Consistent and asymptotically normal parameter estimates for Hidden Markov Mixtures of Markov Models

Pierre VANDEKERKHOVE

University of Marne-la-Vallée
Laboratoire d'Analyse et Mathématiques Appliquées

Abstract

In this paper we introduce a new missing data model, based on a K–mixture of Markov processes, and consider the general problem of its parametric identification. We point out in detail the main difficulties of the statistical inference for such models: complete likelihood calculation, stationary distribution parametrization, and identifiability. Nevertheless, we propose a general tractable approach to estimate these models (admitting stationary distribution parametrization and identifiability) and check in detail that our assumptions are fully satisfied for a 2–Markov mixture of linear AR(1) models with gaussian noises. Finally a Monte-Carlo method is proposed to calculate the split data likelihood of this model when analytic expression of the invariant probability densities of the Markov processes is unknown.

Abbreviated title. Mixed Markov Models estimation.

AMS 1991 classification. Primary 62M09, 62L20.

Key words and phrases. Hidden Markov chain, incomplete data, Markov chain, mixture, statistical inference.

1 Introduction

In signal processing, or statistics literature, different definitions of Mixtures of Markov Models (MMM) can be found. In the first topic, the study of MMMs is also associated to identification problem of mixed stationary, markovian or not, sources. Let consider an observable finite sequence of K-dimensionnal random variables, $X = (X_k)_{1\leq k\leq T}$, coming from an instantaneous mixture of K colored sources $S = (S_k)_{1\leq k\leq T}$, *i.e.*

$$\forall k = 1, \cdots, T, \quad X_k = AS_k \tag{1}$$

where A is a square and invertible matrix, called mixing matrix. The goal in such a framework is to recover the sources S from X, by estimating $B = A^{-1}$. See for example, Pham and Garat (1997), or Dégerine and Zaïdi (2002) for respectively, pseudo and exact likelihood approach in the case of mixtures of gaussian autoregressive (AR) sources, and detailed references.

In the second topic, see Jalali and Pemberton (1995), Wong and Li (2000) and (2001), Benesch (2001), the MMMs are defined from a distributional point of view. In the spirit of definition (2.1) for MAR (Mixed Autoregressive Model) in Wong and Li (2001), we should say that a process $X = (X_n)_{n\geq 1}$ is an MMM if, for all $n \geq d+1$, $F(x_n|\mathcal{F}_{n-1})$ the conditional cumulative distribution function of X_n given the past information, takes the form

$$F(x_n|\mathcal{F}_{n-1}) = \sum_{k=1}^{d} \alpha_k \Phi_k(x_n; x_{n-1}, \cdots, x_{n-d}; \vartheta_k), \tag{2}$$

with $\sum_{k=1}^{d} \alpha_k = 1$, $\alpha_k > 0$, for $k = 1, \cdots, d$, and where $\Phi_k(\cdot; x_{n-1}, \cdots, x_{n-d}; \vartheta_k)$ is a cumulative distribution function depending on parameters $(x_{n-1}, \cdots, x_{n-d}; \vartheta_k)$ (observations from the d-past, and a statistical parameter). According to Wong and Li (2000) the MAR is usefull to model times series with multimodal marginal or conditional distributions, see Tong (1990), and Chan and Tong (1998). Application to biological real data, about Canadian lynx, is done in these previous papers. Other models with smooth changes time varying have been proposed in econometry , and applied later to other areas: the so-called autoregressive processes with Markov regime, which dynamic is *mixed* by a Markov chain. Denoting by X such a process, one basic definition is

$$\forall\ n \geq d+1, \quad X_n = \sum_{i=1}^{d} a_i(U_n) X_{n-i} + \sigma(U_n)\varepsilon_n, \tag{3}$$

where $(\varepsilon)_{n\geq 1}$ is sequence of i.i.d. random variables, $U = (U_n)_{n\geq 1}$ is a Markov chain with continuous or discrete state space, and $(a_i(\cdot))_{i=1,\cdots,d}$, and $\sigma(\cdot)$ are functions defined on the state space of U. This model was used by Hamilton (1989) to model US gros national product (the U's modelling the economic/business cycles), see Hamilton and Susmel (1994), Cai (1994), Garcia and Perron (1996), for recent extensions. Linear autoregressive processes with

Markov regime are also widely used in electrical engineering, see Bar-Shalom and Li (1993), failure detection, see Tugnait (1982), automatic control Ji *et al.* (1993), Kryshnamurty and Rydén (1998). An other important class of autoregressive Markov processes with Markov regime are the Hidden Markov Models (HMMs), for which the conditional distribution of X_n does not depend on lagged X's but only on U_n. HMMs are used in many different areas, including speech recognition, see Juang an Rabiner (1991), neurophysiology, see Fredkin and Rice (1987), econometrics, see Chib *et al.* (1998). Most works on maximum likelihood estimation in these models have focused on numerical method for approximation of the maximum likelihood estimator (MLE). In sharp contrast, statistical issue on asymptotic poperties of the MLE for HMMs and autoregressive processes with Markov regime took a long time in obtaining significative overhangs. Concerning HMMs, see Baum and Petrie (1966), Leroux (1992), Bakry *et al.* (1997), Bickel *et al.* (1998), LeGland and Mevel (2000), Douc and Matias (2000), and about autoregressive processes with Markov regime, see also Krisnamurthy and Rydén (1998), and Francq and Roussignol (1998), when U takes value in a finite set, and Douc *et al.* (2004), when U takes value in a continuous state space.

Let us observe finally that probabilistic works on characterization of mixtures of Markov chains distribution, have been initiated by de Finetti (1959), and continued, among many others, by Freedman (1962), Diaconis and Freedman (1980) or more recently by Fortini *et al.* (2002).

In this paper we introduce, an other possible definition of MMM. Let us consider $X^{[i]} = (X_n^{[i]})_{n\geq 1}$, $1 \leq i \leq K$, K independent stationary discrete time Markov processes taking values in a measurable state space $(E, \mathcal{E})$ with respectively probability transition densities Q^i, $1 \leq i \leq K$, with respect to a common finite dominating measure λ. The MMM we consider, induces an observed process $Z = (Z_n)_{n\geq 0}$ based on the collection of the K mutually independent processes $X^{[i]} = (X^{[i]})_{n\geq 0}$, and defined by:

$$\forall n \geq 1, \quad Z_n = \sum_{i=1}^{K} \mathbf{1}_{\{U_n = i\}} X_n^{[i]}, \tag{4}$$

where $(U_n)_{n\geq 0}$ is a stationary positive recurrent Markov chain valued in $\mathcal{U} = \{1, \cdots, K\}$. We suppose in addition that the chain $(U_n)_{n\geq 0}$ is not observed, which corresponds to a situation where mixtures of sample path (due to a markovian process selection) coming from independent markov sources are only observed. To differentiate this model from other MMMs, we propose to call it: Hidden Markov Mixture of Markov Models (HMMMM, or shorter H4M). Let us remark that our MMM is not Markovian, and is clearly different from other MMMs. On the other hand it is good to notice that Hidden Markov Models (HMMs) belong to the class of H4Ms. To check this point it is enough to consider independent sequences for the $X^{[i]}$'s in (4). Notice at this step that HMMs are at the corner of H4Ms and the class of autoregressive models with Markov regime, when the underlying Markov chain U is supposed to belong to a finite state space. From the previous remark the H4Ms are naturaly dedicated to find applications in areas where HMMs hold, see the previous paragraph given over to HMMs, and section 6.

These preliminary remarks being done, we want to point out the ability of our model to describe discrete time series with: i) abrupt changes; when U switchs of state ii) local stationarity; during stages where U remains in the same state iii) multimodal marginal distributions; from mixture structure, and iv) *phase-type feedback* effects, see Neuts (1994, p.46) for definition of phase-type distribution. Let precise point iv). Consider two sample paths of length $n \geq 3$: $u_1^n = (u_1, \cdots, u_n)$ from U and $z_1^n = (z_1, \cdots, z_n)$ from Z, and fix $u_{n+1} = i$. Let suppose that there exists an index $\tau_n \geq 2$ such that $u_{n+1-\tau_n} = i$, $u_{n+1-k} \neq i$ for all $k = 1, \cdots, \tau_n - 1$, *i.e.* corresponding to the time separating the current observation of U at state i and the last observed value of U at state i. From the definition of Z, we can check that the conditional law $\mathcal{L}(Z_{n+1}|U_1^{n+1} = u_1^{n+1}, Z_1^n = z_1^n)$, satisfy:

$$\mathcal{L}(Z_{n+1}|U_1^{n+1} = u_1^{n+1}, Z_1^n = z_1^n) = \mathcal{L}(Z_{n+1}|U_{n+1-\tau_n}^{n+1} = u_{n+1-\tau_n}^{n+1}, Z_{n+1-\tau_n} = z_{n+1-\tau_n}), \tag{5}$$

which only depends on i, τ_n, and $z_{n-\tau_n}$, from independence of the $X^{[i]}$'s, and their markovian structure. Writting (5) shows that the law of the process Z at time $n+1$, is precoloured by a particular observation of its past which index time is unknown (feedback effect). Let us add a last point v) quasi independence of the homogeneous phases; when long stages of U in the same states occur and if the $X^{[i]}$'s are strongly mixing.

The goal of this paper is now to propose a $\sqrt{n}$-consistent method, based on the maximum split data likelihood estimate (MSDLE) introduced by Rydén (1994), to estimate the parameters driving the transition density kernels of the $X^{[i]}$'s, and the transition matrix of U. The sequel of this paper is organized as follows. In section 2 we present precisely the MSDLE for H4Ms, and the main assumptions (see also ergodicity, and identifiability). In section 3 we prove consistency and asymptotic normality of the MSDLE under mild conditions. In section 4 we propose a Monte-Carlo approach to estimate the log of the split data likelihood (SDL), when analytical expression of the $X^{[i]}$'s invariant density is unknown under fixed parametrization of the transitions. Section 5 is devoted to a detailed study of a hidden Markov mixture of two linear autoregressive processes of order 1. In section 6, we indicate some possible applications for the H4Ms, in areas like neurophysiology and kinetics. The last paragraphe of section is given over to sample path simulations of different HMM and H4Ms with same marginal distribution, and same underlying markov chain U. A short empirical comparison of the obtained patterns is made, and similarities with Alpha and Theta waves found in kinetics are noticed.

2 Assumptions and parametrization

To alleviate the notations, and without loss of generality, we propose to consider the case $K = 2$, and to denote by $X = X^{[1]}$ and $Y = X^{[2]}$. The transition density kernels of X and Y

will be parametrized by $\theta \in \Phi^1$ for Q^1 and $\phi \in \Phi^2$ for Q^2, whith Φ^i for $i = 1, 2$, compact sets of $\mathbb{R}^q$, and are assumed to belong respectively to the parametric families $\mathcal{K}^1 = \{Q^1_\theta(\cdot,\cdot), \theta \in \Phi^1\}$, and $\mathcal{K}^2 = \left\{Q^2_\phi(\cdot,\cdot), \phi \in \Phi^2\right\}$. We suppose that for each $\theta \in \Phi^1$ (resp. $\phi \in \Phi^2$) the probability transition kernel Q^1_θ (resp. Q^2_ϕ) induces a recurrent positive Markov process, and admits a unique invariant probability measure with density q^1_θ (resp. q^2_ϕ). Notice that in general, analytic expressions of these densities are not explicitly known except in the case of linear autoregressive models with gaussian noises, see section 5 and 6. Nevertheless let us recall that for each $\theta \in \Phi^1$ and $\phi \in \Phi^2$, q^1_θ and q^2_ϕ are the unique solutions of the functional fixed point problems

$$\int_E q^1_\theta(x_1) Q^1_\theta(x_1,\cdot)\lambda(dx_1) = q^1_\theta(\cdot), \quad \text{and} \quad \int_E q^2_\phi(y_1) Q^2_\phi(y_1,\cdot)\lambda(dy_1) = q^2_\phi(\cdot). \tag{6}$$

The transition matrix Π of U will be parametrised by $\gamma = (\alpha, \beta) \in [\delta, 1-\delta]^2$, with $0 < \delta < 1$, as follows

$$\Pi_\gamma = \begin{pmatrix} \pi_\gamma(1,1) & \pi_\gamma(1,2) \\ \pi_\gamma(2,1) & \pi_\gamma(2,2) \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}. \tag{7}$$

The invariant probability vector associated to Π_γ is denoted by $(\pi_\gamma(1), \pi_\gamma(2)) = (\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$. Finally the *global* parameter which is to be estimated can be written as $\vartheta = (\gamma; \theta, \phi) \in \Theta = [\delta, 1-\delta]^2 \times \Phi^1 \times \Phi^2$. From now on we use the notation $\mathbf{Z}^n_1 = \{Z_k \ ; \ 1 \le k \le n\}$ for all processes. Let suppose that U should be observed and consider $u^n_1 = (u_1, \cdots, u_n)$ a sample path of length n from U , and $z^n_1 = (z_1, \cdots, z_n)$ a sample path of length n from Z. In that case the likelihood function for (U, Z) based on (u^n_1, z^n_1) can be written as $p_\vartheta(u^n_1, z^n_1) = p_\vartheta(z^n_1|u^n_1)p_\vartheta(u^n_1)$, where $p_\vartheta(u^n_1) = P_\vartheta(\mathbf{U}^n_1 = u^n_1)$, and $p_\vartheta(z^n_1|u^n_1)$ denotes the density of the Z's conditionally on $\{\mathbf{U}^n_1 = u^n_1\}$, which expressions are respectively given by:

$$p_\vartheta(u^n_1) = \pi_\gamma(u_1) \prod_{j=1}^{n-1} \pi_\gamma(u_j, u_{j+1}),$$

and

$$\begin{aligned} p_\vartheta(z^n_1|u^n_1) \;\; = \;\; & \int_{E^n} q^1_\theta(x_1) \prod_{j=1}^{n-1} Q^1_\theta(x_j, x_{j+1}) \otimes_{j\in\{1,\cdots,n\}/u_j=2} \lambda(dx_j) \otimes_{j\in\{1,\cdots,n\}/u_j=1} \delta_{z_j}(dx_j) \\ & \times \int_{E^n} q^2_\phi(y_1) \prod_{j=1}^{n-1} Q^2_\phi(y_j, y_{j+1}) \otimes_{j\in\{1,\cdots,n\}/u_j=1} \lambda(dy_j) \otimes_{j\in\{1,\cdots,n\}/u_j=2} \delta_{z_j}(dx_j), \end{aligned}$$

where we recognize the joint density of the independent random vectors $\mathbf{X}^n_1$ and $\mathbf{Y}^n_1$, integrated componentwise when U is not in state 1, resp. U is not in state 2. To compute the likelihood function for the Z's alone, it remains to sum $p_\vartheta(u_1^n, z_1^n)$ over all the possible values of u_1^n, then it comes

$$\begin{aligned} p_\vartheta(z_1^n) &= \sum_{(u_1,u_2,\cdots,u_n)\in\{1,2\}^n} \pi_\gamma(u_1)\prod_{j=1}^{n-1}\pi_\gamma(u_j,u_{j+1}) \\ &\times \int_{E^n} q_\theta^1(x_1)\prod_{j=1}^{n-1} Q_\theta^1(x_j,x_{j+1}) \otimes_{j\in\{1,\cdots,n\}/u_j=2}\lambda(dx_j)\otimes_{j\in\{1,\cdots,n\}/u_j=1}\delta_{z_j}(dx_j) \\ &\times \int_{E^n} q_\phi^2(y_1)\prod_{j=1}^{n-1} Q_\phi^2(y_j,y_{j+1}) \otimes_{j\in\{1,\cdots,n\}/u_j=1}\lambda(dy_j)\otimes_{j\in\{1,\cdots,n\}/u_j=2}\delta_{z_j}(dy_j). \end{aligned} \tag{8}$$

Let us remark that, contrary to discrete HMMs, the likelihood of H4Ms do not benefit of recurrence formula based on the filter, since successive Z_i's are not independent conditionally on a finite past of U. This technique allows surprisingly to compute the huge likelihood of HMMs in linear times (with respect to n), see Rabiner (1989). For this reason, and the very big complexity of the likelihood function of H4Ms, we rather propose, in a first step, to consider a maximum split data likelihood estimate (MSDLE) in the spirit of Rydén (1994), instead of the very hard to handle maximum likelihood estimate (MLE).

For an integer m conveniently choosen, we define the m-dimensional MSDLE based on $\mathbf{Z}_1^{km}$, with $k \geq 1$, as follows

$$\hat\vartheta_k = \arg\max_{\vartheta\in\Theta}\prod_{j=1}^{k} p_\vartheta(\mathbf{Z}_{(j-1)m+1}^{jm}). \tag{9}$$

The true parameter value will be denoted by ϑ_0, the law of Z over $E^{\mathbb{N}^*}$ will be denoted for simplicity by P_0, index 0 recalling that ϑ_0 defines entirely the law of Z, and expectation under P_0 will be denoted by $E_0(\cdot)$. The following conditions will be used throughout the paper.

C1. The true parameter ϑ_0 is an interior point of Θ, compact set of $\mathbb{R}^{2q+2}$.

C2. The parametric family $\mathcal{F}^m = \{p_\vartheta(z_1,\cdots,z_m)\ ;\ \vartheta\in\Theta\}$ is identifiable in the sense:

$$\forall(\vartheta,\vartheta')\in\Theta^2\ /\ p_\vartheta(z_1,\cdots,z_m) = p_{\vartheta'}(z_1,\cdots,z_m)\quad \lambda^{\otimes m}-a.e. \Rightarrow \vartheta=\vartheta'.$$

C3. There exist two functions g_1 and g_2 from E^m into $\mathbb{R}$ such that

$$g_1(z_1,\cdots,z_m) \leq p_\vartheta(z_1,\cdots,z_m) \leq g_2(z_1,\cdots,z_m),\quad \forall(z_1,\cdots,z_m;\vartheta)\in E^m\times\Theta,$$

$$\text{and}\qquad \int_{E^m} |\log(g_i(z_1,\cdots,z_m))| p_{\vartheta_0}(z_1,\cdots,z_m)\lambda(dz_1^m) < M,\quad i=1,2.$$

C4. The function $\vartheta\mapsto p_\vartheta(z_1,\cdots,z_m)$ is $\lambda^{\otimes m}$–$a.e.$ twice differentiable on Θ.

C5. Write $\vartheta = (\alpha, \beta; \theta_1, \cdots, \theta_q; \phi_1, \cdots, \phi_q) = (\vartheta_1, \vartheta_2; \vartheta_3, \cdots, \vartheta_{q+2}; \vartheta_{q+3}, \cdots, \vartheta_{2q+2})$, and let $||\cdot||$ be the Euclidean norm on $\mathbb{R}^{2q+2}$. There exists $\xi_0 > 0$ such that

(i) for $1 \leq i \leq 2q+2$, and all $(z_1, \cdots, z_m) \in E^m$, there exists a function $g^{(1)}$ from E^m into $\mathbb{R}$ such that

$$\sup_{||\vartheta - \vartheta_0|| \leq \xi_0} \left| \frac{\partial}{\partial \vartheta_i} \log p_\vartheta(z_1, \cdots, z_m) \right| \leq g^{(1)}(z_1, \cdots, z_m),$$

with

$$\int_{E^m} g^{(1)}(z_1, \cdots, z_m) p_{\vartheta_0}(z_1, \cdots, z_m) \lambda(dz_1^m) < \infty,$$

and for $\kappa > 0$,

$$\int_{E^m} \left(g^{(1)}(z_1, \cdots, z_m)\right)^{2+\kappa} p_{\vartheta_0}(z_1, \cdots, z_m) \lambda(dz_1^m) < \infty.$$

(ii) for all $1 \leq i, j \leq 2q+2$, and all $(z_1, \cdots, z_m) \in E^m$, there exists a function $g^{(2)}$ from E^m into $\mathbb{R}$ such that

$$\sup_{||\vartheta - \vartheta_0|| \leq \xi_0} \left| \frac{\partial^2}{\partial^2 \vartheta_i \vartheta_j} \log p_\vartheta(z_1, \cdots, z_m) \right| \leq g^{(2)}(z_1, \cdots, z_m),$$

and

$$\int_{E^m} g^{(2)}(z_1, \cdots, z_m) p_{\vartheta_0}(z_1, \cdots, z_m) \lambda(dz_1^m) < \infty.$$

C6. The partial derivative of order $0, 1, 2$ of the function $\vartheta \mapsto p_\vartheta(z_1, \cdots, z_m)$ are $\mathcal{E}^{\otimes m}$-measurable for each $\vartheta \in \Theta$.

C7. The Markov processes X and Y are supposed stationary and geometrically α-mixing (or β-mixing).

Definition of α-mixing coefficients for a stationary process is given in (13), and see also Doukhan (1995), p. 88, for a simple definition in the case of stationary Markov processes.

3 Consistency and asymptotic normality

In this section we prove under mild conditions that the MSDLE defined in (9) is consistent and asymptotically normal. For this purpose, we give immediately a technical Lemma usefull to treat the asymptotic behavior of the split data likelihood (and its derivatives).

Lemma 1 *(i) Under condition C7, for all measurable function $\varphi(\cdot)$ from E^m into $\mathbb{R}^d$, with $d \geq 1$, the sequence $(\varphi(\mathbf{Z}_{(k-1)m+1}^{mk}))_{k\geq 1}$ is stationary and geometrically α-mixing.*

(ii) Under the previous assumptions, for all $\varphi \in L_1(P_0)$ we have the strong law of large numbers, i.e.

$$M_k = \frac{1}{k}\sum_{j=1}^{k} \varphi(\mathbf{Z}_{(j-1)m+1}^{jm}) \underset{k\to\infty}{\longrightarrow} E_0(\varphi(\mathbf{Z}_1^m)), \quad P_0 - a.s. \tag{10}$$

(iii) Suppose that $E(\varphi(\mathbf{Z}_1^m)) = 0$, $E|\varphi(\mathbf{Z}_1^m)|^{2+\kappa} < \infty$ for some real number $\kappa > 0$, and that condition C7 is satisfied, then

$$\Sigma \overset{def.}{=} E(\varphi(\mathbf{Z}_1^m)^2) + 2\sum_{k=1}^{\infty} kE(\varphi(\mathbf{Z}_1^m)\varphi^T(\mathbf{Z}_{(k-1)m+1}^{mk})) < \infty, \tag{11}$$

and, if $\Sigma \neq 0_{d\times d}$,

$$\sqrt{k}M_k \underset{k\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,\Sigma). \tag{12}$$

Proof. (i) Without loss of generality we consider the case $m = 2$. The function $\varphi(\cdot)$ being $\mathcal{E}^{\otimes 2}$-measurable, it is enough to consider the α-mixing coefficient associated to the Markov process $W = (W_k)_{k\geq 1} = (\mathbf{X}_{(k-1)2+1}^{2k}, \mathbf{Y}_{(k-1)2+1}^{2k}, \mathbf{U}_{(k-1)2+1}^{2k})_{k\geq 1}$. At this step, let us define for all stationary process $\tilde{X}$, and all $(t,k) \in \mathbb{N}^* \times \mathbb{N}$ the sequence of α-mixing coefficients associated to $\tilde{X}$ by:

$$\alpha^{\tilde{X}}(k) = \sup_{A\in\mathcal{F}_{\tilde{X},1}^{t}, B\in\mathcal{F}_{\tilde{X},t+k+1}^{\infty}} |P(A\cap B) - P(A)P(B)|. \tag{13}$$

where $\mathcal{F}_{\tilde{X},t_1}^{t_2}$ denotes, for all $t_2 > t_1 \geq 1$, the σ-algebra generated by $(\tilde{X}_{t_1}, \cdots, \tilde{X}_{t_2})$. Following (13), the sequence of α-mixing coefficients associated to W is defined by

$$\alpha^W(k) = \sup|P((A_1,A_2,A_U)\cap(B_1,B_2,B_U)) - P(A_1,A_2,A_U)P(B_1,B_2,B_U)| \tag{14}$$

where the supremum is taken over all $(A_1,A_2,A_U) \in \mathcal{F}_{X,1}^{2n} \otimes \mathcal{F}_{Y,1}^{2n} \otimes \mathcal{F}_{U,1}^{2n}$ and $(B_1,B_2,B_U) \in \mathcal{F}_{X,2n+2k+1}^{\infty} \otimes \mathcal{F}_{Y,2n+2k+1}^{\infty} \otimes \mathcal{F}_{U,2n+2k+1}^{\infty}$.

From mutual independence of X, Y and U, the modulus of the difference of probabilities in the right hand side of (14) satisfies

$$\begin{aligned}
&|P(A_1\cap B_1)P(A_2\cap B_2)P(A_U\cap B_U) - P(A_1)P(B_1)P(A_2)P(B_2)P(A_U)P(B_U)| \\
&= P(A_U)P(A_1)P(A_2)|P(B_1|A_1)P(B_2|A_2)[P(B_U|A_U) - P(B_U)] \\
&\quad + P(B_1|A_1)P(B_U)[P(B_2|A_2) - P(B_2)] \\
&\quad + P(B_2)P(B_U)[P(B_1|A_1) - P(B_1)]| \\
&\leq |P(A_U\cap B_U) - P(A_U)P(B_U)| + |P(A_1\cap B_1) - P(A_1)P(B_1)| \\
&\quad + |P(A_2\cap B_2) - P(A_2)P(B_2)|.
\end{aligned}$$

From the markovian structure of U, X and Y, condition C7, and the last inequality, it comes

$$\alpha^W(k) \leq \alpha^U(2k) + \alpha^X(2k) + \alpha^Y(2k) \leq \rho^k, \tag{15}$$

for a certain $0 < \rho < 1$, and k large enough. Notice now that the mapping s from $E^2 \times E^2 \times \{1,2\}^2$ into E^2, such that $\mathbf{Z}_1^2 = s(\mathbf{X}_1^2; \mathbf{Y}_1^2; \mathbf{U}_1^2)$, see (4), and defined by

$$s(x_1, x_2; y_1, y_2; u_1, u_2) = ((2-u_1)x_1 + (u_1 - 1)y_1; (2-u_2)x_2 + (u_2-1)y_2),$$

is measurable, hence $\varphi \circ s$ is a measurable function from $E^2 \times E^2 \times \{1,2\}^2$ into $\mathbb{R}^d$, which induces that the α–mixing coefficients of the sequence $(\varphi(\mathbf{Z}_{(k-1)m+1}^{mk}))_{k\geq 1}$ are inferior or equal to the coefficients induced by W, which using (15) concludes the proof of (i).

(ii) This result is a direct consequence of the maximal ergodic theorem for stationary processes, see Stout (1974), p.145.

(iii) This central limit theorem is a classical result, see (29.10) p.387, in Billingsley, which is proved by considering central limit theorem for real α–mixing sequences of random variables, see theorem 3.2.1 in Zhengyan, and Chuanrong (1996), and Crámer-Wold device, see theorem 29.4 p.383, in Billingsley.

□

Theorem 1 *Under assumptions C1–7, the MSDLE defined in (9) is strongly consistent, i.e.*

$$\hat{\vartheta}_k \underset{k\to\infty}{\longrightarrow} \vartheta_0 \quad P_0 - a.s. \tag{16}$$

where ϑ_0 is the true value of the parameter.

PROOF. The proof is based on the proof given by Dacunha-Castelle and Duflo (1993, p. 94–96). First of all the MSDLE can be defined as a minimum contrast estimator, as follows

$$\hat{\vartheta}_k = \arg\min_{\vartheta\in\Theta} U_k(\vartheta),$$

where

$$U_k(\vartheta) = -k^{-1}\ell_\vartheta(\mathbf{Z}_1^{mk}), \quad \text{and} \quad \ell_\vartheta(\mathbf{Z}_1^{mk}) = \log \prod_{j=1}^{k} p_\vartheta(\mathbf{Z}_{(j-1)m+1}^{mj}) = \sum_{j=1}^{k} \log p_\vartheta(\mathbf{Z}_{(j-1)m+1}^{mj}), \tag{17}$$

where $\ell_\vartheta(z_1^{km})$ denotes the log of the split data likelihood (SDL). From Lemma 1 and conditions C3 and C7, we get

$$U_k(\vartheta) = -k^{-1}\ell_\vartheta(\mathbf{Z}_1^{mk}) \underset{k\to\infty}{\longrightarrow} \mathcal{E}_0(\vartheta) = -E_0(\log p_\vartheta(\mathbf{Z}_1^m)), \quad P_0 - a.s. \tag{18}$$

It is clear from Jensen inequality and condition C2, that

$$\mathcal{E}_0(\vartheta_0) \leq \mathcal{E}_0(\vartheta), \quad \text{and } \mathcal{E}_0(\vartheta_0) = \mathcal{E}_0(\vartheta) \Rightarrow \vartheta = \vartheta_0.$$

We consider now the Kullback distance $K(\vartheta_0,\vartheta) = \mathcal{E}_0(\vartheta) - \mathcal{E}_0(\vartheta_0) \geq 0$, with $K(\vartheta_0,\vartheta) = 0 \Leftrightarrow \vartheta_0 = \vartheta$. Let us consider D a countable dense set in Θ, by this way $\inf_{\vartheta\in\Theta} U_k(\vartheta) = \inf_{\vartheta\in\Theta\cap D} U_k(\vartheta)$, is a $\mathcal{F}^k_{Z,1}$-measurable random variable. We define in addition the random variable $W(k,\eta) = \sup\{|U_k(\vartheta) - U_k(\vartheta')|;\ (\vartheta,\vartheta')\in D^2,\ |\vartheta-\vartheta'|\leq\eta\}$, and recall that $\mathcal{K}(\vartheta_0,\vartheta_0) = 0$. Let us consider a non empty open ball B_0 centered in ϑ_0 such that $K(\vartheta_0,\vartheta)$ is bounded from below by a positive real number 2ε on $\Theta\backslash B_0$. Let us consider a sequence $(\eta_r)_{r\geq 0}$ decreasing toward zero, and cover $\Theta\backslash B_0$ by a finite number ℓ of balls $(B_i)_{1\leq i\leq\ell}$, centered respectively in $(\vartheta_i)_{1\leq i\leq\ell}$, and of radius less than η_r for one r fixed arbitrarly. For all $\vartheta\in B_i$, then

$$\begin{aligned} U_k(\vartheta) &\geq U_k(\vartheta_i) - |U_k(\vartheta) - U_k(\vartheta_i)| \\ &\geq U_k(\vartheta_i) - \sup_{\vartheta\in B_i}|U_k(\vartheta) - U_k(\vartheta_i)|, \end{aligned}$$

which leads to

$$\inf_{\vartheta\in\Theta\backslash B_0} U_k(\vartheta) \geq \inf_{1\leq i\leq\ell} U_k(\vartheta_i) - W(k,\eta_r).$$

As a consequence we have the following events inclusions

$$\begin{aligned} \left\{\hat\vartheta_k \notin B_0\right\} &\subseteq \left\{\inf_{\vartheta\in\Theta\backslash B_0} U_k(\vartheta) < \inf_{\vartheta\in B_0} U_k(\vartheta)\right\} \\ &\subseteq \left\{\inf_{\vartheta\in\Theta\backslash B_0} U_k(\vartheta) < U_k(\vartheta_0)\right\} \\ &\subseteq \left\{\inf_{1\leq i\leq\ell} U_k(\vartheta_i) - W(k,\eta_r) < U_k(\vartheta_0)\right\} \\ &\subseteq \{W(k,\eta_r) > \varepsilon\} \cup \left\{\inf_{1\leq i\leq\ell}(U_k(\vartheta_i) - U_k(\vartheta_0)) \leq \varepsilon\right\}. \end{aligned}$$

Thus we have

$$\limsup_k \left\{\hat\vartheta_k \notin B_0\right\} \subseteq \limsup_k \{W(k,\eta_r) > \varepsilon\} \cup \limsup_k \left\{\inf_{1\leq i\leq\ell}(U_k(\vartheta_i) - U_k(\vartheta_0)) \leq \varepsilon\right\}. \quad (19)$$

By the strong law of large number established in (18) we have

$$P_0\left(\limsup_k \left\{\inf_{1\leq i\leq\ell}(U_k(\vartheta_i) - U_k(\vartheta_0)) \leq \varepsilon\right\}\right) = 0. \quad (20)$$

In addition according to assumption C3 there exists a random variable $h(\mathbf{Z}_1^m)$ such that

$$\sup_{\vartheta\in\Theta} |\log p_\vartheta(\mathbf{Z}_1^m)| \leq h(\mathbf{Z}_1^m),$$

with $E_0[h(\mathbf{Z}_1^m)] < \infty$, where $h = |\log g_1| + |\log g_2|$, does not depend on ϑ. Let us consider the following random variable

$$H_\eta(\mathbf{Z}_1^m) = \sup_{(\vartheta,\vartheta')\in\Theta^2}\left\{|\log p_\vartheta(\mathbf{Z}_1^m) - \log p_{\vartheta'}(\mathbf{Z}_1^m)|;\ |\vartheta-\vartheta'|\leq\eta\right\}.$$

It comes, using the previous uniform upper bound and continuity assumption C4, that

$$H_\eta\left(\mathbf{Z}_1^m\right) \leq 2h\left(\mathbf{Z}_1^m\right), \text{ and } \lim_{\eta\to 0} E_0\left[H_\eta\left(\mathbf{Z}_1^m\right)\right] = 0.$$

Hence for r' large enough we have $E_0(H_{\eta_{r'}}(\mathbf{Z}_1^m)) \leq \varepsilon$, and $W(k,\eta_{r'}) \leq k^{-1}\sum_{j=1}^k H_{\eta_{r'}}(\mathbf{Z}_{(j-1)m+1}^{mj})$ P_0–almost surely , therefore

$$\limsup_k \{W(k,\eta_{r'}) > \varepsilon\} \subseteq \limsup_k \left\{k^{-1}\sum_{j=1}^k H_{\eta_{r'}}(\mathbf{Z}_{(j-1)m+1}^{mj}) > \varepsilon\right\},$$

and $P_0\left(\limsup_k \left\{k^{-1}\sum_{j=1}^k H_{\eta_{r'}}(\mathbf{Z}_{(j-1)m+1}^{mj}) > \varepsilon\right\}\right) = 0$ which leads to

$$P_0(\limsup_k \{W(k,\eta_{r'}) > \varepsilon\}) = 0. \tag{21}$$

By (19)–(21), we prove the strong consistency of the MSDLE $\hat{\vartheta}_k$. □

Write $V_j(\vartheta) = \log p_\vartheta(\mathbf{Z}_{(j-1)m+1}^{mj})$ for $j = 1,\cdots,k$, and let us denote respectively, for any function v depending on ϑ, its gradient vector, and Hessian matrix, by :

$$\dot{v}\,(\vartheta) = \frac{\partial v}{\partial\vartheta}(\vartheta) \quad \text{and} \quad \ddot{v}\,(\vartheta) = \frac{\partial^2 v}{\partial\vartheta\partial\vartheta^T}(\vartheta). \tag{22}$$

Let us denote by $\ell_k(\vartheta) = \ell_\vartheta(\mathbf{Z}_1^{2k})$, to follow notation in (22). From writing (17), it comes

$$k^{-1/2}\,\dot{\ell}_k\,(\vartheta) = k^{-1/2}\sum_{j=1}^k \dot{V}_j\,(\vartheta). \tag{23}$$

Lemma 2 *Under assumption C4–7, we have* $k^{-1/2}\,\dot{\ell}_k\,(\vartheta_0) \xrightarrow[k\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_0)$, *where*

$$\Sigma_0 = E_0(\dot{V}_1\,(\vartheta_0)\,\dot{V}_1^T\,(\vartheta_0)) + 2\sum_{k=2}^\infty kE_0(\dot{V}_1\,(\vartheta_0)\,\dot{V}_k^T\,(\vartheta_0)) < \infty.$$

Proof. This result is a direct consequence of condition C4–7, and Lemma 1, taking $\varphi(\cdot) = \frac{\partial}{\partial\vartheta}\log p_{\vartheta_0}(\cdot)$. □

Lemma 3 *Let* $(\vartheta_k^*)_{k\geq 0}$ *be any arbitrary sequence converging* P_0*–a.s. toward* ϑ_0*. Under assumptions C4–C7, whe have* $k^{-1}\,\ddot{\ell}_k\,(\vartheta_k^*) \underset{k\to\infty}{\longrightarrow} A_0 = E_0(\ddot{V}_1\,(\vartheta_0))$ *in* P_0*–probability.*

Proof. By Lemma 1 and assumption C4 we know that $(k^{-1}\,\ddot{\ell}_k\,(\vartheta_0))_{k\geq 1}$ converges P_0–almost surely to $E_0\left[\ddot{V}_1\,(\vartheta_0)\right]$. Now let us prove that $(k^{-1}\,\ddot{\ell}_k\,(\vartheta_k^*))_{k\geq 1}$ and $(k^{-1}\,\ddot{\ell}_k\,(\vartheta_0))_{k\geq 1}$ are asymptotically equivalent in P_0–probability, that is:

$$\forall\ \eta > 0, \quad \lim_{k\to\infty} P_0\left(\left|\frac{1}{k}\,\ddot{\ell}_k\,(\vartheta_k^*) - \frac{1}{k}\,\ddot{\ell}_k\,(\vartheta_0)\right| > \eta\right) = 0,$$

where we denote simply (among this proof) by $|\cdot|$ the norm on the real matrices defined, for all $d\times d$ real matrix $A=(A_{i,j})_{i,j=1,\cdots,d}$, by $|A|=\max_{i,j=1,\cdots,d}|A_{i,j}|$, taking the convention that the norm of a scalar coincides with its modulus. For this purpose we notice that for all $0<\xi<\xi_0$ (for definition of ξ_0, see assumption C5), we can write

$$P_0\left(\left|\frac{1}{k}\ddot{\ell}_k(\vartheta_k^*)-\frac{1}{k}\ddot{\ell}_k(\vartheta_0)\right|>\eta\right) \le P_0\left(\frac{1}{k}\sum_{j=1}^{k}\sup_{\vartheta\in B(\vartheta_0,\xi)}\left|\ddot{V}_j(\vartheta)-\ddot{V}_j(\vartheta_0)\right|>\eta\right) +P_0(\vartheta_k^*\notin B(\vartheta_0,\xi)),$$

where $B(\vartheta_0,\xi)$ denotes the ball centered in ϑ_0, with radius equal to ξ. The second term of the right hand side goes to zero as k goes to infinity by strong consistency of ϑ_k^*. For the first term of right hand side we notice

$$\varrho(\xi;\mathbf{Z}_1^m)=\sup_{\vartheta\in B(\vartheta_0,\xi)}\left|\ddot{V}_j(\vartheta)-\ddot{V}_j(\vartheta_0)\right|\underset{\xi\to 0}{\longrightarrow}0 \quad a.e.$$

In addition there exists, from C5, a P_0–integrable function $g^{(2)}$, such that, for all $j=1,\cdots,k$, the components of the matrix $\ddot{V}_j(\vartheta)$ are all dominated in modulus by $g^{(2)}(\mathbf{Z}_1^m)$ on $B(\vartheta_0,\xi_0)$, which implies that $\varrho(\xi;\mathbf{Z}_1^m)\le 2g^{(2)}(\mathbf{Z}_1^m)$. Now, using the Lebesgue's continuity theorem, it comes that

$$E_0\left[\varrho(\xi;\mathbf{Z}_1^m)\right]\underset{\xi\to 0}{\longrightarrow}0. \tag{24}$$

For all $\varepsilon>0$, and all $\xi>0$ small enough such that $0<E_0\left[\varrho(\xi;\mathbf{Z}_1^m)\right]<\varepsilon$, we have, using Tchebychev inequality for positive random variables:

$$\begin{aligned}P_0\left(\frac{1}{k}\sum_{j=1}^{k}\varrho(\xi;\mathbf{Z}_{(j-1)m+1}^{mj})\ge\varepsilon\right) &\le P_0\left(\frac{1}{k}\sum_{j=1}^{k}\varrho(\xi;\mathbf{Z}_{(j-1)m+1}^{mj})\ge\varepsilon-E_0\left[\varrho(\xi;\mathbf{Z}_1^m)\right]\right)\\ &\le \frac{1}{k[\varepsilon-E_0[\varrho(\xi;\mathbf{Z}_1^m)]]}\sum_{j=1}^{k}E_0\left[\varrho(\xi;\mathbf{Z}_{(j-1)m+1}^{mj})\right]\\ &= \frac{1}{\varepsilon-E_0\left[\varrho(\xi;\mathbf{Z}_1^m)\right]}E_0\left[\varrho(\xi;\mathbf{Z}_1^m)\right],\end{aligned}$$

which goes to zero, from (24), as ξ goes to 0. □

Theorem 2 *Under assumptions C1–C7, and assuming that A_0 is nonsingular, we get*

$$k^{1/2}(\hat{\vartheta}_k-\vartheta_0)\underset{k\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}(0,A_0^{-1}\Sigma_0A_0^{-1}).$$

PROOF. For k large enough $\hat{\vartheta}_k$ is an interior point of Θ, and $\|\hat{\vartheta}_k-\vartheta_0\|<\xi_0$, and then by a Taylor expansion of $\dot{\ell}_\vartheta(\mathbf{Z}_1^{km})$ about ϑ_0 we get,

$$k^{1/2}(\hat{\vartheta}_k-\vartheta_0)=[-k^{-1}\ddot{\ell}_k(\vartheta_k^*)]^{-1}k^{-1/2}\dot{\ell}_k(\vartheta_0),$$

where ϑ_k^* is a point of the line segment between ϑ_0 and $\hat{\vartheta}_k$. Therefore using Theorem 1, Lemma 3 and 2 we obtain the asymptotic normality of the MSDLE. □

4 Monte-Carlo estimate of the log SDL

We noticed in section 2, that except in the case of linear gaussian autoregressive models, the invariant probability densities q_θ and q_ϕ involved in (8) are analytically unknown, but are solutions of the fixed point problems in (6). The goal of this section is to propose a general tractable approach to approximate the log of the SDL, see (17), for a given fixed sample path z_1^{mk}, $k \geq 1$. The methodology presented here is inspired by Chauveau and Vandekerkhove (2001). Let consider for simplicity the case $k = 1$. In this framework the crucial point is to estimate numerically, for each $\theta \in \Phi^1$, and each $\phi \in \Phi^2$, the quantities $q_\theta^1(z_1)$ and $q_\phi^2(z_1)$. We just illustrate our method on $q_\theta^1(z_1)$, the same work holding for $q_\phi^2(z_1)$ (the associate estimate will be denoted by $\hat{q}_\phi^2(z_1)$. Let suppose that for each $\theta \in \Phi^1$, we are able to simulate an ergodic Markov process $X^\theta = (X_N^\theta)_{N\geq 1}$, from Q_θ^1 (knowledge of the stationary initial distribution is not needed in practice, a long burning-run of the chain suffices). From strong law of large numbers for ergodic Markov chains, and (6), it comes

$$\hat{q}_\theta^1(z_1) = \frac{1}{N}\sum_{i=1}^{N} Q_\theta^1(X_i^\theta, z_1) \underset{N\to\infty}{\longrightarrow} \int_E q_\theta^1(x) Q_\theta^1(x, z_1)\lambda(dx) = q_\theta^1(z_1), \quad P_\theta - a.s.$$

Hence $\hat{q}_\theta^1(z_1)$ is a strongly convergent estimate of $q_\theta^1(z_1)$. From this, it is easy to construct a consistent plug-in estimator $\hat{\ell}_\vartheta(z_1^m)$ of $\ell_\vartheta(z_1^m)$, replacing $q_\theta^1(z_1)$ and $q_\phi^2(z_1)$ respectively by $\hat{q}_\theta^1(z_1)$ and $\hat{q}_\phi^2(z_1)$, in (8). In addition, central limit theorem for $\hat{\ell}_\vartheta(z_1^m)$ can be established. In fact write for simplicity $\hat{\ell}_\vartheta(z_1^m)$ and $\ell_\vartheta(z_1^m)$ under the form

$$\hat{\ell}_\vartheta(z_1^m) = \log(\hat{q}_\theta^1(z_1)c_1 + \hat{q}_\phi^2(z_1)c_2), \quad \text{and } \ell_\vartheta(z_1^m) = \log(q_\theta^1(z_1)c_1 + q_\phi^2(z_1)c_2),$$

where c_1 and c_2 are constants depending on z_2^m and ϑ. Let suppose a condition C7' equivalent to C7, but true for all $\vartheta \in \Theta$ (and not only for ϑ_0). In that case, supposing moments conditions of Lemma 1 on $Q_\theta^1(z_1, \cdot)$ and $Q_\phi^2(z_1, \cdot)$, it comes

$$\sqrt{N}(\hat{q}_\theta^1(z_1) - q_\theta^1(z_1)) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \Sigma^1), \quad \text{and} \quad \sqrt{N}(\hat{q}_\phi^2(z_1) - q_\phi^2(z_1)) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \Sigma^2). \tag{25}$$

where Σ^1 and Σ^2 are variance terms defined in (11). Finally by a Taylor expansion of the log function, about $q_\theta^1(z_1)c_1 - q_\phi^2(z_1)c_2$ we have for all $N \geq 1$:

$$\sqrt{N}(\hat{\ell}_\vartheta(z_1^m) - \ell_\vartheta(z_1^m)) = \frac{1}{\ell_N^*}(\sqrt{N}(\hat{q}_\theta^1(z_1) - q_\theta^1(z_1))c_1 + \sqrt{N}(\hat{q}_\phi^2(z_1) - q_\phi^2(z_1))c_2),$$

where ℓ_N^* is a point of the line between $\hat{q}_\theta^1(z_1)c_1 - \hat{q}_\phi^2(z_1)c_2$ and $q_\theta^1(z_1)c_1 - q_\phi^2(z_1)c_2$. From (25) and convergence in P_ϑ-probability of ℓ_n^* towards $q_\theta^1(z_1)c_1 - q_\phi^2(z_1)c_2$, it comes

$$\sqrt{N}(\hat{\ell}_\vartheta(z_1^m) - \ell_\vartheta(z_1^m)) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = [q_\theta^1(z_1)c_1 - q_\phi^2(z_1)c_2]^{-2}[c_1^2\Sigma^1 + c_2^2\Sigma^2]$. In conclusion we have proposed a $\sqrt{N}$-consistent method to calculate the terms of the form $\log p_\vartheta(z_{(j-1)m+1}^{mj})$, $1 \leq j \leq k$, and hence for k fixed, and each $\vartheta \in \Theta$ a $\sqrt{N}$-consistent method exists to calculate the log of the SDL, see (17), at point z_1^{mk} (the asymptotic variance growing linearly with k). From the pratical point of view, this approach allows at least to compute pointwise the log SDL, and to implement a discretized version of the MSDLE on a grid over Θ.

5 Hidden Markov Mixture of two AR(1)s

In this section we check in detail that conditions of mixing, identifiability, regularity, and integrability supposed in section 2, see C2–7, are satified for a Hidden Markov Mixture of two autoregressive of order 1 (in abbreviate AR(1)) processes. More precisely the processes X and Y considered in this section are defined, for all $n \geq 1$, by the dynamic

$$X_{n+1} = a_1 X_n + \varepsilon_{n+1}, \text{ and } \quad Y_{n+1} = a_2 Y_n + \varepsilon'_{n+1}, \tag{26}$$

where $(a_1, a_2) \in (0,1)$, $(\varepsilon_n)_{n\geq 1}$ and $(\varepsilon'_n)_{n\geq 1}$ are two mutually independent sequences of independent gaussian random variables with respectively μ_1 and μ_2 mean, and variance equal to σ_1^2, and σ_2^2. The mixture process U is a $\{1,2\}$-Markov chain with transition matrix defined in (7). The parameter in such a set up is:

$$\vartheta = (\alpha, \beta, \mu_1, \mu_2, a_1, a_2, \sigma_1^2, \sigma_2^2).$$

We do not describe precisely at this step the form of the parametrical space Θ since it will be deduced from the coming discussion about identifiability.

Mixing. Condition $C7$ is clearly satisfied for U, and the same hold for X and Y since these processes are geometrically β-mixing (and hence α-mixing), see for example Baraud *et al.* (2001) for general conditions.

Identifiability. Processes X, and Y defined (26), with initial conditions x_1, and y_1 satisfy

$$X_{n+1} = a_1^n x_1 + \sum_{k=0}^{n-1} a_1^k \varepsilon_{n+1-k}, \quad Y_{n+1} = a_2^n y_1 + \sum_{k=0}^{n-1} a_2^k \varepsilon'_{n+1-k}. \tag{27}$$

From this writing and properties on gaussian random vectors, it is easy to identify the density of the stationary distribution of X and Y processes. In fact for X we obtain the density of a $\mathcal{N}(m_1, s_1)$ distribution, when for Y we have the density of a $\mathcal{N}(m_2, s_2)$ distribution, with

$$m_1 = \frac{\mu_1}{1-a_1}, \; s_1 = \frac{\sigma_1^2}{1-a_1^2}, \; m_2 = \frac{\mu_2}{1-a_2}, \text{ and } s_2 = \frac{\sigma_2^2}{1-a_2^2}. \tag{28}$$

In order to prove identifiability, we propose to consider $m = 2$, and to check that for all ϑ and ϑ' in Θ (which needs to be defined), we have:

$$p_\vartheta(z_1, z_2) = p_{\vartheta'}(z_1, z_2) \quad \lambda^{\otimes 2} - a.e. \Rightarrow \vartheta = \vartheta'. \tag{29}$$

Nethertheless partial information on identifiability will be given considering the marginal equality

$$p_\vartheta(z_2) = \int_{\mathbb{R}} p_\vartheta(z_1, z_2) dz_1 = \int_{\mathbb{R}} p_{\vartheta'}(z_1, z_2) dz_1 = p_{\vartheta'}(z_2) \quad \lambda - a.e. \tag{30}$$

Denoting by f_{μ,σ^2} the density function of a $\mathcal{N}(\mu, \sigma^2)$, we have for all $\vartheta \in \Theta$:

$$p_\vartheta(z_2) = \pi_\vartheta(1) f_{(m_1,s_1)}(z_2) + \pi_\vartheta(2) f_{(m_2,s_2)}(z_2). \tag{31}$$

Teicher (1963) establishes identifiability property for mixtures of various density families. A mixture of at most r elements of $\mathcal{G} = \{g(z;\theta);\ \theta \in \Phi\}$ is identifiable if for θ_i and θ_i', for $i = 1, \cdots, r$, in Φ, $(c_1, \cdots, c_r)$ and $(c_1', \cdots, c_r')$ probability vectors we have (δ_θ denotes the point mass at θ)

$$\sum_{i=1}^{r} c_i g(z;\theta_i) = \sum_{i=1}^{r} c_i' g(z;\theta_i') \quad \lambda - a.e. \Rightarrow \sum_{i=1}^{r} c_i \delta_{\theta_i} = \sum_{i=1}^{r} c_i' \delta_{\theta_i'}.$$

This writing is equivalent to the following statement: there exists a unique permutation σ on $\{1, \cdots, r\}$ such that for all $i = 1, \cdots, r$, $(c_i, \theta_i) = (c'_{\sigma(i)}, \theta'_{\sigma(i)})$. Teicher (1963) shows in particular that mixtures of gaussian densities (where $\theta_i = (m_i, s_i)$, for $i = 1, \cdots, r$, with m_i denoting the mean parameter, and s_i denoting the variance parameter of the i-th component of the mixture) are identifiable, and propose to order the parametrical space to avoid the previous permutation ambiguities, *i.e.* by imposing $m_1 < \cdots < m_r$ if $s_i = s_j$, for all $i, j = 1, \cdots, r$, or $(m_i, s_i) < (m_j, s_j)$ if $s_i < s_j$ or $m_i < m_j$ if $s_i = s_j$. In practice equality of the variance parameters is supposed to build a simple parametrical for $(m_1, \cdots, m_r)$. The same approach can be done on $(s_1, \cdots, s_r)$ if the variance parameters are supposed to be all differents (without any order constraints on $(m_1, \cdots, m_r)$). Imposing in Θ that $s_1 < s_2$, the mixture equality (30) and (31), leads to a first partial identification

$$\pi_\vartheta(1) = \pi_{\vartheta'}(1), \quad \pi_\vartheta(2) = \pi_{\vartheta'}(2), \quad m_1 = m_1', \quad m_2 = m_2', \quad s_1 = s_1', \quad \text{and} \quad s_2 = s_2'. \tag{32}$$

For simplicity let us denote $\pi(\cdot) = \pi_\vartheta(\cdot)$, $\pi'(\cdot) = \pi_{\vartheta'}(\cdot)$, $\pi(\cdot,\cdot) = \pi_\vartheta(\cdot,\cdot)$, and $\pi'(\cdot,\cdot) = \pi_{\vartheta'}(\cdot,\cdot)$. Using this first identification in (29), it comes the following relation which is to be discussed:

$$\begin{aligned} & \pi(1)\pi(1,1) f_{(m_1,s_1)}(z_1) f_{(a_1 z_1 + \mu_1, \sigma_1^2)}(z_2) + \pi(1)\pi(1,2) f_{(m_1,s_1)}(z_1) f_{(m_2,s_2)}(z_2) \\ & + \pi(2)\pi(2,1) f_{(m_2,s_2)}(z_1) f_{(m_1,s_1)}(z_2) + \pi(2)\pi(2,2) f_{(m_2,s_2)}(z_1) f_{(a_2 z_1 + \mu_2, \sigma_2^2)}(z_2) \\ = \\ & \pi(1)\pi'(1,1) f_{(m_1,s_1)}(z_1) f_{(a_1' z_1 + \mu_1', \sigma_1'^2)}(z_2) + \pi(1)\pi'(1,2) f_{(m_1,s_1)}(z_1) f_{(m_2,s_2)}(z_2) \\ & + \pi(2)\pi'(2,1) f_{(m_2,s_2)}(z_1) f_{(m_1,s_1)}(z_2) + \pi(2)\pi'(2,2) f_{(m_2,s_2)}(z_1) f_{(a_2' z_1 + \mu_2', \sigma_2'^2)}(z_2). \end{aligned} \tag{33}$$

Taking memberwise the Fourrier transform with respect to z_2, it comes

$$\begin{aligned} & \pi(1)\pi(1,1)f_{(m_1,s_1)}(z_1)e^{it(a_1z_1+\mu_1)-\sigma_1^2t^2} + \pi(1)\pi(1,2)f_{(m_1,s_1)}(z_1)e^{itm_2-s_2t^2} \\ & +\pi(2)\pi(2,1)f_{(m_2,s_2)}(z_1)e^{itm_1-s_1t^2} + \pi(2)\pi(2,2)f_{(m_2,s_2)}(z_1)e^{it(a_2z_1+\mu_2)-\sigma_2^2t^2} \\ = & \\ & \pi(1)\pi'(1,1)f_{(m_1,s_1)}(z_1)e^{it(a'_1z_1+\mu'_1)-\sigma'^2_1t^2} + \pi(1)\pi'(1,2)f_{(m_1,s_1)}(z_1)e^{itm_2-s_2t^2} \\ & +\pi(2)\pi'(2,1)f_{(m_2,s_2)}(z_1)e^{itm_1-s_1t^2} + \pi(2)\pi'(2,2)f_{(m_2,s_2)}(z_1)e^{it(a'_2z_1+\mu'_2)-\sigma'^2_2t^2}. \end{aligned} \tag{34}$$

Let consider the case $\sigma_1^2 \neq \sigma_2^2$. Let consider now the subcase $\sigma_2^2 < \sigma'^2_2 < s_2$, $\sigma_1^2 < \sigma'^2_1 < s_1$, and $\sigma_2^2 < \sigma_1^2$. Multiplying the two hand sides of (34) by $e^{\sigma_2 t^2}$, and taking the limit when t goes to infinity, we obtain the absurd result $\pi(2,2) = 0$. Let consider the more complicated subcase $s_1 < \sigma_2^2 < \sigma'^2_2 < s_2$, and $\sigma_1 = \sigma'_1 < s_1$ (which induces that $a_1 = a'_1$, and $\mu_1 = \mu'_1$, from (32)). Considering the previous constraints in (34) and multiplying both sides of (34) by $e^{\sigma_1^2t^2}$ and taking limit when t goes to infinity, we obtain the necessary condition $\pi(1,1) = \pi'(1,1)$ (hence terms in $e^{-\sigma_1^2t^2}$ disappear from (34)). After that let multiply the remaining hand sides of (34) by $e^{s_1t^2}$, and take the limit as t goes to infinity, it comes that necessary $\pi(2,1) = \pi'(2,1)$ (hence terms in $e^{-s_1^2t^2}$ disappear from (34)). Finally, let multiply the remaining hand sides of (34) (with only two terms at this step) by $e^{\sigma_2^2t^2}$, and take the limit as t goes to infinity, it comes that necessary $\pi(2,2) = 0$ which is absurd. In any situation such that $(\sigma_1^2, \sigma_2^2) \neq (\sigma'^2_1, \sigma'^2_2)$ the same technique should be applied leading in any case to an absurd conclusion. By this way it is established that necessary $(\sigma_1^2, \sigma_2^2) = (\sigma'^2_1, \sigma'^2_2)$. From this remark and (32), it comes that $a_1 = a'_1$, $a_2 = a'_2$, $\mu_1 = \mu'_1$, and $\mu_2 = \mu'_2$. Including the various identifications obtained until now in (34), we have:

$$\begin{aligned} 0 = & \ \pi(1)(\pi(1,1) - \pi'(1,1))f_{(m_1,s_1)}(z_1)e^{it(a_1z_1+\mu_1)-\sigma_1^2t^2} \\ & +\pi(1)(\pi(1,2) - \pi'(1,2))f_{(m_1,s_1)}(z_1)e^{itm_2-s_2t^2} \\ & +\pi(2)(\pi(2,1) - \pi'(2,1))f_{(m_2,s_2)}(z_1)e^{itm_1-s_1t^2} \\ & +\pi(2)(\pi(2,2) - \pi'(2,2))f_{(m_2,s_2)}(z_1)e^{it(a_2z_1+\mu_2)-\sigma_2^2t^2}. \end{aligned} \tag{35}$$

The right hand side of the previous equality being a linear combination of linearly independent functions it comes that $\pi(i,j) = \pi'(i,j)$, for all i and j in $\{1,2\}$, which concludes the proof for this first case. The other cases $\sigma_1^2 = \sigma_2^2$ (and $a_1 < a_2$) or $a_1 = a_2$ (and $\sigma_1^2 < \sigma_2^2$) are solved in the same way by using (32)–(35).

Remarks. i) In conclusion, if assumption $s_1 < s_2$ is made (which is reasonnable in practice), the parameter $\vartheta = (\alpha, \beta, \mu_1, \mu_2, a_1, a_2, \sigma_1^2, \sigma_2^2)$ should be supposed to belong to a compact set $[\delta, 1-\delta]^2 \times [-M, M]^2 \times \mathcal{S}$, where $\mathcal{S}$ is any compact subset of $[0, 1-\delta]^2 \times [\delta, V]^2$, where $0 < \delta < 1$ denotes an arbitrary small positive value, and $0 < M < \infty$, $0 < V < \infty$ are arbitrary positive bounds, such that:

$$\forall (a_1, a_2, \sigma_1^2, \sigma_2^2) \in \mathcal{S} : \quad \frac{\sigma_1^2}{1 - a_1^2} < \frac{\sigma_2^2}{1 - a_2^2}. \tag{36}$$

ii) The previous result can be extended with an extra work to cases corresponding to $K \geq 2$ (the previous technique do not use the fact that $\pi(1,1) = 1 - \pi(1,2)$ or $\pi(2,1) = 1 - \pi(2,2)$ when $K = 2$), but the set $\mathcal{S}$ becomes then much more tricky to build.

iii) Finally it is good to notice that the case $a_i = 0$ for some i's in $\{1, \cdots, K\}$, is compatible with this identifiability approach. Thus Markov mixtures of AR(1) processes and sequences of i.i.d. gaussian random variables, according to expression (4), leads to an identifiable model.

iv) If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and $a_1 = a_2 = a$, which leads to $s_1 = s_2$, the model is still identifiable. In fact the same kind of proof can be employed using the ordering $\mu_1 < \mu_2$.

Regularity, integrability. In this part we check essentially that conditions C3–6 are satisfied. In order to simplify the expressions, and without loss of generality, we consider $\sigma_1^2 = \sigma_2^2 = 1$, and $a_1 = a_2 = a$ (which corresponds to iv) of the previous set of remarks). Let us denote by $\vartheta = (\alpha, \beta, \mu_1, \mu_2, a) = (\vartheta_1, \cdots, \vartheta_5)$. Let us write the 2-likelihood for this parametrization:

$$\begin{aligned} p_\vartheta(z_1, z_2) &= \frac{(1-\alpha)\beta}{\alpha+\beta} T_1(z_1, z_2; \vartheta) + \frac{\alpha\beta}{\alpha+\beta} T_2(z_1, z_2; \vartheta) \\ &\quad + \frac{\alpha\beta}{\alpha+\beta} T_3(z_1, z_2; \vartheta) + \frac{\alpha(1-\beta)}{\alpha+\beta} T_4(z_1, z_2; \vartheta), \end{aligned}$$

where

$$\begin{aligned} T_1(z_1, z_2; \vartheta) &= f_{\left(\mu_1, \frac{1}{1-a^2}\right)}(z_1) f_{\left(az_1+\mu_1, \frac{1}{1-a^2}\right)}(z_2), \\ T_2(z_1, z_2; \vartheta) &= f_{\left(\mu_1, \frac{1}{1-a^2}\right)}(z_1) f_{\left(\frac{\mu_2}{1-a}, \frac{1}{1-a^2}\right)}(z_2), \\ T_3(z_1, z_2; \vartheta) &= f_{\left(\frac{\mu_2}{1-a}, \frac{1}{1-a^2}\right)}(z_1) f_{\left(\frac{\mu_1}{1-a}, \frac{1}{1-a^2}\right)}(z_2), \\ T_4(z_1, z_2; \vartheta) &= f_{\left(\frac{\mu_2}{1-a}, \frac{1}{1-a^2}\right)}(z_1) f_{(az_1+\mu_2, a^2)}(z_2). \end{aligned}$$

Concerning C3, the uniform P_0-integrability of the family $\{\log p_\vartheta(z_1, z_2) \ ; \ \vartheta \in \Theta\}$ it is enough to notice that for all $(z_1, z_2) \in \mathbb{R}^2$, and all $\vartheta \in \Theta$, we have

$$\frac{\alpha\beta}{\alpha+\beta} T_2(z_1, z_2; \vartheta) \leq p_\vartheta(z_1, z_2) \leq 4 \max_{z \in \mathbb{R}} f^2_{\left(0, \frac{1}{1-a^2}\right)}(z),$$

hence

$$\log\left(\frac{\delta^2(1-\delta/2))}{2\pi}\right) - [\frac{1}{2}(z_1^2 + z_2^2) + \frac{M}{\delta}(|z_1| + |z_2|) + \left(\frac{M}{\delta}\right)^2] \leq \log p_\vartheta(z_1, z_2) \leq \log \frac{2}{\pi}.$$

The two side of the previous inequality being independent of ϑ and P_0-integrable, we thus obtain the wanted result. Condition C4 is direct to prove.

Let us recall now that for all $i, j = 1, \cdots, 5$, the expressions of the first and second order partial derivatives are given by:

$$
\begin{aligned}
\frac{\partial}{\partial \vartheta_i} \log p_\vartheta(z_1, z_2) &= \frac{\frac{\partial}{\partial \vartheta_i} p_\vartheta(z_1, z_2)}{p_\vartheta(z_1, z_2)}, \\
\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log p_\vartheta(z_1, z_2) &= \frac{\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} p_\vartheta(z_1, z_2) p_\vartheta(z_1, z_2) - \frac{\partial}{\partial \vartheta_i} p_\vartheta(z_1, z_2) \frac{\partial}{\partial \vartheta_j} p_\vartheta(z_1, z_2)}{(p_\vartheta(z_1, z_2))^2},
\end{aligned}
$$

where for $\vartheta_1 = \alpha$, $\vartheta_3 = \mu_1$ (the same calculation holding for $\vartheta_2 = \beta$, $\vartheta_4 = \mu_2$), and $\vartheta_5 = a$, we get

$$
\begin{aligned}
\frac{\partial}{\partial \alpha} p_\vartheta(z_1, z_2) &= \frac{1}{(\alpha+\beta)^2} \left[-\beta(2\alpha+\beta) T_1(z_1, z_2; \vartheta) + \beta^2 (T_2(z_1, z_2; \vartheta) + T_3(z_1, z_2; \vartheta)) \right. \\
&\quad \left. + (1-\beta)\beta^2 T_4(z_1, z_2; \vartheta) \right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \mu_1} p_\vartheta(z_1, z_2) &= \frac{(1-\alpha)\beta}{\alpha+\beta} \left[\left(z_1 - \frac{\mu_1}{(1-a)} \right) \frac{1-a^2}{1-a} + z_2 - a z_1 - \mu_1 \right] T_1(z_1, z_2; \vartheta) \\
&\quad \frac{\alpha\beta}{\alpha+\beta} \left[\left(z_1 - \frac{\mu_1}{(1-a)} \right) \frac{1-a^2}{1-a} \right] (T_2(z_1, z_2; \vartheta) + T_3(z_1, z_2; \vartheta)),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial a} p_\vartheta(z_1, z_2) &= \frac{(1-\alpha)\beta}{\alpha+\beta} \left[\frac{-2a}{1-a^2} + a \left(z_1 - \frac{\mu_1}{1-a} \right)^2 \right. \\
&\quad \left. + \left(z_1 - \frac{\mu_1}{1-a} \right) \frac{\mu_1}{(1-a)^2} (1-a^2) + z_1 (z_2 - a z_1 - \mu_1) \right] T_1(z_1, z_2; \vartheta) \\
&\quad \frac{\alpha\beta}{\alpha+\beta} \left[\frac{-2a}{1-a^2} - \frac{1-a^2}{(1-a)^2} \left(\mu_1 (z_1 - \frac{\mu_1}{1-a}) + \mu_2 (z_2 - \frac{\mu_2}{1-a}) \right) \right. \\
&\quad \left. + a \left(\left(\frac{z_1 - \mu_1}{1-a} \right)^2 + \left(\frac{z_2 - \mu_2}{1-a} \right)^2 \right) \right] T_2(z_1, z_2; \vartheta) \\
&\quad \frac{\alpha\beta}{\alpha+\beta} \left[\frac{-2a}{1-a^2} - \frac{1-a^2}{(1-a)^2} \left(\mu_1 (z_1 - \frac{\mu_2}{1-a}) + \mu_2 (z_2 - \frac{\mu_1}{1-a}) \right) \right. \\
&\quad \left. + a \left(\left(\frac{z_1 - \mu_2}{1-a} \right)^2 + \left(\frac{z_2 - \mu_1}{1-a} \right)^2 \right) \right] T_3(z_1, z_2; \vartheta) \\
&\quad \frac{\alpha(1-\beta)}{\alpha+\beta} \left[\frac{-2a}{1-a^2} + a \left(z_1 - \frac{\mu_2}{1-a} \right)^2 \right. \\
&\quad \left. + \left(z_1 - \frac{\mu_2}{1-a} \right) \frac{\mu_2}{(1-a)^2} (1-a^2) + z_1 (z_2 - a z_1 - \mu_2) \right] T_4(z_1, z_2; \vartheta).
\end{aligned}
$$

We do not calculate here for simplicity the second order partial derivatives, but from this calculations it can be shown that the absolute values of the 2-dimensional likelihood partial derivatives of order 1 and 2 are always dominated by a bivariate function taking the form:

$$\mathrm{Pol}_{\vartheta}^{4}(|z_1|,|z_2|)\left(\sum_{i=1}^{4} T_i(z_1,z_2;\vartheta)\right),$$

where $\mathrm{Pol}_{\vartheta}^{4}(\cdot,\cdot)$ is a bivariate polynomial of order 4 which coefficients depend on ϑ and are uniformly bounded over Θ. On the other hand for all $(z_1,z_2)\in E^2$, $p_{\vartheta}(z_1,z_2)\geq \delta^2\sum_{i=1}^{4} T_i(z_1,z_2;\vartheta)$. In conclusion the partial derivatives of order 1, and 2 of the 2-dimensional log-likelihood function (with respect to the various component of ϑ) are dominated by a bivariate polynomial of order 4 which is $p_{\vartheta_0}(\cdot,\cdot)$-integrable. Finally it is good to notice that the MSDLE for a 2-mixture of gaussian linear AR(1) models is easy to implement since the gradient function of the log-likelihood is analytically known, and classical optimization procedures can be employed to solve $\dot{\ell}_{\vartheta}\,(z_1^{2k})=0$ over Θ, using various initialization conditions

6 Applications

The goal of this section is to present 3 possible areas of application for H4Ms. The topics are neurophysiology (epileptic electroencephalogram, in abbreviate EEG, signal, and Alpha and Theta waves), and kinetic (single ion channel analysis). We will propose, for each subject, precise references, and motivate application by comparing the expectations of the specialists with H4M's properties i) from v) described in introduction.

Epilectic EEG signal. Among the wide class of electrical brain signals activity, the epileptic EEG signals remains one of the most misunderstood. Various authors have proposed different kind of models, from stochastic models to dynamical models, to capture the huge complexity of epileptic EEG data series. For a first reading about this subject see for example Sackellares *et al.* (2004), see also, in this reference, some figures of epileptic EEG data series, Franaszczuk and Bergey (1998), Bergey and Franaszczuk (2001), and references theirin. The previous papers analyse precisely the behaviour of epileptic EEG, and present two different modelling approachs: one based on nonlinear chaotic models, and the other one on simple linear models. According to conclusions of these works, epileptic EEG modelling is an extraordinary difficult problem which is still open (each method having their advantages, and inconvenients). Let us describe shortly some fundamental points concerning epileptic EEG signals. All cerebral activity detectable by EEG is a reflection of synchronous neuronal activity , state considered as normal. Epileptic seizures, however, are abnormal, temporary manifestations of dramatically increased neuronal synchrony, either occuring regionally (partial seizures), or bilaterally (generalized seizure) in the brain. The period between seizures (interictal period) the EEG patterns is decribed as low to medium voltage, irregular and arrhythmic, this contrasts with the organized, rythmic, and self-sustained characteristics of EEG patterns during period out of seizure (ictal period). Iasemidis and Sackellarres (1991), study a refinement of the states, by considering the repetitive process of dynamical transitions from the interictal, to the preictal (prior to seizure), to the ictal, and to the postictal state (after seizure). Bergey and Franaszczuk (2001) exhibit that the changes occuring at the beginning of a seizure (onset) have not been studied because of the rapidly changing nature of the signal. One of the problems inherent in applying standard signal analysis methods, is that most of linear and non linear methods require long periods of relative stationary activity. From this description of epilectic EEG patterns, one can propose a H4M model taking in account these main caracteristics. Formally speaking, the Markov chain U should have a state space $\mathcal{U}$ with 5 states, representing the 5 regimes: interictal, preictal, onset, interictal, and postictal, and a highly structured matrix transition (with a small probability to remain on the onset state), to reflect the switching possibilities between these states, and the mixed Markov processes shoud be autoregressive processes (see Franaszczuk and Bergey, 1998, for intereictal state modelling with autoregressive processes), with calibrated coefficients (scale and location parameter of the noise, and coefficients of the regression from the past).

Alpha and Theta waves. Waves analysis and classification are crucial in neurophysiology, since they reflect the normality or not of the brain activity. Most waves of 7.5 Hz and higher are normal findings in the EEG of an awake adult. Waves with frequency of 7 Hz or less, are classified as abnormal for awake adult, althought they normally can be seen in children or in adults who are asleep. In certain situations EEG, waveforms of an appropriate frequency for age and state of alertness are considered abnormal because they occur at an inappropriate scalp location or demonstrate irregularities in rhytmicity or amplitude. Some waves are recognised by their shape, head distribution, and symmetry. As a result EEG signals are divided into two groups: according to their frequency context and morphomogy characteristics. In Novák *et al.* (2001), an exhaustive classification (with figures) of the existing waves forms is presented. We focus our attention on Alpha waves and Theta waves, which switch respectively between 3, and 2 frequency levels with short and long stationary stages. From Fig. 8 in Novák *et al.* (2001), HMM modelling for the Alpha wave with a 3 state Markov chain seems reasonable, when H4M modelling seems much more appropriate for the Theta wave, because of the abrupt changes and the various piecewise stationary patterns occuring in Fig.10 (with trend and notable phase-type feedback effects). A very similar sample path is simulated in appendix, using a 2-state H4M.

Single ion channel analysis. Ion channels catalyse the diffusion of ions accross membra electrical currents in the order of pico Amperes (10^{-12}A). The recording of single channel currents shows current levels corresponding to the closed and open state respectively. Transitions between these two states are very fast and in order of fractions of millisecond, and appear in the recording as rectangular jumps from one level to the other. Normally, the channels stay open for only a fraction of seconds, allowing the flux of tens of thousands of ions through the pore. HMMs provide an efficient approach for analysis of single channel currents. In fact different states of current levels are supposed during the "open" state, and are considered as hidden by the noise due to the recording instruments. It is particularly usefull for records where the signal-to-noise ratio is low or the channel kinetics is rapid, see Chung *et al.* (1990), Fredkin and Rice (1992), Chung and Gage (1998), see also an excellent overview of the HMM approach to single channel analysis in Quin *et al.* (2000a,b).

Quin *et al.* (2000b) have adressed their issues by modelling the background noise by an autoregressive process, under the strong assumption that the noise depends only on the current state. Venkataramanan and Sigworth (2002) model also the noise as an autoregressive process but make use of a more general description of state-dependent noise. The use of H4M in this context should be motivated by regarding the *global signal* (current flux plus noise) as a Markov process (current state discretization is not needed from now on), and by considering the open/close mecanism of the pore as a markovian censoring process of the global signal, following exactly the principle described in formula (4).

Sample paths simulation. In this paragraph we show, by considering two states HMM and H4Ms, choiced with the same marginal distribution and same switching source, the morphologic pattern differences one can obtain. The models considered have the same underlying chain U, with transition matrix:

$$\Pi = \begin{pmatrix} 0.92 & 0.08 \\ 0.08 & 0.92 \end{pmatrix}. \tag{37}$$

For the HMM the conditional law with respect to state 1 is a $\mathcal{N}(0,1)$, and the conditional law with respect to state 2 is a $\mathcal{N}(\mu_2, 1)$, with $\mu_2 = 1.5$. For the H4Ms, the AR(1) process X has $a_1 = a$ and noise distribution equal to $\mathcal{N}(0, 1-a^2)$, when Y has $a_2 = a$ and noise distribution equal to $\mathcal{N}(\mu_2(1-a), 1-a^2)$. By construction the HMM and H4Ms previously defined have the same marginal distribution. The following figures, show a HMM sample path (Fig.2), and H4Ms sample paths, of length $n = 200$, with different choice of a, respectively $a = 0.9$ (weakly mixing case), and $a = 0.7$ (medium mixing case), see resp. Fig.3 and 4, using both the same chain U (represented in Fig.1).

We observe that in Fig.2 the HMM pattern is very noisy but the global switching scheme is almost clear. In Fig.3 the abrupt changes and the small variance of the jumps from each AR sources make de switch design clearer. In Fig.4 the observed sample path is much more difficult to interpret, since the concatened locally stationary sequences are not different enough (because of the importance of the jumps, and the history of each AR sources) to detect clearly the instants corresponding to changes of regime. This situation is more ambiguous, in some sense, than in Fig.2, because the resulting process looks like a self-sustained process, where important jumps occur sometimes, and do not have the well known morphology of a noised state space model. Finally let us remark that pattern shown in Fig.3 imitates quite well the Theta wave pattern given in Novák *et al.* (2001).

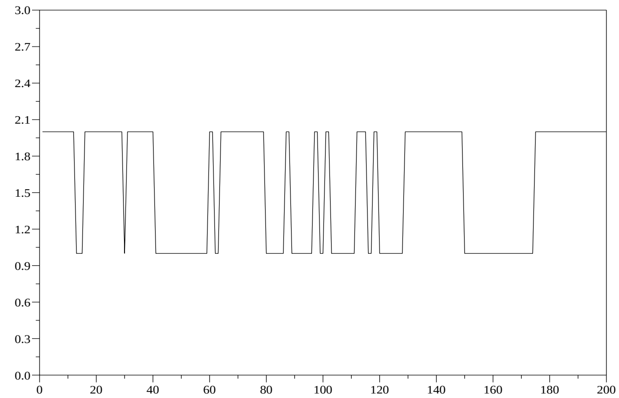

Figure 1: Sample path simulation of the underlying Markov chain U.

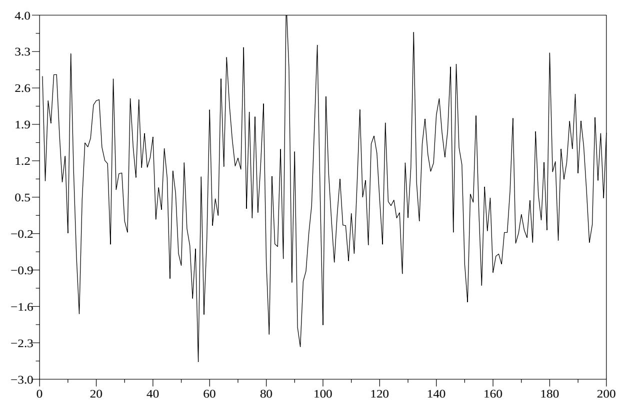

Figure 2: Sample path simulation of a HMM, $\mu_1 = 0$, $\mu_2 = 1.5$, $\sigma_i^2 = 1$, for $i = 1, 2$.

7 conclusion

In this paper we have introduced a new missing data model, called Hidden Markov Mixture of Markov model (H4M), whose observations come from different independent Markov sources, which selection at time n is done randomly according to a discrete markov chain U_n. We noticed that such a process is not Markovian, differs clearly from other Mixture of Markov Models, and do not belongs to the class of Hidden Markov models (successive observations are not independent conditionally on a finite past of U's). We have proved under mild conditions that the MSDLE, proposed by Rydén and adapted to our case, is consistent and

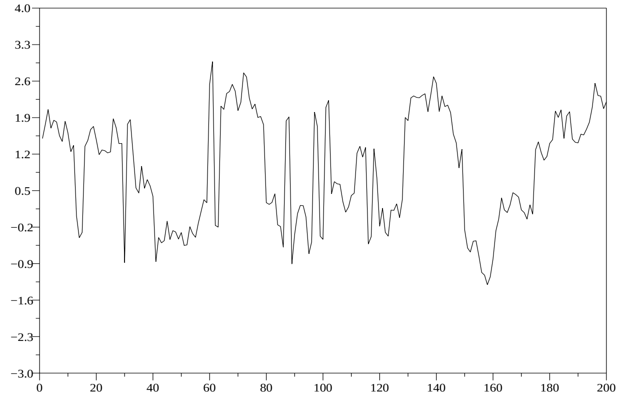

Figure 3: Sample path simulation of a H4M with two AR(1) sources, $a_i = 0.9$, for $i = 1, 2$.

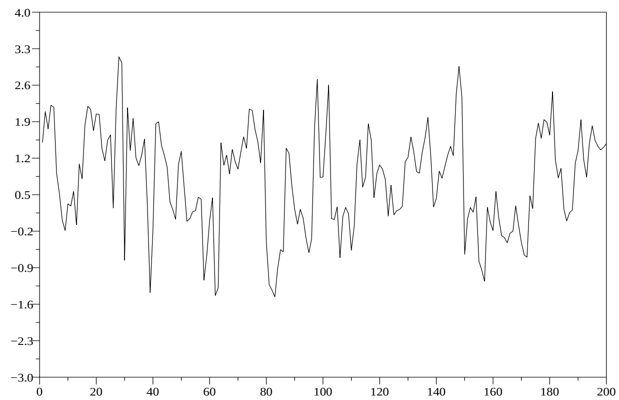

Figure 4: Sample path simulation of a H4M with two AR(1) sources, $a_i = 0.7$, for $i = 1, 2$.

asymptotically distributed. But we also pointed out that identifiability and invariant probability densities parametrization are not easy to obtain. To answer partially to the second difficulty, we proposed a Monte-Carlo approach to estimate the split data likelihood when invariant probability densities parametrization is not explicit. However we exhibit one class of models, the hidden Markov mixture of K linear autoregressive processes of order 1, $K \geq 2$, with gaussian noises, for which all the conditions needed for $\sqrt{n}$-consistency of the MSDLE are satisfied, except the classical singularity of covariance matrix involved in the asymptotic normality result. Finally it seems that H4Ms are hopefull models in areas like neurophysiology, and kinetics, which deal with data series with abrupt changes, and locally stationary sequences.

This preliminary work should be extended in the futur in two directions: (i) the research of simple conditions on dynamical systems equations and noises family distribution insuring identifiability of certain classes of H4Ms, and (ii) the study of the very challenging *exact* maximum likelihood estimator, where we claim that invariant probability densities parametrization is less crucial (in case of uniform exponential forgetting of the $X^{[i]}$'s initial conditions, for example).

Acknowledgments. The author would like to thank Laurence Denat for its encouragments and insightfull comments. He is also gratefull to the referees for constructive comments that led to a clearer presentation, and more complete analysis of the problem.

References

Bakry, D., Milhaud, X., and Vandekerkhove, P. (1997). Statistique de chaînes de Markov cachées à espace d'états fini. Le cas non stationnaire. *C. R. Acad. Sci. Paris*, Série I, **325**, 203–206.

Bar-Shalom, Y. and Li, X. R. (1993). *Estimation and Tracking : Principles, Technics, and Software.* Artech House, Boston, London.

Baraud, Y., Comte, F., Viennet, G. (2001). Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.,* **29**, 839–875.

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

Benesch, T. (2001). The Baum–Welch algorithm for parameter estimation of Gaussian autoregressive mixture models. *J. Math. Sci., New York.* **105**, 2515–2518.

Bickel, P.J., Ritov, Y., and Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Stat.* **26**, 1614–1635.

Billingsley, P. (1995). *Probability and measure.* Third Edition, Wiley series.

Bergey, G. K., and Franaszczuk, P. J. (2001). Epileptic Seizures are characterized by Changing Signal Complexity. *Clinical Neurophysiology,* **112**, 241–249.

Cai, J. (1994). A Markov unconditional variance in ARCH. *Journal of business and Economic Statistics.*

Chan, K. S., and Tong, H. (1998). A note on testing for multi-modality with dependent data. Unpublished.

Chauveau, D., Vandekerkhove, P. (2004). A Monte Carlo estimation of the entropy for Markov chains . *Preprint.*

Dacunha-Castelle, D., and Duflo, M. (1993). *Probabilités et statistiques. Problèmes à temps mobile.* Masson.

de Finetti, B. (1959). In: *Centro Internazionale Matematico Estiro (CIME), Induzione Statistica. Cremonese.* La probabilità e la stadistica nei rapporti con l'induzione secondo i diversi punto di vista. , Roma. 1–115.

Dégerine, S., Zaïdi, A. (2002). Separation of an intantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach. *submitted to IEEE, Trans. Signal. Process.*

Diaconis, P., and Freedman, D. (1980). De Finetti's theorem for Markov chains. *Ann. Probab.* **8**, 115–130.

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

Douc, R., Moulines, E., Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime. *To appear in Ann. Statist.*

Doukan, P. (1994). *Mixing properties and Examples.* Lecture Notes in Statistics. Springer, New York.

Chib, S., Kim, S., and Shepard, N. (1995). Stochastic Volatility: Likelihood inference and comparison with ARCH models. *Rev. Economic Studies,* **65**, 361–394.

Chung, S. H., Moore, J., Xia, L., Premkumar, L. S., and Gage, P. W. (1990). Characterization of single channel currents using a digital signal processing techniques based on hidden Markov models. *Phil. Trans. R. Soc. London B.,* **329**, 265–285.

Fortini, S., Ladelli, L., Petris, G., and Regazzini, E. (2002). On mixtures of distributions of Markov chains. *Stochastic Proc. Appl.,* **100**, 147–165.

Francq, C., and Roussignol, M. (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum likelihood estimator. *Statistics,* **32**, 151–173.

Franaszczuk, P. J., Bergey, G. K. (1999). An autoregressive method for the measurement of synchronization of interictaland ictal EEG signals. *Biol. Cybern.*, **81**, 3–9.

Fredkin, D. R., and Rice, J. A. (1987). Correlation functions of a function of finite-state Markov process with application to chanel kinetics. *Math. Biosci.*, **87**, 161–172.

Fredkin, D. R., and Rice, J. A. (1992). Maximum likelihood estimation and identification from single chanels recordings. *Proc. R. Soc. Lond. B.,* **249**, 125–132.

Freedman, D. (1962). Mixture of Markov processes. *Ann. Math. Statist.,* **33**, 114–118.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shift. *The review of Economics and Statistics.*

Glasbey, C. A. (2001). Non-linear autoregressive time series with multivariate gaussian mixtures as marginal distributions. *App. Statist.*, **50**, 143–154.

Hamilton, J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica,* **57**, 357–384.

Hamilton, J.D. and Susmel. R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, **64**, 307–333.

Iasemidis, L. D., and Shackellares, J.C. (1991). In: *Measuring Chaos in the human brain. The evolution with time of the spatial distribution of the largest Lyapounov exponent on the human epileptic cortex.* World Scientific, Singapore.

Leroux, B.G. (1992). Maximum likelihood estimation for Hidden markov models. *Stoch. Processes and their Applications,* **20,** 545–558.

Jalali, A. and Pemberton, J. (1995). Mixture models for time series. *J. Appl. Prob.,* **32**, 123–138.

Ji, C., Snapp, R. and Psaltis, D. (1990). Generalizing smoothness constraints from discrete samples. *Neural Computation,* **2 (2)**, 188–197.

Juang, B.H., and Rabiner, L.R. (1991). Hidden Markov models for speech recognition. *Technometrics,* **33**, 251–272.

Krishnamurthy, V., and Rydén, T. (1998). Consistent estimation of linear and non-linear autoregressive models with Markov regime. *Journal of Times Series Analysis*, **19**, 291–307.

LeGland, F., and Mevel, L. (2000). Exponential forgetting and geometric ergodicity in Hidden Markov Models. *Math. Control Signals Syst.*, **13**, No. 1, 63–93.

Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models. An algorithmic approach.* Johns Hopkins Series in the Mathematical Sciences.

Novák, D., Lhotská, L., Eck, V., and Sorf, M. (2001). EEG and VEP signal processing. *Preprint of Czech Technical University.*

Pham, D-T., Garat, P. (1997). Blind separation of mixture of independent sources through a quasi-maximum likelihood approach, *IEEE, Trans. on Signal Processing,* **4**, 1712–1725.

Quin, F., Auerbach, A., and Sachs, F. (2000a). A direct optimization approach to hidden Markov modelling for single channel kinetics. *Biophys. J.,* **79**, 1915–1927.

Quin, F., Auerbach, A., and Sachs, F. (2000b). Hidden Markov modelling for single channel kinetics with filtering and correlated noise. *Biophys. J.,* **79**, 1928–1944.

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

Rydén, T. (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.,* **22**, 1884–1895.

Sackellares, J. C., Iasemidis, L. D., Shiau, D-S., Gilmore, R., and Roper, S. N. (2004). In:*Chaos in the brain? Epilepsy-When chaos fails.* World Scientific, Singapore, in press.

Stout, W. (1977). Almost sure convergence. *Academic Press.*

Teicher, H. (1963). Identifiability of finite mixture. *Ann. Math. Statist.,* **34**, 1265–1269.

Tong, H. (1990). *Non-linear Time Series.* New York: Oxford University Press.

Tugnait, J. K. (1982). Detection and estimation for abruptly changing systems. *Automatica,* **18**, 607–615.

Wong, C. S., and Li, W. K. (2000). On a mixture autoregressive model. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **62**, 95–115.

Venkataramanan, L. and Sigworth, F. J. (2002). Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.,* **82**, 1930–1942.

Wong, C. S., and Li, W. K. (2001). On a logistic mixture autoregressive model. *Biometrika*

88, 833–846.

Zhengyan, L., and Chuanrong, L. (1996). *Limit Theory for Mixing Dependent Random Variables*. Kluwer Academic Publishers.