology: Applications to protein modeling. *J. Molecular Biology* 235, 1501–1531.
- Leroux, B. G., 1992. Maximum-likelihood estimation for hidden markov models. *Stochastic Process. Appl.* 40, 127–143.
- Limnios, N., Oprisan, G., 2001. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston.
- Petrie, T., 1969. Probabilistic functions of finite state Markov chains. *Ann.*

Math. Stat. 40, 97–115.

Rabiner, L. R., 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–284.

Rabiner, L. R., Juang, B. H., 1993. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ.

Sansom, J., Thomson, P., 2001. Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability* 38A, 142–157.