Available online at www.sciencedirect.com

SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 342 (2006) 201–205

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

Maximum likelihood estimation for hidden semi-Markov models

Vlad Barbu, Nikolaos Limnios

Laboratoire de mathématiques appliquées de Compiègne, Université de technologie de Compiègne, BP 20529, 60205 Compiègne, France

Received 29 June 2005; accepted after revision 29 November 2005

Presented by Paul Deheuvels

Abstract

In this Note we consider a discrete-time hidden semi-Markov model and we prove that the nonparametric maximum likelihood estimators for the characteristics of such a model have nice asymptotic properties, namely consistency and asymptotic normality. ***To cite this article: V. Barbu, N. Limnios, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

Résumé

Estimation du maximum de vraisemblance d'un modèle semi-markovien caché. Dans cette Note nous introduisons un modèle semi-markovien caché à temps discret et nous prouvons que les estimateurs du maximum de vraisemblance non-paramétrique d'un tel modèle ont de bonnes propriétés asymptotiques, à savoir la convergence et la normalité asymptotique. ***Pour citer cet article : V. Barbu, N. Limnios, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

Version française abrégée

Après avoir été introduits par Baum et Petrie [3], les modèles markoviens cachés (MMC) sont devenus très populaires dans des domaines très variés, comme la biologie [5], la reconnaissance du texte et de la parole [16], etc. Les processus markoviens cachés ont un désavantage important : les temps de séjour du processus caché suivent des lois géométriques ou exponentielles. Pour résoudre ce problème, Ferguson [9] a proposé un nouveau modèle qui permet d'avoir des temps de séjour arbitrairement répartis pour le processus caché.

À notre connaissance, tous les travaux existants sur l'estimation des modèles semi-markoviens cachés consistent seulement en des approches algorithmiques (voir [11,10,17]). Il n'existe pas de littérature concernant les propriétés asymptotiques théoriques de l'estimateur du maximum de vraisemblance des modèles semi-markoviens cachés, comme c'est le cas pour les modèles markoviens cachés.

Le travail présenté dans cet article porte sur les propriétés asymptotiques de l'estimateur du maximum de vraisemblance d'un modèle semi-markovien caché.

Soient $Z = (Z_n)_{n\in\mathbb{N}}$ une chaîne semi-markovienne homogène, d'espace d'état fini $E = \{1, \ldots, s\}$, et $(J, S) = (J_n, S_n)_{n\in\mathbb{N}}$ la chaîne de renouvellement markovien associée [1]. Notons $q(k) := (q_{ij}(k);\ i, j \in E)$ le noyau semi-markovien, $f(k) := (f_{ij}(k);\ i, j \in E)$ les lois conditionnelles de temps de séjour et $p := (p_{ij})_{i,j\in E}$ la matrice de transition de la chaîne de Markov $(J_n)_{n\in\mathbb{N}}$.

E-mail addresses: vlad.barbu@dma.utc.fr (V. Barbu), nikolaos.limnios@utc.fr (N. Limnios).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.12.013

Supposons que la chaîne Z n'est pas directement observée. Les observations sont décrites par une suite $Y = (Y_n)_{n\in\mathbb{N}}$ de variables aléatoires conditionnellement indépendantes, à valeurs dans $A := \{1, \ldots, d\}$ fini. Posons $R = (R_{i;a};\ i \in E, a \in A)$ pour la loi conditionnelle de la suite Y, $R_{i;a} := \mathbb{P}(Y_n = a \mid Z_n = i)$. La chaîne $(Z, Y) = (Z_n, Y_n)_{n\in\mathbb{N}}$ est appelée *chaîne semi-markovienne cachée*, ou, plus précisément, *chaîne cachée SM1–M0*.

Soit $U = (U_n)_{n\in\mathbb{N}}$ le temps de récurrence en arrière de la chaîne semi-markovienne $(Z_n)_{n\in\mathbb{N}}$, $U_n := n - S_{N(n)}$. La chaîne $(Z_n, U_n)_{n\in\mathbb{N}}$ est une chaîne de Markov d'espace d'état $E \times \mathbb{N}$ (voir [14]). Notons $\tilde{p} := (p_{(i,t_1)(j,t_2)})_{i,j\in E,\ t_1,t_2\in\mathbb{N}}$ sa matrice de transition.

Nous allons supposer dans le reste de cet article que pour tout $i, j \in E$, les lois conditionnelles des temps de séjour $f_{ij}(\cdot)$ ont des supports finis. Par conséquent, nous pouvons choisir un ensemble fini d'entiers positifs $D := \{0, \ldots, \tilde{n}\} \subset \mathbb{N}$ tel que $\operatorname{supp} f_{ij}(\cdot) \subset D$, pour tout $i, j \in E$. Grâce à la condition d'indépendance conditionnelle (2), nous avons associé la chaîne markovienne cachée $M1$–$M0$ $((Z_n, U_n), Y_n)_{n\in\mathbb{N}}$ à la chaîne semi-markovienne cachée initiale. Soit $\Theta := \mathcal{P} \times \mathcal{R}$ l'espace de paramètres du modèle markovien caché, où $\mathcal{P} := \{\tilde{p} = (p_{(i,t_1)(j,t_2)})_{i,j\in E,\ t_1,t_2\in D} \in \mathcal{M}_{s(\tilde{n}+1)\times s(\tilde{n}+1)} \mid \tilde{p}$ matrice stochastique$\}$ et $\mathcal{R} := \{R = (R_{ia})_{i\in E,\ a\in A} \in \mathcal{M}_{s\times d} \mid R_{ia} > 0$, $\sum_{a\in A} R_{ia} = 1\}$. Notons θ^0 la vraie valeur du paramètre et $\{Y_0 = y_0, \ldots, Y_M = y_M\}$ un échantillon des observations.

Théorème 0.1. *Sous des hypothèses de régularité* (*voir* A1 *et* A2), *l'estimateur du maximum de vraisemblance* $\hat{\theta}(M)$ *est fortement convergent vers* θ^0, *lorsque* M *tend vers l'infini.*

À partir de ce théorème, nous obtenons la convergence des estimateurs du maximum de vraisemblance des valeurs vraies du noyau semi-markovien $(q^0_{ij}(k))_{i,j\in E,\ k\in D}$, de la matrice de transition de la chaîne immergée $(p^0_{i,j})_{i,j\in E}$ et de la loi conditionnelle $(R^0_{ia})_{i\in E,\ a\in A}$ de Y.

Soit $\sigma(\theta^0)$ la matrice d'information de Fisher asymptotique, calculée dans θ^0 (voir Éq. (3)).

Théorème 0.2. *Sous des hypothèses de régularité* (*voir* A1, A2 *et* A3), *le vecteur aléatoire*

$$\sqrt{M}\left[\left(\widehat{R}_{ia}(M)\right)_{i\in E,\ a\in A} - (R_{ia})_{i\in E,\ a\in A}\right]$$

est asymptotiquement normal, lorsque $M \to \infty$, *d'espérance nulle et de matrice de covariance* $\sigma(\theta^0)^{-1}_{22}$ *donnée par l'Éq.* (5).

Théorème 0.3. *Sous des hypothèses de régularité, le vecteur aléatoire* (*voir* A1, A2 *et* A3)

$$\sqrt{M}\left[\left(\tilde{q}_{ij}(k, M)\right)_{i,j\in E,\ k\in D} - \left(q^0_{ij}(k)\right)_{i,j\in E,\ k\in D}\right]$$

est asymptotiquement normal, lorsque $M \to \infty$, *d'espérance nulle et de matrice de covariance* σ_q *donnée par l'Éq.* (7).

1. Introduction and preliminaries

After having been introduced by Baum and Petrie [3], the hidden Markov models (HMM) have become very popular in a wide range of applications, such as biology [5], speech and text recognition [16], etc.

In parallel with the extensive use of HMM in applications and with the related statistical inference work (see [3,13,4,15,12,7,8]), a new type of model is derived, initially in the domain of speech recognition. In fact, the hidden Markov processes has an important disadvantage, namely the fact that the sojourn times of the hidden process are geometrically or exponentially distributed. In order to solve this problem, Ferguson [9] was the first to propose a new model, which allows arbitrary sojourn time distributions for the hidden process. As the unobserved process becomes semi-Markovian, this model is called a hidden semi-Markov model (HSMM). Recent papers proposed computational techniques for HSMM (see [11,10,17]). To the best of our knowledge, all the existing works in HSMM area consist in the construction of different types of models and in the design of adapted algorithms (usually of EM type). Therefore, we do not have in the HSMMs' literature studies of asymptotic properties of the maximum likelihood estimator (MLE), like it is the case in the HMM context. In the sequel we investigate the asymptotic properties of the MLE for a HSMM, following the line of [4] and [3].

Firstly, let us set the basic notations of a discrete-time semi-Markov model. For more details, see [1,2] or [14]. Let $Z = (Z_n)_{n\in\mathbb{N}}$ be an homogeneous semi-Markov chain with finite state space $E = \{1, \ldots, s\}$. Put $(J, S) = (J_n, S_n)_{n\in\mathbb{N}}$

for the associated Markov renewal chain and $X=(X_n)_{n\in\mathbb{N}^*}$ for the inter-jumps chain. Thus, for all $n\in\mathbb{N}$ we have $Z_n=J_{N(n)}$, where $N(k):=\max\{n\geqslant 0 \mid S_n\leqslant k\}$ is the discrete-time counting process of the number of jumps in $[1,k]\subset\mathbb{N}$.

For n given sets $E_1,\dots,E_n$, we denote by $\mathcal{M}_{E_1\times\cdots\times E_n}$ the set of nonnegative n-dimensional arrays on $E_1\times\cdots\times E_n$ and by $\mathcal{M}_{E_1}(\mathbb{N})$ the set of matrix-valued functions : $\mathbb{N}\to\mathcal{M}_{E_1\times E_1}$.

Let us denote by $q(k):=(q_{ij}(k);\ i,j\in E)\in\mathcal{M}_E(\mathbb{N})$ the discrete semi-Markov kernel ($q_{ij}(k):=\mathbb{P}(J_{n+1}=j, X_{n+1}=k\mid J_n=i)$), by $p:=(p_{ij};\ i,j\in E)\in\mathcal{M}_E$ the transition matrix of the embedded Markov chain $(J_n)_{n\in\mathbb{N}}$ ($p_{ij}:=\mathbb{P}(J_{n+1}=j\mid J_n=i)$) and by $f(k):=(f_{ij}(k);\ i,j\in E)\in\mathcal{M}_E(\mathbb{N})$ the conditional distribution of sojourn times ($f_{ij}(k)=\mathbb{P}(X_{n+1}=k\mid J_n=i,\ J_{n+1}=j)$). We suppose that $q_{ij}(k)=0$ if $i=j$ or $k=0$. The following assumption is needed in the sequel:

A1. The Markov chain $(J_n)_{n\in\mathbb{N}}$ is irreducible.

2. Results

Suppose that we have a stationary semi-Markov chain Z, which is not directly observed. The observations are described by a chain $Y=(Y_n)_{n\in\mathbb{N}}$, with the state space $A:=\{1,\dots,d\}$. The process $Y=(Y_n)_{n\in\mathbb{N}}$ is a sequence of conditionally independent random variables, i.e., for all $a\in A,\ j\in E,\ n\in\mathbb{N}^*$, the following relation holds true

$$\mathbb{P}(Y_n=a\mid Y_{n-1}=\cdot,\dots,Y_0=\cdot,Z_n=i,\dots,Z_0=\cdot)=\mathbb{P}(Y_n=a\mid Z_n=i). \tag{1}$$

Define $R=(R_{i;a})_{i\in E,\ a\in A}\in\mathcal{M}_{E\times A}$ the conditional distribution of the chain Y, $R_{i;a}:=\mathbb{P}(Y_n=a\mid Z_n=i)$.

The process $(Z,Y)=(Z_n,Y_n)_{n\in\mathbb{N}}$ is called *a hidden semi-Markov chain*, or, more precisely, *a hidden SM1–M0 chain*, where the index 1 stands for the order of the semi-Markov chain Z and the index 0 stands for the order of Y, regarded as a Markov chain.

2.1. Consistency of MLE

Let $U=(U_n)_{n\in\mathbb{N}}$ be the backward-recurrence times of the semi-Markov chain $(Z_n)_{n\in\mathbb{N}}$, $U_n:=n-S_{N(n)}$. The chain $(Z_n,U_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $E\times\mathbb{N}$ (see [14]). We denote by $\tilde{p}:=(p_{(i,t_1)(j,t_2)})_{i,j\in\mathbf{E},\ t_1,t_2\in\mathbb{N}}$ the transition matrix of the Markov chain (Z,U).

Reduced state space: We shall consider that all the conditional distributions of sojourn times, $f_{ij}(\cdot),\ i,j\in E$, have finite support. Consequently, we can choose a finite set of integers $D:=\{0,\dots,\tilde{n}\}$ such that $\operatorname{supp} f_{ij}(\cdot)\subseteq D$, for all $i,j\in E$.

Note that, especially from a practical point of view, there is only little loss of generality which follows from this assumption, since in applications we always take into account just a finite support for a distribution. In conclusion, all the work in the rest of the paper will be done under the assumption:

A2. The conditional sojourn time distributions have finite support $\subset D$.

Taking into account the conditional independence relation (1), for all $a\in A$, $j\in E$ and $t\in D$ we have

$$R_{i;a}=\mathbb{P}(Y_n=a\mid Z_n=i)=\mathbb{P}(Y_n=a\mid Z_n=i,\ U_n=t). \tag{2}$$

Consequently, starting from the initial hidden semi-Markov chain $(Z_n,Y_n)_{n\in\mathbb{N}}$, we have an associated hidden Markov chain $((Z_n,U_n),Y_n)_{n\in\mathbb{N}}$. Let $\Theta:=\mathcal{P}\times\mathcal{R}$ be the parameter space of the hidden Markov model, where $\mathcal{P}:=\{\tilde{p}=(p_{(i,t_1)(j,t_2)})_{i,j\in E,\ t_1,t_2\in D}\in\mathcal{M}_{s(\tilde{n}+1)\times s(\tilde{n}+1)}\mid \tilde{p}$ stochastic matrix$\}$ and $\mathcal{R}:=\{R=(R_{ia})_{i\in E,\ a\in A}\in\mathcal{M}_{s\times d}\mid R_{ia}>0,\ \sum_{a\in A}R_{ia}=1\}$. Note θ^0 the true value of the parameter.

For technical reasons, we consider that the time scale of all the processes is $\mathbb{Z}$ instead of $\mathbb{N}$. In conclusion, our work is carried out within the framework of [3] and we have the following results.

Theorem 2.1. *Under assumptions* A1 *and* A2, *given a sample path of observations* $\{Y_0=y_0,\dots,Y_M=y_M\}$, *the MLE* $\widehat{\theta}(M)$ *of* θ^0 *is strongly consistent, as* M *tends to infinity.*

We stress the fact that the convergence in the preceding theorem is in the quotient topology, in order to render the model identifiable. Thus, we have obtained the consistency of the maximum likelihood estimators of $(R^0_{ia})_{i\in E,\ a\in A}$

and $(p^0_{(i,t_1)(j,t_2)})_{i,j\in E,\ t_1,t_2\in D}$, denoted by $(\widehat{R}_{ia}(M))_{i\in E,\ a\in A}$ and $(\widehat{p}_{(i,t_1)(j,t_2)}(M))_{i,j\in E,\ t_1,t_2\in D}$. The following theorem uses these results in order to prove the consistency of the MLE of $(q^0_{ij}(k))_{i,j\in E,\ k\in D}$ and of $(p^0_{i,j})_{i,j\in E}$.

Theorem 2.2. *Under assumptions* A1 *and* A2, *given a sample path of observations* $\{Y_0 = y_0, \ldots, Y_M = y_M\}$, *the MLEs* $(\widehat{q}_{ij}(k, M))_{i,j\in E,\ k\in D}$ *and* $(\widehat{p}_{i,j}(M))_{i,j\in E}$ *of* $(q^0_{ij}(k))_{i,j\in E,\ k\in D}$ *and of* $(p^0_{i,j})_{i,j\in E}$ *are strongly consistent, as M tends to infinity.*

2.2. Asymptotic normality of MLE

For $\{Y_0 = y_0, \ldots, Y_n = y_n\}$ a sample path of observations, denote by $p_\theta(Y_0^n)$ the associated likelihood function and by $\sigma_{Y_0^n}(\theta^0) := -\mathbb{E}_{\theta^0}(\frac{\partial^2 \log p(Y_0^n)}{\partial\theta_u\partial\theta_v}|_{\theta=\theta^0})_{u,v}$ the Fisher information matrix computed in θ^0.

Let

$$\sigma(\theta^0) = (\sigma_{u,v}(\theta^0))_{u,v} := -\mathbb{E}_{\theta^0}\left(\frac{\partial^2 \log \mathbb{P}_\theta(Y_0 \mid Y_{-1}, Y_{-2}, \ldots)}{\partial\theta_u\partial\theta_v}\bigg|_{\theta=\theta^0}\right)_{u,v} \tag{3}$$

be the asymptotic Fisher information matrix computed in θ^0 (see [3] or [8] for the definition of $\sigma(\theta^0)$ as a limiting matrix of Fisher information matrices).

From Theorem 3 of [6] we know that $\sigma(\theta^0)$ is nonsingular if and only if there exists an integer $n \in \mathbb{N}$ such that $\sigma_{Y_0^n}(\theta^0)$ is nonsingular. Consequently, all our work will be done under the following assumption:

A3. There exists an integer $n \in \mathbb{N}$ such that the matrix $\sigma_{Y_0^n}(\theta^0)$ is nonsingular.

Theorem 2.3. *Under assumptions* A1, A2 *and* A3, *the random vector*

$$\sqrt{M}\left[(\widehat{R}_{ia}(M))_{i\in E,\ a\in A} - (R^0_{ia})_{i\in E,\ a\in A}\right] \tag{4}$$

is asymptotically normal, as $M \to \infty$, *with zero mean and covariance matrix* $\sigma(\theta^0)^{-1}_{22}$, *where we have considered the following partition of the matrix* $\sigma(\theta^0)^{-1}$:

$$\sigma(\theta^0)^{-1} = \begin{array}{c} \overbrace{\quad}^{s^2\times(\tilde{n}+1)^2} \quad \overbrace{\quad}^{sd} \\ \begin{bmatrix} \sigma(\theta^0)^{-1}_{11} & \sigma(\theta^0)^{-1}_{12} \\ \sigma(\theta^0)^{-1}_{21} & \sigma(\theta^0)^{-1}_{22} \end{bmatrix} \end{array} \begin{array}{l} \} \ s^2\times(\tilde{n}+1)^2 \\ \} \ sd \end{array}. \tag{5}$$

Before giving the result concerning the asymptotic normality of the semi-Markovian kernel estimator, we need to introduce some notations. For all $k = 2, \ldots, \tilde{n}$ and $i \in E$, let us consider the scalars $\alpha_{i;k} := \prod_{t=0}^{k-2} p_{(i,t)(i,t+1)}$ and the matrices $B_{i;k} \in \mathcal{M}_{(s-1)\times(\tilde{n}-1)}$, whose elements $B_{i;k}(u, v)$, $u = 1, \ldots, s-1$, $v = 1, \ldots, \tilde{n}-1$, are defined by

$$B_{i;k}(u,v) := \begin{cases} 0 & \text{if } v > k-1, \\ p_{(i,k-1)(u,0)} \prod_{t=0}^{k-2} p_{(i,t)(i,t+1)} / p_{(i,v-1)(i,v)} & \text{if } v \leqslant k-1, u < i, \\ p_{(i,k-1)(u+1,0)} \prod_{t=0}^{k-2} p_{(i,t)(i,t+1)} / p_{(i,v-1)(i,v)} & \text{if } v \leqslant k-1, u \geqslant i. \end{cases}$$

Let us put $\tilde{\sigma} := \sigma(\theta^0)^{-1}$. We consider the following partition of the matrix $\tilde{\sigma}$:

$$\left[\begin{array}{cccccc} \overbrace{\tilde{\sigma}_{1,1}}^{s-1} & \ldots & \overbrace{\tilde{\sigma}_{1,s\tilde{n}}}^{s-1} & \overbrace{\tilde{\sigma}_{1,s\tilde{n}+1}}^{\tilde{n}-1} & \ldots & \overbrace{\tilde{\sigma}_{1,s\tilde{n}+s}}^{\tilde{n}-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\sigma}_{s\tilde{n},1} & \ldots & \tilde{\sigma}_{s\tilde{n},s\tilde{n}} & \tilde{\sigma}_{s\tilde{n},s\tilde{n}+1} & \ldots & \tilde{\sigma}_{s\tilde{n},s\tilde{n}+s} \quad \tilde{\sigma}_A \\ \tilde{\sigma}_{s\tilde{n}+1,1} & \ldots & \tilde{\sigma}_{s\tilde{n}+1,s\tilde{n}} & \tilde{\sigma}_{s\tilde{n}+1,s\tilde{n}+1} & \ldots & \tilde{\sigma}_{s\tilde{n}+1,s\tilde{n}+s} \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\sigma}_{s\tilde{n}+s,1} & \ldots & \tilde{\sigma}_{s\tilde{n}+s,s\tilde{n}} & \tilde{\sigma}_{s\tilde{n}+s,s\tilde{n}+1} & \ldots & \tilde{\sigma}_{s\tilde{n}+s,s\tilde{n}+s} \\ & & & \tilde{\sigma}_B & & \quad\quad \tilde{\sigma}_C \end{array}\right] \begin{array}{l} \} \ s-1 \\ \vdots \\ \} \ s-1 \\ \} \ \tilde{n}-1 \\ \vdots \\ \} \ \tilde{n}-1 \\ \\ \end{array}.$$

Theorem 2.4. *Under assumptions* A1, A2 *and* A3, *the random vector*

$$\sqrt{M}\left[\left(\tilde{q}_{ij}(k,M)\right)_{i,j\in E,\ k\in D}-\left(q^0_{ij}(k)\right)_{i,j\in E,\ k\in D}\right] \tag{6}$$

is asymptotically normal, as $M\to\infty$, *with zero mean and covariance matrix* σ_q *given by*

$$\sigma_q=\begin{bmatrix} M_{1,1} & \dots & M_{1,\tilde{n}} \\ \vdots & & \vdots \\ M_{\tilde{n},1} & \dots & M_{\tilde{n},\tilde{n}} \end{bmatrix}, \tag{7}$$

where the block matrices $M_{u,v}\in\mathcal{M}_{s(s-1)\times s(s-1)}$ *are as follows*:

$$M_{1,1}=\begin{bmatrix} \tilde{\sigma}_{1,1} & \dots & \tilde{\sigma}_{1,s} \\ \vdots & & \vdots \\ \tilde{\sigma}_{s,1} & \dots & \tilde{\sigma}_{s,s} \end{bmatrix},$$

$$\text{for } 2\leqslant v\leqslant\tilde{n},\quad M_{1,v}=\begin{bmatrix} \alpha_{1;v}\tilde{\sigma}_{1,sv-s+1}+\tilde{\sigma}_{1,s\tilde{n}+1}B^\top_{1;v} & \dots & \alpha_{s;v}\tilde{\sigma}_{1,sv}+\tilde{\sigma}_{1,s\tilde{n}+s}B^\top_{s;v} \\ \vdots & & \vdots \\ \alpha_{1;v}\tilde{\sigma}_{s,sv-s+1}+\tilde{\sigma}_{s,s\tilde{n}+1}B^\top_{1;v} & \dots & \alpha_{s;v}\tilde{\sigma}_{s,sv}+\tilde{\sigma}_{s,s\tilde{n}+s}B^\top_{s;v} \end{bmatrix},$$

$$\text{for } 2\leqslant u\leqslant\tilde{n},\quad M_{u,1}=\begin{bmatrix} \alpha_{1;u}\tilde{\sigma}_{su-s+1,1}+B_{1;u}\tilde{\sigma}_{s\tilde{n}+1,1} & \dots & \alpha_{1;u}\tilde{\sigma}_{su-s+1,s}+B_{1;u}\tilde{\sigma}_{s\tilde{n}+1,s} \\ \vdots & & \vdots \\ \alpha_{s;u}\tilde{\sigma}_{su,1}+B_{s;u}\tilde{\sigma}_{s\tilde{n}+s,1} & \dots & \alpha_{s;u}\tilde{\sigma}_{su,s}+B_{s;u}\tilde{\sigma}_{s\tilde{n}+s,s} \end{bmatrix},$$

for $2\leqslant u,v\leqslant\tilde{n}$, $M_{u,v}=(M_{u,v}(l,r))_{l,r=1,\dots,s}$ *is a block matrix, having the matrix*

$$M_{u,v}(l,r):=(\alpha_{l;u}\tilde{\sigma}_{su-s+l,sv-s+r}+B_{l;u}\tilde{\sigma}_{s\tilde{n}+l,sv-s+r})\alpha_{r;v}+(\alpha_{l;u}\tilde{\sigma}_{su-s+l,s\tilde{n}+r}+B_{l;u}\tilde{\sigma}_{s\tilde{n}+l,s\tilde{n}+r})B^\top_{r;v}$$

on the position (l,r), $1\leqslant l,r\leqslant s$.

References

[1] V. Barbu, M. Boussemart, N. Limnios, Discrete time semi-Markov model for reliability and survival analysis, Comm. Statist. Theory Methods 33 (11) (2004) 2833–2868.

[2] V. Barbu, N. Limnios, Discrete time semi-Markov processes for reliability and survival analysis – a nonparametric estimation approach, in: M. Nikulin, N. Balakrishnan, M. Meshbah, N. Limnios (Eds.), Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis and Quality of Life, Birkhäuser, Boston, 2004, pp. 487–502.

[3] L.E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Statist. 37 (1966) 1554–1563.

[4] P.J. Bickel, Y. Ritov, T. Rydén, Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models, Ann. Statist. 26 (1998) 1614–1635.

[5] G. Churchill, Hidden Markov chains and the analysis of genome structure, Computers Chem. 16 (1992) 107–115.

[6] R. Douc, Non singularity of the asymptotic Fisher information matrix in hidden Markov models, École Polytechnique, Preprint.

[7] R. Douc, C. Matias, Asymptotics of the maximum likelihood estimator for general hidden Markov models, Bernoulli 7 (3) (2001) 381–420.

[8] R. Douc, E. Moulines, T. Rydén, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, Ann. Statist. 32 (5) (2004) 2254–2304.

[9] J.D. Ferguson, Variable duration models for speech, in: Proc. of the Symposium on the Application of Hidden Markov Models to Text and Speech, Princeton, NJ, 1980, pp. 143–179.

[10] Y. Guédon, Estimating hidden semi-Markov chains from discrete sequences, J. Comput. Graph. Statist. 12 (3) (2003) 604–639.

[11] Y. Guédon, C. Cocozza-Thivent, Explicit state occupancy modelling by hidden semi-Markov models: application of Derin's scheme, Computer Speech and Language 4 (1990) 167–192.

[12] J.L. Jensen, N.V. Petersen, Asymptotic normality of the maximum likelihood estimator in state space models, Ann. Statist. 27 (1999) 514–535.

[13] B.G. Leroux, Maximum-likelihood estimation for hidden Markov models, Stochastic Process. Appl. 40 (1992) 127–143.

[14] N. Limnios, G. Oprişan, Semi-Markov Processes and Reliability, Birkhäuser, Boston, 2001.

[15] F. Muri-Majoube, Comparaison d'algorithmes d'identification de Chaînes de Markov Cachées et application à la détection de régions homogènes dans les séquences d'ADN, Ph.D. Thesis, University Paris 5, 1997.

[16] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (1989) 257–284.

[17] J. Sansom, P.J. Thomson, Fitting hidden semi-Markov models to breakpoint rainfall data, J. Appl. Probab. 38A (2001) 142–157.