Statistical inference for partially hidden Markov models

Laurent BORDES[1] Pierre VANDEKERKHOVE[2]

[1]University of Technology of Compiègne

Laboratoire de Mathématiques Appliquées de Compiègne

[2]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 standard parametric Hidden Markov Model (HMM), for which informations on the latent Markov chain are given since this one reaches a fixed state (and until it leaves this state). We study, under mild conditions, the consistency and asymptotic normality of the maximum likelihood estimator. We point out also that the underlying Markov chain does not need to be ergodic, and that identifiability of the model is not tractable in a simple way (unlike standard HMMs), but can be studied using various technical arguments.

Abbreviated title. Consistency of MLE for general HMM.

AMS 1991 classification. Primary 62M09, 62L20.

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

1 Introduction

Hidden Markov models (HMMs) form a wide class of discrete-time stochastic processes, intensively used in many areas as speech recognition (for a good introduction, see Rabiner 1989, Juang and Rabiner 1991), biology for heterogeneous DNA sequences analysis (Churchill 1989), neurophysiology (Freidkin and Rice 1992), econometrics (Kim *et al.* 1998), time series analysis (DeJong and Shepard 1995, Chan and Ledolter 1995, MacDonald and Zucchini 1997), and image segmentation (Choi and Baraniuk 2001, for a recent work). The main focus of these efforts have been algorithms for fitting of these models. For finite hidden state space models, the first contribution is due to Baum *et al.* (1970) who proposed an early and elegant application of the expectation-maximization principle (Dempster *et al.* 1977), known as the "forward-backward" procedure. The more difficult issue of hidden Markov models with continuous state space has been also studied during the 1990s, using preferably simulation-based approaches allowed by the recent developments of Markov chain Monte Carlo methods (Chib *et al.* 1998, Durbin and Koopman 1997, Cappé *et al.* 2002). There has been comparatively little work on the study of inferential properties of the likelihood methods in these models. Baum and Petrie (1966) have shown the consistency and asymptotic normality of the maximum likelihood estimator (MLE) in the case of finite valued observable and latent variables. These results have been extended recently in various papers by Leroux (1992), Bickel and Ritov (1996), Bickel *et al.* (1998), Bakry *et al.* (1997), Legland and Mevel (2000). The most recent paper on statistical inference for general HMMs is due to Douc and Matias (2001), who prove the consistency and asymptotic normality of the MLE when the hidden Markov chain is not necessary stationary and takes value in a general topological space.

In the present paper we introduce a new class of missing data Markov model, the so-called partially hidden Markov models (PHMM, note however that this term is used by Forchhammer and Rissanen 1996 in another context), naturally connected to HMMs, but which do not belong entirely to this class of models. Typically the PHMM is built to model discrete or continuous observations whose law depends on a discrete Markov chain, exactly as the usual discrete HMM, except for the fact that information on the state of the latent Markov chain is given from time to time. This model should find possible applications in reliability modelling : a large literature on degradation, deteriorating, or damage processes is available. Singpurwalla (1995) discussed a class of degradation models based on Gamma processes, Bagdonavičius and Nikulin (2001) considered the same class of processes, and they introduced random time scale governed by covariates. Other kind of processes was used to modelize the same type of phenomenon, like markov additive processes, Gaussian processes with trend, marked point processes (see e.g. Khale and Wendt, 2000). In many situation, monitoring is done at periodic times, and the measurements are only symptoms of the true unobservable degradation process of the system under study. However, generally, the only reachable degradation state is the system failure. This is the reason why PHMM could be an alternative model to some existing degradation models of the statistical literature.

This model could be also of interest in software reliability modelling. Chen and Singpurwalla (1997) gave an overview of software reliability models based on self-exciting random processes. Durand and Gaudouin (2003) introduced a new class of models by considering that interarrival times of bugs were exponentially distributed with random parameters taking values in a finite set, and governed by a time homogeneous Markov chain. These assumptions lead naturally to an HMM model that the authors estimate by using a EM algorithm. In such a case we can consider that the debbuging state as a specific observable state leading naturally to a PHMM.

The same kind of phenomena are studied in medicine, see Guihenneuc-Jouyaux *et al.* (2000), Jackson and Sharples (2002), to model markers of disease progression by a hidden Markov model, the "failure" state being, in that case, the death of the patient which occurs at a random time. However our model should be adapted to this context in order to take into account several pieces of trajectories possibly censored.

An important fact concerning PHMM is that the visits to a specific state allow regeneration of the underlying Markov chain. This leads to factorization of the likelihood function into independent and identically distributed pieces of sample paths with random length, allowing the study of the MLE in a easier way from classical HMM framework (see Leroux, 1992; Bickel *et al.*, 1998).

In the sequel of this paper we present precisely, in Section 2, the PHMM itself, asymptotic properties of functionals of its trajectories, and the main assumptions. In Section 3 we discuss some identifiability conditions for this model, in Section 4 and 5 we prove respectively the consistency and the asymptotic normality of the MLE under mild conditions. It is shown in

Appendix A that some basic models can be identified.

2 Notations and preliminary results

A hidden Markov model (HMM) is a discrete-time stochastic process $(X_n, Y_n)_{n\geq 1}$ such that (i) $(X_n)_{n\geq 1}$ is a finite-state Markov chain, and (ii) given $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$ is a sequence of conditionally independent random variables and the conditional distribution of Y_n depends on $(X_n)_{n\geq 1}$ only through X_n. It is easy to check that $(X_n, Y_n)_{n\geq 1}$ is a Markov chain, whereas it is not longer true for $(Y_n)_{n\geq 1}$ alone. The name HMM is motivated by the assumption that $(X_n)_{n\geq 1}$ is not observable, so that inference has to be based on $(Y_n)_{n\geq 1}$ alone. Suppose that $E = \{1, 2, \dots, a\}$ is the state-space of the chain $(X_n)_{n\geq 1}$ with $a \geq 3$. Suppose also that (X_n, Y_n) is observable conditionally on $\{X_n = a\}$; otherwise only Y_n can be observed. This is a situation where the Markov model is partially observed and then such a model will be called a partially hidden Markov model (PHMM).

Let us consider $(X_n)_{n\geq 1}$ a homogeneous irreducible Markov chain on E with the transition probability matrix $\alpha = (\alpha_{ij})_{1\leq i,j\leq a}$ (with $\alpha_{ij} = \alpha(i,j)$ is the probability that $\{X_{n+1} = j\}$ given $\{X_n = i\}$). The transition probabilities will be parametrized by a parameter $\phi \in \Phi$, i.e. $\alpha_{ij} = \alpha_{ij}(\phi)$, where $\Phi \subset \mathbb{R}^q$. The fully observed process $(Y_n)_{n\geq 1}$ is assumed to take values in some measurable, separable and complete space F, and the conditional distributions of Y_n are all assumed to be dominated by some σ-finite measure μ on the Borel σ-field $\mathcal{B}(F)$. Moreover, the corresponding conditional densities are assumed to belong to some parametric family $\mathcal{G} = \{g(\cdot;\theta) \ : \ \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$, and the parameters of those densities are functions of X_n as well as of ϕ, and then the conditional density of Y_n given $\{X_n = i\}$ is $g(\cdot;\theta_i(\phi))$. The most common parameterization is $\phi = (\alpha_{11}, \alpha_{12}, \dots, \alpha_{aa}, \theta_1, \dots, \theta_a)$ with $\alpha_{ij}(\cdot)$ and $\theta_i(\cdot)$ being the coordinate projections (called the "usual parameterization" in Rydén, 1997). An other example of parameterization can be the coordinate projections for $\alpha_{ij}(\cdot)$ and $\theta_i = (i, \sigma_i^2)$ with $g(\cdot;\theta_i)$ the Gaussian density with mean i, variance σ_i^2. We consider, for more generality, the framework $\alpha(a,a) \neq 0$, the adaptation of all our proofs to the case $\alpha(a,a) = 0$ being straightforward.

The probabilistic model is given by

$$\left((E \times F)^{\mathbb{N}}, (\mathcal{P}(E) \otimes \mathcal{B}(F))^{\otimes \mathbb{N}}, (X_n, Y_n)_{n\geq 1}, \alpha_\phi, \phi \in \Phi\right).$$

where $\mathcal{P}(E)$ is the subset family of E. Let us consider the following initial time τ_1 such that $X_{\tau_1-1} \neq a$ and $X_{\tau_1} = a$, and

$$\tilde{\tau}_1 = \inf\{n \geq \tau_1 \ : \ X_{n-1} = a, X_n < a\},$$

then, let us define for $p \geq 2$

$$\tau_p = \inf\{n > \tilde{\tau}_{p-1} \ : \ X_n = a\},$$

and

$$\tilde{\tau}_p = \inf\{n > \tau_p \ : \ X_n < a\}.$$

Sequences $(\tau_p)_{p\geq 1}$ and $(\tilde{\tau}_p)_{p\geq 1}$ are entry times in $\{a\}$ and $E\backslash\{a\}$ respectively. Define for $p \geq 1$, $N_p = \tilde{\tau}_p - \tau_p$ and $\tilde{N}_p = \tau_{p+1} - \tilde{\tau}_p$ and therefore $(N_p)_{p\geq 1}$ and $(\tilde{N}_p)_{p\geq 1}$ are the sequences of sojourn times in $\{a\}$ and $E\backslash\{a\} \stackrel{\text{def.}}{=} E_a$ respectively.
Observations of the PHMM consist in $(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})_{k\geq 1} = \big((Y_n)_{\tau_1 \leq n \leq \tau_{k+1}-1}, (\tau_i)_{1\leq i \leq k+1}, (\tilde{\tau}_i)_{1\leq i\leq k}\big)_{k\geq 1}$, and we denote by $\mathcal{F}_k$ the σ-field generated by $\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}$. Such information contains the fact that between τ_i and $\tilde{\tau}_i - 1$ the Markov chain X is observed at state a, whereas it is at state E_a between $\tilde{\tau}_i$ and $\tau_{i+1} - 1$. For convenience we define

$$\mathbf{Y}_k^l \stackrel{\text{def.}}{=} \{(Y_j)_{k\leq j\leq l}\}.$$

P1. By a general result on regenerative cycles for discrete Markov chains (see, e.g., Brémaud, 1998, page 86-87), the pieces of trajectories $(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}, \mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1})_{j\geq 1}$ are independent and identically distributed. For simplicity we consider the random variable $(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{\tilde{N}+N})$ having the same law as $(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}, \mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1})$, where N and $\tilde{N}$ have respectively the same law as N_i and $\tilde{N}_i$.

P2. $(N_p)_{p\geq 1}$ and $(\tilde{N}_p)_{p\geq 1}$ are two sequences of independent and identically distributed random variables:

$$P_\phi(N_i = n_i, \tilde{N}_i = \tilde{n}_i : i = 1, \ldots, k) = \prod_{i=1}^{k} \left(\sum_{x_1^{\tilde{n}_i} \in (E_a)^{\tilde{n}_i}} \alpha(a, x_1) \ldots \alpha(x_{\tilde{n}_i}, a) \right) \times \alpha(a,a)^{n_i - 1}.$$

Thus we have:

$$\begin{aligned} P_\phi(N_i = n) &= (\alpha(a,a))^{n-1}(1 - \alpha(a,a)), \\ P_\phi(\tilde{N}_i = \tilde{n}) &= \frac{1}{1-\alpha(a,a)} \sum_{x_1^{\tilde{n}} \in (E_a)^{\tilde{n}}} \alpha(a, x_1) \ldots \alpha(x_{\tilde{n}}, a). \end{aligned}$$

We can check that $\tilde{N}_i$ are non degenerated integer valued random variables, since a is a recurrent point for the chain X. Indeed, denoting the first time of return in state a for the chain X, *i.e.* $T_a = \inf\{n \geq 1; X_n = a | X_0 = a\}$, the $\tilde{N}_i$'s have the same distribution as $T_a - 1$ conditionally on $\{T_a > 1; X_0 = a\}$. Now since a is a recurrent state for the chain X, we have $P_\phi(T_a < \infty) = 1$ and $P_\phi(T_a > 1) = 1 - \alpha(a,a)$, and then we have following relations

$$P_\phi(\tilde{N}_i = \tilde{n}) = P_\phi(T_a = \tilde{n}+1 | T_a > 1) = \frac{P_\phi(T_a = \tilde{n}+1)}{P_\phi(T_a > 1)} = \frac{P_\phi(T_a = \tilde{n}+1)}{1-\alpha(a,a)}.$$

Finally we get $P_\phi(1 \leq \tilde{N}_i < \infty) = 1$.

P3. The law of the random element $(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{\tilde{N}+N}; N, \tilde{N}) \in \mathbb{R}^{\mathbb{N}} \times \mathbb{N}^2$ is given by:

$$p_\phi(y_1^n, y_{n+1}^{n+\tilde{n}}; N = n, \tilde{N} = \tilde{n}) = \alpha(a,a)^{n-1} \prod_{j=1}^{n} g(y_j; \theta_a(\phi))$$

$$\times \sum_{x_{n+1}^{n+\tilde{n}} \in (E_a)^{\tilde{n}}} \alpha(a, x_{n+1}) g(y_{n+1}; \theta_{x_{n+1}}(\phi)) \cdots \alpha(x_{n+\tilde{n}-1}, x_{n+\tilde{n}}) g(y_{n+\tilde{n}}; \theta_{x_{n+\tilde{n}}}(\phi)) \alpha(x_{n+\tilde{n}}, a).$$

P4. It follows from P1 that $(\log p_\phi(\mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1}))_{j\geq 1}$ and $(\log p_\phi(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}))_{j\geq 1}$ are respectively two sequences of i.i.d. random variables.

P5. Obvious calculations lead to independence of $(\mathbf{Y}_1^N, N)$, and $(\mathbf{Y}_{N+1}^{\tilde{N}+N}; \tilde{N})$.

From now on we use the notation $\mathbf{Z}_1^n = \{Z_k \ ; \ 1 \leq k \leq n\}$ for all processes and we note by $p_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$ the likelihood function of the observations. We have

$$p_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) = \prod_{i=1}^{k} p_\phi(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i-1} | N_i) p_\phi(\mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}, \tilde{N}_i) p_\phi(N_i),$$

by P1, and P**??**. The log-likelihood function $\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$ can be written

$$\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) = \sum_{j=1}^{k} \log p_\phi\left(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i-1}, \mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}\right), \tag{1}$$

where

$$\log p_\phi\left(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i-1}, \mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}\right) = \log p_\phi(\mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}, \tilde{N}_i) + \log p_\phi(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i-1} | N_i) + \log p_\phi(N_i). \tag{2}$$

We propose to state now the assumptions for future reference. For simplicity we will denote by $\dot{\ell}_k(\phi) = \frac{\partial}{\partial \phi}\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$, and $\ddot{\ell}_k(\phi) = \frac{\partial^2}{\partial \phi \partial \phi^T}\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$, where ϕ^T denotes the transposed vector of the column vector ϕ.

We call a *squeleton* of a stochastic matrix α on $E \times E$ the set of locations (i,j) in $E \times E$ such that $\alpha(i,j) \neq 0$. We write φ the application which to a stochastic matrix associates its squeleton, and $\mathcal{H}$ the family of all squeletons of E-square stochastic matrices. Let δ be a real number in $(0,1)$, for $I \in \mathcal{H}$, we write

$$\Phi_I = \{\phi \in \mathbb{R}^q \ ; \ \varphi(\alpha(\phi)) = I, \ \alpha_{ij}(\phi) \geq \delta \ , \ \forall (i,j) \in I, \text{ and } \theta_i(\phi) < \theta_j(\phi) \text{ if } 1 \leq i < j \leq a\},$$

where the order relation between $\theta_i(\phi)$ and $\theta_j(\phi)$ must be understood with respect to the lexical order.

C1. For all $I \in \mathcal{H}$ the set Φ_I is supposed to be a compact set. The full parametrical space Φ is equal to $\cup_{I \in \mathcal{H}} \Phi_I$.

C2. The true parameter ϕ_0 is an interior point of Φ, and $\alpha(\phi_0)$ is irreducible.

C3. There exist deterministic functions g_1 and g_2 defined on F such that:

$$g_1(y) \leq g(y; \theta_i(\phi)) \leq g_2(y), \quad \forall (y, \phi, i) \in F \times \Phi \times E,$$

and

$$\int_F |\log(g_i(y))| \, g_2(y) d\mu(y) \leq M < +\infty, \quad i = 1, 2.$$

C4. The functions $\phi \mapsto g(\cdot; \theta_i(\phi))$ are $\mu - a.e.$ twice continuously differentiable on Φ, and the functions $\phi \mapsto \log \alpha_{ij}(\phi)$ are twice continuously differentiable on Φ, for all $1 \leq i, j \leq a$.

C5. Write $\phi = (\phi_1, \ldots, \phi_q)$, and let $\| \cdot \|$ be the Euclidean norm on $\mathbb{R}^q$. There exists a $\xi > 0$ such that

(i) there exists a function $g^{(1)}$, such that for all $1 \leq i \leq q$, $x \in E$, and $y \in F$,

$$\sup_{\|\phi - \phi_0\| < \xi} \left| \frac{\partial}{\partial \phi_i} \log g(y; \theta_x(\phi)) \right| \leq g^{(1)}(y), \tag{3}$$

with

$$\int_F g^{(1)}(y) g_2(y) dy < \infty, \quad \text{and} \quad \int_F (g^{(1)}(y))^2 g_2(y) dy < \infty. \tag{4}$$

(ii) there exists a function $g^{(2)}$, such that for all $1 \leq i, j \leq q$, $x \in E$, and $y \in F$,

$$\sup_{\|\phi-\phi_0\|<\xi} \left| \frac{\partial^2}{\partial\phi_i\partial\phi_j} \log g(y;\theta_x(\phi)) \right| \leq g^{(2)}(y), \text{ such that } \int_F g^{(2)}(y) g_2(y) dy < \infty. \tag{5}$$

Remarks. In general, if ϕ_0 is the true parameter of a discrete HMM, there exist other parameters ϕ, equivalent to ϕ_0, *i.e.* the law of the Y's is indistinguishable under ϕ and ϕ_0. A well known class of parameters equivalent to ϕ_0, is the class of all the parameters ϕ such as, for all $(i,j) \in E^2$,

$$\left\{ \begin{array}{lll} \alpha(i,j)(\phi) & = & \alpha(\sigma(i),\sigma(j))(\phi_0), \\ \theta_i(\phi) & = & \theta_{\sigma(i)}(\phi_0). \end{array} \right. \tag{6}$$

where σ is an arbitrary permutation on E. To avoid this confusion we suppose (as it is suggested in Rydén, 1994) in conditions C1–2, that the true parameter ϕ_0 is such that $\theta_1(\phi_0) < \theta_2(\phi_0) < ... < \theta_a(\phi_0)$, which does not lead to a loss of generality according to the previous remark (this explains the definition of the parametrical spaces Φ_I). It is easy to construct such compact parametrical space satisfying the previous order restriction.

Let us denote by $\tilde{N}_\phi(\Omega)$ the support of $\tilde{N}$ under ϕ, and note that from C**??**, it is an infinite subset of $\mathbb{N}$. We separate from now on Φ in two disjoint parts: $\Phi_1 = \cup_{I\in\mathcal{H}_0}\Phi_I$ and $\Phi_2 = \cup_{I\in\mathcal{H}_0^c}\Phi_I$, where

$$\mathcal{H}_0 = \left\{ I \in \mathcal{H} \; : \; \forall\phi \in \Phi_I, \; \tilde{N}_0(\Omega) \subseteq \tilde{N}_\phi(\Omega) \right\}. \tag{7}$$

We point out here that according to the positions of zero values in the Markov transition matrix, the support of $\tilde{N}$ could not coincide under ϕ and ϕ_0 (if for example the chain is aperiodic under ϕ and periodic under ϕ_0). We mention this fact because it is the source of log-likelihood degeneracy for ϕ belonging to Φ_2 (see Lemma 3).

Let us notice that C**??** (uniform irreducibility of α) leads to the finite expectation of $N + \tilde{N}$, and the finite expectation of each kind of sojourn time. On the other hand, conditions C**??**–C**??** are equivalent to the standard conditions found in HMM literature, see for example Bickel *et al.* (1998, p. 1618). Nevertheless condition A1 in Bickel *et al.* (1998), insuring stationarity, and hence ergodicity, of the latent Markov chain is alleviated into weaker conditions C**??**–C**??** which just deal with the graph structure of the Markov chain, but do not involve aperiodicity.

3 Identifiability

In this section we will discuss the identifiability conditions in closer detail. From now on, when $\phi = \phi_0$ we denote P_ϕ and E_ϕ by P_0 and E_0 respectively. Suppose the following conditions are satisfied.

I1. (Identifiability). The family of mixtures of at most r elements of $\{g(y;\theta);\ \theta \in \Theta\}$ is identifiable. This condition means that if $\theta_i \in \Theta$ and $\theta'_i \in \Theta$, for $i = 1, \ldots, a$, with $\theta_i \neq \theta_j$ and $\theta'_i \neq \theta'_j$, if $i \neq j$, and $(\beta_1, \ldots, \beta_a)$ and $(\beta'_1, \ldots, \beta'_a)$ are probability vectors, then

$$\sum_{i=1}^{a} \beta_i g(y;\theta_i) = \sum_{j=1}^{a} \beta'_j g(y;\theta'_j)\ \mu - a.e. \Rightarrow \exists\ !\ \sigma \in \mathcal{P}_a : (\beta_i, \theta_i) = (\beta'_{\sigma(i)}, \theta'_{\sigma(i)}),\ i = 1, \ldots, a,$$

where $\mathcal{P}_a$ denotes the set of all the permutations on $\{1, \ldots, a\}$, that is we can identify the components of the mixing distribution.

I2. The mapping from $[0,1]^{a^2}$ to $[0,1]^{\mathbb{N}}$ defined by

$$\alpha \mapsto \left\{ \prod_{i=0}^{n} \alpha(x_i, x_{i+1});\ x_0^{n+1} \in \{a\} \times (E_a)^n \times \{a\},\ n \geq 1 \right\},$$

where $\alpha = (\alpha_{ij})_{1 \leq i,j \leq a}$ is a $a \times a$ stochastic matrix, is injective.

The first assumption holds, for example for the Poisson family, the negative exponential family and the normal family with fixed variance, see Prakasa Rao (1992) and Lindsay (1995) for more details on mixture identifiability. Still now we have just discussed identifiability of one-dimensional distributions, but it turns out that this property carries over to multidimensional ones, see Teicher (1967); that is, the family of at most r elements of $\prod_{k=1}^{m} g(y_k;\theta_k)$ (over Θ^m) is identifiable.

The second assumption, which appears to be nontrivial to check, is needed to prove MLE consistency. Let us remark that this kind of assumption does not occur in standard HMMs. In fact, as it is pointed out in Leroux (1992), or Rydén (1994), the true parameter ϕ_0 of a stationary HMM is entirely determined by the 2-dimensional distribution of X over Φ.

However as it appears on the 3 following examples, to check the injectivity condition I2 we need to consider trajectories length of order a. We claim that the 3 following models satisfy the injectivity condition I2, the proofs of these claims are relegated to appendix A.

M1 Strictly positive transitions in dimension 3: 3×3 stochastic matrices for which all the components are non null.

M2 Periodic transitions in dimension 4: the four states model of Figure **??** is clearly periodic with period equal to 2.

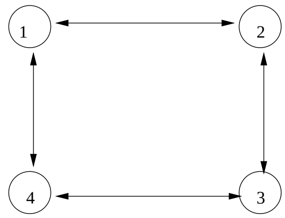

Figure 1: A periodic model with $a = 4$ states

M3 Pure degradation model: the Figure **??** describes the graph of a degradation model

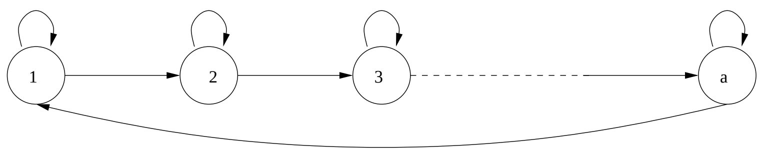

Figure 2: A degradation model with $a \geq 2$ degradation states

for which non null terms of the transition matrix α are $\alpha(i,i)$ and $\alpha(i,i+1)$ for $i = 1,\ldots,a-1$, and $\alpha(a,1)$ and $\alpha(a,a)$ on its last row.

4 Consistency

This section is devoted to the proof of MLE consistency. For this aim we need to establish some technical lemmas. The first one deals with the majorisation of a conditional likelihood, by a convenient quantity. This result is crucial for the proof of Lemma **??**.

Lemma 1 *For all $\phi \in \Phi$, and $\mathbf{y} = (y_{i_1}, y_{i_2}, \ldots, y_{i_j}) \in F^j$, where $1 \leq i_1 < i_2 < \cdots < i_j \leq \tilde{n}$, we have*

$$P_\phi(\tilde{N} = \tilde{n}) \prod_{k=1}^{j} g_1(y_{i_k}) \leq p_\phi(\mathbf{y}, \tilde{N} = \tilde{n}) \leq P_\phi(\tilde{N} = \tilde{n}) \prod_{k=1}^{j} g_2(y_{i_k}).$$

Proof. To avoid technicalities we prove the result for $j = 1$ and $i_1 = l$. For simplicity we note $x_1^m(k) = (x_i)_{1\le i\le m, i\neq k}$ and $dy_1^m(k) = \bigotimes_{j=1,j\neq k}^m dy_j$.

$$
\begin{aligned}
& p_\phi(y_l, \tilde{N} = \tilde{n}) \\
= & \int_{F^{\tilde{n}-1}} \sum_{x_0^{\tilde{n}} \in \{a\}\times E_a^{\tilde{n}}} \left(\prod_{j=0}^{\tilde{n}-1} \alpha(x_j, x_{j+1}) g(y_{j+1}, \theta_{x_{j+1}}(\phi)) \right) \frac{\alpha(x_{\tilde{n}}, a)}{1-\alpha(a,a)} dy_1^{\tilde{n}}(l) \\
= & \sum_{x_l \in E_a} \left[\sum_{x_0^{\tilde{n}-1}(l) \in \{a\}\times E_a^{\tilde{n}-1}} \left(\prod_{j=0, j\neq l-1}^{\tilde{n}-1} \alpha(x_j, x_{j+1}) \int_F g(y_{j+1}, \theta_{x_{j+1}}(\phi)) dy_{j+1} \right) \right. \\
& \left. \times \alpha(x_{l-1}, x_l) g(y_l, \theta_{x_l}(\phi)) \right] \frac{\alpha(x_{\tilde{n}}, a)}{1-\alpha(a,a)} \\
\le & \sum_{x_l \in E_a} \left[\sum_{x_0^{\tilde{n}-1}(l) \in \{a\}\times E_a^{\tilde{n}-1}} \prod_{j=0, j\neq l-1}^{\tilde{n}-1} \alpha(x_j, x_{j+1}) \alpha(x_{l-1}, x_l) g_2(y_l) \right] \frac{\alpha(x_{\tilde{n}}, a)}{1-\alpha(a,a)} \\
\le & P_\phi(\tilde{N} = \tilde{n}) g_2(y_l).
\end{aligned}
$$

Following the above lines we get the left hand side of the double side inequality. □

The two following lemmas establish the P_0–integrability of functionals of $\tilde{N}$. It will be useful results for proving integrability assumptions.

Lemma 2 *Under conditions C1–3 the* $\mathbb{N}$*-valued random variable* $\tilde{N}$ *belongs to* $L^m(P_0)$ *for all* $m > 0$.

Proof. Following Neuts (1994, p.46) $\tilde{N}$ has a discrete phase-type distribution, the moments of which are all finite. □

Lemma 3 *Under conditions C1–3, for all* $\phi \in \Phi_1$*, the* $\mathbb{R}$*-valued random variable* $\log P_\phi(\tilde{N})$ *is* P_0*–integrable.*

Proof. Let ϕ be a parameter of Φ_1, for all $\tilde{n} \in \tilde{N}_\phi(\Omega) \supseteq N_0(\Omega)$ there exists at least one path $(x_1, \ldots, x_{\tilde{n}}) \in E_a^{\tilde{n}}$ such that $P_\phi(X_{\tilde{n}+1} = a, X_{\tilde{n}} = x_{\tilde{n}}, \ldots, X_1 = x_1 | X_0 = a) > 0$, hence from the definition of Φ_1, it comes for all $\phi \in \Phi_1$

$$
\begin{aligned}
P_\phi(\tilde{N} = \tilde{n}) & \ge \frac{1}{1-\alpha(a,a)} \alpha(a, x_1)\alpha(x_1, x_2)\ldots\alpha(x_{\tilde{n}}, a) \\
& \ge \frac{\delta^{\tilde{n}}}{1-\delta}.
\end{aligned}
$$

Thus for all $\tilde{n} \in \tilde{N}(\Omega)$ we can write for all $\phi \in \Phi_1$

$$|\log P_\phi(\tilde{N} = \tilde{n})| \leq \tilde{n}|\log(\delta)| - \log(1-\delta). \tag{8}$$

Then by Lemma **??** $\log P_\phi(\tilde{N})$ is P_0–integrable. □

Lemma 4 *i) Under conditions C**??**–**??**, for all $\phi \in \Phi_1$, and all $i \geq 1$, the $\mathbb{R}$-valued random variables $\log p_\phi\left(\mathbf{Z}_{\tau_i}^{\tau_{i+1}-1}\right)$ are P_0–integrable. In addition, we have the strong law of large numbers on the $\log p_\phi\left(\mathbf{Z}_{\tau_i}^{\tau_{i+1}-1}\right)$'s, i.e.*

$$\frac{1}{k}\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) \underset{k\to\infty}{\longrightarrow} \mathcal{E}_0(\phi) \quad P_0 - a.s.,$$

where $\mathcal{E}_0(\phi) = E_0\left[\log p_\phi\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right)\right]$.

*ii) Otherwise, under conditions C**??**–**??**, we have the following degenerative behaviour*

$$\sup_{\phi\in\Phi_2}\frac{1}{k}\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) \underset{k\to\infty}{\longrightarrow} -\infty \quad P_0 - a.s.$$

Proof. i) From the factorization of the likelihood it is sufficient to show the P_0–integrability of each term in (**??**). We begin with the simplest term $\log p_\phi(N)$. Since the following inequality holds

$$\begin{aligned} |\log p_\phi(N)| &\leq (N-1)|\log\alpha(a,a)| + |\log(1-\alpha(a,a))|, \\ &\leq N\log(\delta), \end{aligned} \tag{9}$$

the P_0–integrability of $\log p_\phi(N)$ is a consequence of the finite P_0-expectation of N, and the fact that $\alpha(a,a) \neq 0$ and 1 (general assumptions and C**??**). For the second term it comes directly from C**??**

$$|\log p_\phi(\mathbf{Y}_1^N)| \leq \sum_{j=1}^N |\log g_2(Y_j)|. \tag{10}$$

Then we check directly from C**??** that

$$E_0[|\log p_\phi(\mathbf{Y}_1^N|N)|] \leq ME_0(N) < \infty.$$

We treat now the first term in (**??**). Let us consider for $\phi \in \Phi_1$ and $\tilde{n} \in \tilde{N}_0(\Omega)$

$$I(\tilde{n}) \overset{\text{def.}}{=} E_0\left[|\log p_\phi(\mathbf{Y}_1^{\tilde{N}}, \tilde{N})| \,\middle|\, \tilde{N} = \tilde{n}\right].$$

By Lemma **??**, we get that for all $\tilde{n} \in \tilde{N}_0(\Omega)$

$$\sum_{j=1}^{\tilde{n}} \log g_1(Y_j) + \log P_\phi(\tilde{N} = \tilde{n}) \leq \log p_\phi(\mathbf{Y}_1^{\tilde{N}}, \tilde{N} = \tilde{n}) \leq \sum_{j=1}^{\tilde{n}} \log g_2(Y_j).$$

Then we get for all $\tilde{n} \in \tilde{N}_0(\Omega)$

$$\begin{aligned}
I(\tilde{n}) &\leq \sum_{i=1}^{2}\sum_{j=1}^{\tilde{n}} E_0\left[|\log g_i(Y_j)| \,\Big|\, \tilde{N} = \tilde{n}\right] + |\log P_\phi(\tilde{N} = \tilde{n})| \\
&\leq \sum_{i=1}^{2}\sum_{j=1}^{\tilde{n}} \int_{F^{\tilde{n}}} \sum_{x_0^{\tilde{n}} \in \{a\}\times E_a^{\tilde{n}}} |\log g_i(y_j)| \prod_{l=1}^{\tilde{n}} \alpha_0(x_{l-1}, x_l) g(y_l; \theta_{x_l}(\phi_0)) \\
&\quad \times \frac{\alpha_0(x_{\tilde{n}}, a)}{1-\alpha_0(a,a)} dy_1^{\tilde{n}} \; + |\log P_\phi(\tilde{N} = \tilde{n})| \\
&\leq \sum_{i=1}^{2}\sum_{j=1}^{\tilde{n}} \int_{F} \sum_{x_0^{\tilde{n}} \in \{a\}\times E_a^{\tilde{n}}} |\log g_i(y_j)| \alpha_0(x_{j-1}, x_j) g_2(y_j) \prod_{l=1;\; l\neq j}^{\tilde{n}} \alpha_0(x_{l-1}, x_l) \\
&\quad \times \frac{\alpha_0(x_{\tilde{n}}, a)}{1-\alpha_0(a,a)} dy_j + |\log P_\phi(\tilde{N} = \tilde{n})| \\
&\leq M \sum_{i=1}^{2}\sum_{j=1}^{\tilde{n}} 1 + |\log P_\phi(\tilde{N} = \tilde{n})| \\
&\leq 2M\tilde{n} + \tilde{n}|\log(\delta)| + |\log(1-\delta)|,
\end{aligned}$$

where the last inequality arises from inequality (**??**). As $\tilde{N}$ is P_0–integrable, we obtain the existence of the desired expectation.

The strong law of large numbers on the $\log p_\phi\left(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i - 1}, \mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}\right)$'s for $\phi \in \Phi_1$ is now a direct consequence of the above integrability result and remark P1.

ii) For each $I \in \mathcal{H}_0^c$ and $\phi \neq \phi_0$, there exists at least one value $\tilde{n}_I \in \tilde{N}_0(\Omega)$ such that $P_0(\tilde{N} = \tilde{n}_I) > 0$ and $P_\phi(\tilde{N} = \tilde{n}_I) = 0$. When k goes to infinity, one event of the kind $\left\{\tilde{N}_k = \tilde{n}_I\right\}$ will be P_0-almost surely observed and then $\log P_\phi(\tilde{N} = \tilde{n}_I) = \log 0 = -\infty$. Notice now that $\mathrm{Card}(\mathcal{H}_0^c) < \infty$ and

$$\limsup \left\{ \sup_{\phi\in\Phi_2} \frac{1}{k} \ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) \neq -\infty \right\} \subseteq \bigcup_{I\in\mathcal{H}^c} \limsup \left\{ \tilde{N}_k \neq \tilde{n}_I \right\},$$

and we obtain

$$P_0\left(\limsup \left\{ \sup_{\phi\in\Phi_2} \frac{1}{k} \ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}) \neq -\infty \right\}\right) \leq \sum_{I\in\mathcal{H}^c} P_0\left(\limsup\left\{\tilde{N}_k \neq \tilde{n}_I\right\}\right) = 0,$$

which concludes the proof. □

Lemma 5 *Under conditions C1–3 and I1–2 the Kullback distance* $\mathcal{K}(\phi_0,\phi) \overset{def.}{=} \mathcal{E}_0(\phi_0) - \mathcal{E}_0(\phi)$ *satisfies the contrast property, i.e.*

$$\mathcal{K}(\phi_0,\phi) \geq 0, \quad \textit{for all } \phi \in \Phi, \quad \textit{and} \quad \mathcal{K}(\phi_0,\phi) = 0 \Longleftrightarrow \phi = \phi_0.$$

Proof. Define

$$\begin{aligned}
\mathcal{E}_0^{(1)}(\phi) &= E_0\left[\log p_\phi(N)\right], \\
\mathcal{E}_0^{(2)}(\phi) &= E_0\left[\log p_\phi(\mathbf{Y}_1^N|N)\right], \\
\mathcal{E}_0^{(3)}(\phi) &= E_0\left[\log p_\phi(\mathbf{Y}_1^{\tilde{N}},\tilde{N})\right],
\end{aligned}$$

and the corresponding Kullback distances:

$$\mathcal{K}^{(i)}(\phi_0,\phi) \overset{\text{def.}}{=} \mathcal{E}_0^{(i)}(\phi_0) - \mathcal{E}_0^{(i)}(\phi), \quad \text{for } i = 1,2,3,$$

then

$$\mathcal{K}(\phi_0,\phi) = \sum_{i=1}^{3} \mathcal{K}^{(i)}(\phi_0,\phi).$$

Now, by the Jensen inequality we have

$$\mathcal{K}^{(1)}(\phi_0,\phi) = E_0\left[\log\left(\frac{p_\phi(N)}{p_0(N)}\right)\right] \geq \log\left(E_0\left[\frac{p_\phi(N)}{p_0(N)}\right]\right) = 0,$$

for all $\phi \in \Phi$. Applying the conditional Jensen inequality, we obtain the same inequalities for $\mathcal{K}^{(2)}(\phi_0,\phi)$ and $\mathcal{K}^{(3)}(\phi_0,\phi)$. As a consequence we have

$$\mathcal{K}(\phi_0,\phi) = 0 \Longleftrightarrow \mathcal{K}^{(i)}(\phi_0,\phi) = 0 \ \text{ for } i = 1,2,3.$$

Direct calculations show that:

$$\mathcal{K}^{(1)}(\phi_0,\phi) = 0 \Longleftrightarrow \alpha_0(a,a) = \alpha(a,a).$$

Furthermore, we can check that $\mathcal{K}^{(2)}(\phi_0,\phi)$ is the product of $E_0(N)$ and the Kullback distance between $g(\cdot;\theta_a(\phi))$ and $g(\cdot;\theta_a(\phi_0))$, and then by assumption I1 we get

$$\mathcal{K}^{(2)}(\phi_0,\phi) = 0 \Longleftrightarrow \theta_a = \theta_a(\phi_0).$$

Finally, we have to show that $\mathcal{K}^{(3)}(\phi_0,\phi) = 0$, which allows us to identify the remaining components of α_0 and parameters $\theta_1(\phi_0),\ldots,\theta_{a-1}(\phi_0)$. Now, it is easy to check that

$$\mathcal{K}^{(3)}(\phi_0,\phi) = \sum_{\tilde{n}\in\tilde{N}_0(\Omega)} \mathcal{K}_{\tilde{n}}^{(3)}(\phi_0,\phi) P_0(\tilde{N} = \tilde{n}) + \mathcal{K}^{(4)}(\phi_0,\phi),$$

where $\mathcal{K}^{(4)}(\phi_0, \phi)$ is the Kullback distance between $P_\phi(\tilde{N})$ and $P_0(\tilde{N})$, and $\mathcal{K}^{(3)}_{\tilde{n}}(\phi_0, \phi)$ is the Kullback distance between $p_\phi(\mathbf{Y}_1^{\tilde{n}}|\tilde{N} = \tilde{n})$ and $p_0(\mathbf{Y}_1^{\tilde{n}}|\tilde{N} = \tilde{n})$. Now, if $\mathcal{K}^{(3)}(\phi_0, \phi) = 0$ we have $\mathcal{K}^{(4)}(\phi_0, \phi) = 0$ and $\mathcal{K}^{(3)}_{\tilde{n}}(\phi_0, \phi) = 0$ for all $\tilde{n} \in \tilde{N}_0(\Omega)$ since $P_0(\tilde{N} = \tilde{n}) > 0$. Obviously $\mathcal{K}^{(4)}(\phi_0, \phi) = 0$ implies $P_\phi(\tilde{N}) = P_0(\tilde{N})$. Moreover, for $\tilde{n} \geq 1$ such that $P_0(\tilde{N} = \tilde{n}) > 0$, we have $p_\phi(\mathbf{y}_1^{\tilde{n}}|\tilde{N} = \tilde{n}) = p_0(\mathbf{y}_1^{\tilde{n}}|\tilde{N} = \tilde{n})$ $\mu_{\tilde{n}} - a.e.$ on $\mathbb{R}^{\tilde{n}}$ ($\mu_{\tilde{n}}$ denotes the Lebesgue measure on $\mathbb{R}^{\tilde{n}}$). By I1, and constraints C1–2, it follows that we identify all the

$$\frac{\alpha(a, x_1)\alpha(x_1, x_2)\ldots\alpha(x_{\tilde{n}}, a)}{(1 - \alpha(a, a))P_\phi(\tilde{N} = \tilde{n})} \quad \text{and} \quad \theta_{x_i}(\phi), \quad i = 1, \ldots, \tilde{n},$$

for which $\alpha(a, x_1)\alpha(x_1, x_2)\ldots\alpha(x_{\tilde{n}}, a) > 0$. Because $\mathcal{K}^{(1)}(\phi_0, \phi) = 0$ and $\mathcal{K}^{(4)}(\phi_0, \phi) = 0$ identify the laws of N and $\tilde{N}$ we get the identifiability of

$$\alpha(a, x_1)\alpha(x_1, x_2)\ldots\alpha(x_{\tilde{n}}, a) \quad \text{and} \quad \theta_{x_i}(\phi), \quad i = 1, \ldots, \tilde{n},$$

for which $\alpha(a, x_1)\alpha(x_1, x_2)\ldots\alpha(x_{\tilde{n}}, a) > 0$. By irreducibility of α there exists an $\tilde{n}$ large enough with $P_0(\tilde{N} = \tilde{n}) > 0$ for which all the $\theta_i(\phi)$ ($i = 1, \ldots, a - 1$) are identified. Finally, the injectivity of

$$\alpha \mapsto \left\{ \prod_{i=0}^{n} \alpha(x_i, x_{i+1});\ x_0^{n+1} \in \{a\} \times E_a^n \times \{a\},\ n \geq 1 \right\}$$

given in I2 allows us to identify the matrix α. This completes the proof. □

Theorem 1 *Under assumptions C1–6, and I1–2, the maximum likelihood estimator $\hat{\phi}_k$ defined by:*

$$\hat{\phi}_k = \arg\max_{\phi \in \Phi} \ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1}), \tag{11}$$

converges P_0–almost surely toward ϕ_0, the true of value of the parameter.

Proof. The proof is based on the proof given by Dacunha-Castelle and Duflo (1993, p. 94–96). Let us consider $\hat{\phi}_k$ defined in (**??**) as the minimum contrast estimator

$$\hat{\phi}_k = \arg\min_{\phi \in \Phi} U_k(\phi) = \min(\arg\min_{\phi \in \Phi_1} U_k(\phi), \arg\min_{\phi \in \Phi_2} U_k(\phi)),$$

where $U_k(\phi) = -k^{-1}\ell_\phi(\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$. From Lemma 5 it is clear that $\hat{\phi}_k$ belongs asymptotically, P_0 almost surely to Φ_1. Let us consider a countable dense set D in Φ_1. To this way

$\inf_{\phi\in\Phi_1} U_k(\phi) = \inf_{\phi\in\Phi_1\cap D} U_k(\phi)$, is a $\mathcal{F}_k$-measurable random variable. We define in addition the random variable $W(k,\eta) = \sup\{|U_k(\phi) - U_k(\phi')|;\ (\phi,\phi') \in D^2,\ |\phi - \phi'| \le \eta\}$, and recall that $\mathcal{K}(\phi_0,\phi_0) = 0$. Consider a non empty open ball B_0 centered in ϕ_0 such that $K(\phi_0,\phi)$ is bounded from below by a positive real number 2ε on $\Phi_1\backslash B_0$. Consider a sequence (η_r) decreasing to zero, and for a given $r \ge 1$, a covering of $\Phi_1\backslash B_0$ by a finite number ℓ of balls $(B_i)_{1\le i\le \ell}$ of radius less than η_r. For all $\phi \in B_i$, then

$$\begin{aligned} U_k(\phi) &\ge U_k(\phi_i) - |U_k(\phi) - U_k(\phi_i)| \\ &\ge U_k(\phi_i) - \sup_{\phi\in B_i} |U_k(\phi) - U_k(\phi_i)|, \end{aligned}$$

which leads to

$$\inf_{\phi\in\Phi_1\backslash B_0} U_k(\phi) \ge \inf_{1\le i\le \ell} U_k(\phi_i) - W(k,\eta_r).$$

As a consequence we have the following event inclusions

$$\begin{aligned} \left\{\hat\phi_k \notin B_0\right\} &\subseteq \left\{\inf_{\phi\in\Phi_1\backslash B_0} U_k(\phi) < \inf_{\phi\in B_0} U_k(\phi)\right\} \\ &\subseteq \left\{\inf_{\phi\in\Phi_1\backslash B_0} U_k(\phi) < U_k(\phi_0)\right\} \\ &\subseteq \left\{\inf_{1\le i\le \ell} U_k(\phi_i) - W(k,\eta_r) < U_k(\phi_0)\right\} \\ &\subseteq \{W(k,\eta_r) > \varepsilon\} \cup \left\{\inf_{1\le i\le \ell}(U_k(\phi_i) - U_k(\phi_0)) \le \varepsilon\right\}. \end{aligned}$$

Thus we have

$$\limsup_k \left\{\hat\phi_k \notin B_0\right\} \subseteq \limsup_k \{W(k,\eta_r) > \varepsilon\} \cup \limsup_k \left\{\inf_{1\le i\le \ell}(U_k(\phi_i) - U_k(\phi_0)) \le \varepsilon\right\}. \quad (12)$$

By the strong law of large number established in Lemma **??** we have

$$P_0\left(\limsup_k \left\{\inf_{1\le i\le \ell}(U_k(\phi_i) - U_k(\phi_0)) \le \varepsilon\right\}\right) = 0. \quad (13)$$

In addition according to assumptions C**??**–**??** (see also Lemma **??**, and calculations from (**??**) to (**??**)), there exists a random variable $h(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde N})$ such that

$$\sup_{\phi\in\Phi_1} |\log p_\phi\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde N}\right)| \le h(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde N}),$$

with $E_0[h(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}})] < \infty$, where

$$h\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right) = (N+\tilde{N})|\log\delta| + \sum_{j=1}^{N} |\log g_2(Y_j)| - \log(1-\delta) + \sum_{i=1,2}\sum_{j=N+1}^{N+\tilde{N}} |\log g_i(Y_j)|,$$

does not depend on ϕ. Let us consider the following random variable

$$V_\eta\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right) = \sup_{(\phi,\phi')\in\Phi_1^2} \left\{ |\log p_\phi\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right) - \log p_{\phi'}\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right)|;\ |\phi-\phi'| \le \eta \right\}.$$

Using the previous uniform upper bound and continuity assumption C**??**, we have

$$V_\eta\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right) \le 2h\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right), \text{ and } \lim_{\eta\to 0} E_0\left[V_\eta\left(\mathbf{Y}_1^N, \mathbf{Y}_{N+1}^{N+\tilde{N}}\right)\right] = 0.$$

Hence we have P_0–almost surely $W(k,\eta) \le k^{-1}\sum_{j=1}^k V_{\eta_{r'}}(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}, \mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1})$, and for r' large enough we have $E_0(V_{\eta_{r'}}(\mathbf{Y}_{\tau_1}^{\tilde{\tau}_1-1}, \mathbf{Y}_{\tilde{\tau}_1}^{\tau_2-1})) \le \varepsilon$, therefore

$$\limsup_k \{W(k,\eta_{r'}) > \varepsilon\} \subseteq \limsup_k \left\{ k^{-1}\sum_{j=1}^k V_{\eta_{r'}}(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}, \mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1}) > \varepsilon \right\},$$

and $P_0\left(\limsup_k \left\{ k^{-1}\sum_{j=1}^k V_{\eta_{r'}}(\mathbf{Y}_{\tau_j}^{\tilde{\tau}_j-1}, \mathbf{Y}_{\tilde{\tau}_j}^{\tau_{j+1}-1}) > \varepsilon \right\}\right) = 0$ which leads to

$$P_0(\limsup_k \{W(k,\eta_{r'}) > \varepsilon\}) = 0. \tag{14}$$

By (**??**)–(**??**), we prove the strong consistency of the maximum likelihood estimator $\hat{\phi}_k$. □

5 Asymptotic normality

Write $V_i(\phi) = \log p_\phi\left(\mathbf{Y}_{\tau_i}^{\tilde{\tau}_i-1}, \mathbf{Y}_{\tilde{\tau}_i}^{\tau_{i+1}-1}\right)$ for $i = 1,\ldots,k$. From property P**??** the random variables $V_i(\phi)$'s are independent and identically distributed and have the same distribution as

$$V_1(\phi) = \log\left(\sum_{x_0^{N+\tilde{N}} \in E_a\times\{a\}^N\times E_a^{\tilde{N}}} \prod_{j=1}^{N+\tilde{N}} W_j(\phi) \right),$$

where $W_j(\phi) = \alpha(x_{j-1},x_j)(\phi) g(Y_j;\theta_{x_j}(\phi))$ and $x_0 \stackrel{\text{def.}}{=} x_{N+\tilde{N}}$.
For any function v depending on ϕ, we note

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

Let us recall that

$$k^{-1/2}\ \dot{\ell}_k\ (\phi) = k^{-1/2} \sum_{j=1}^{k} \dot{V}_j\ (\phi), \tag{15}$$

Lemma 6 *Under assumptions C??–C??, we have* $k^{-1/2}\ \dot{\ell}_k\ (\phi_0) \xrightarrow[k\to\infty]{\mathcal{L}} \mathcal{N}(0, \mathcal{I}_0)$.

Proof. ¿From (??), to prove the desired central limit theorem under P_0, it suffices to show that the independent random variables $\dot{V}_j\ (\phi_0)$ are centered and belong to $L^2(P_0)$, or equivalently, that is true for $\dot{V}_1\ (\phi_0)$. We have

$$\dot{V}_1\ (\phi_0) = \left[\sum_{x_0^{N+\tilde{N}} \in E_a \times \{a\}^N \times E_a^{\tilde{N}}} \sum_{k=1}^{N+\tilde{N}} \dot{W}_k\ (\phi_0) \prod_{j=1\ ;\ j\neq k}^{N+\tilde{N}} W_j(\phi_0) \right] \exp(-V_1(\phi_0)),$$

and we notice that for all $(x_{j-1}, x_j) \in E^2$ such that $W_j(\phi_0) \neq 0$ (which is true P_0 almost surely), $\dot{W}_j\ (\phi_0)$ satisfies

$$\begin{aligned} \dot{W}_j\ (\phi_0) &= \frac{\dot{W}_j\ (\phi_0)}{W_j(\phi_0)} W_j(\phi_0) \\ &= \left[\frac{\dot{\alpha}\ (x_{j-1}, x_j)(\phi_0)}{\alpha(x_{j-1}, x_j)(\phi_0)} + \frac{\dot{g}\ (Y_j; \theta_{x_j}(\phi_0))}{g(Y_j; \theta_{x_j}(\phi_0))} \right] W_j(\phi_0) \\ &\leq \left[C + G^{(1)}(Y_j) \right] W_j(\phi_0), \end{aligned}$$

where the inequality holds componentwise, C is a q-dimensional constant arising from C?? and C??, whereas from C??, $G^{(1)}$ is a q-dimensional function the components of which are equal to $g^{(1)}$.

Then

$$\begin{aligned} &\frac{1}{2}\ \dot{V}_1\ (\phi_0)\ \dot{V}_1^T\ (\phi_0) \\ \leq\ & \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \left(C + G^{(1)}(Y_i)\right) \left(C + G^{(1)}(Y_j)\right)^T + \sum_{i=N+1}^{N+\tilde{N}} \sum_{j=N+1}^{N+\tilde{N}} \left(C + G^{(1)}(Y_i)\right) \left(C + G^{(1)}(Y_j)\right)^T \right]. \end{aligned}$$

It remains to show that the right hand side of the above inequality has a finite expectation under P_0. Since all the components of C and $G^{(1)}$ are equal it is sufficient to prove the result for one component of the right hand side of the above inequality, noting then $C = c$ and

$G^{(1)} = g^{(1)}$. This component is therefore equal to

$$c(N^2 + \tilde{N}^2) + 2c\left(N\sum_{j=1}^{N} g^{(1)}(Y_j) + \tilde{N}\sum_{j=N+1}^{N+\tilde{N}} g^{(1)}(Y_j)\right)$$
$$+\sum_{i=1}^{N}\sum_{j=1}^{N} g^{(1)}(Y_i)g^{(1)}(Y_j) + \sum_{i=N+1}^{N+\tilde{N}}\sum_{j=N+1}^{N+\tilde{N}} g^{(1)}(Y_i)g^{(1)}(Y_j).$$

Finiteness of $E_0[\tilde{N}^2]$ holds by Lemma **??** and $E_0[N^2] < +\infty$ since N is geometrically distributed with parameter $1 - \alpha_0(a, a) \in (0, 1)$.

We have by Lemma **??** and **??** and assumption C**??** that:

$$\begin{aligned} E_0\left[\tilde{N}\sum_{l=N+1}^{N+\tilde{N}} g^{(1)}(Y_l)\middle| N = n\right] &= \sum_{\tilde{n}=1}^{+\infty}\int_F \tilde{n}\sum_{l=n+1}^{n+\tilde{n}} g^{(1)}(y_l)p_0(y_l, \tilde{N} = \tilde{n}|N = n)dy_l \\ &\leq \sum_{\tilde{n}=1}^{+\infty}\int_F \tilde{n}\sum_{l=n+1}^{n+\tilde{n}} g^{(1)}(y_l)g_2(y_l)dy_l P(\tilde{N} = \tilde{n}) \\ &\leq \int_F g^{(1)}(y)g_2(y)dy \times \sum_{\tilde{n}=1}^{+\infty}\tilde{n}^2 P(\tilde{N} = \tilde{n}) \\ &\leq C_1 E_0[\tilde{N}^2], \end{aligned}$$

where C_1 is a finite constant, and then the unconditional expectation $E_0\left[\tilde{N}\sum_{l=N+1}^{N+\tilde{N}} g^{(1)}(Y_l)\right]$ is finite. Following the above lines we prove also that $E_0\left[N\sum_{l=1}^{N} g^{(1)}(Y_l)\right]$ is finite.

Again, by Lemma **??** and **??** and assumption C**??** we have by similar calculations:

$$E_0\left[\sum_{j=N+1}^{N+\tilde{N}}\sum_{l=N+1}^{N+\tilde{N}} g^{(1)}(Y_j)g^{(1)}(Y_l)\middle| N = n\right] \leq C_1^2 E_0[\tilde{N}^2] + C_2 E_0[\tilde{N}],$$

where C_2 is a finite constant, and then the unconditional expectation is finite. Following the above lines we also prove that $E_0\left[\sum_{j=1}^{N}\sum_{l=1}^{N} g^{(1)}(Y_j)g^{(1)}(Y_l)\right]$ is finite, which leads to the desired finiteness of $E_0\left[\dot{V}_1(\phi_0)\,\dot{V}_1^T(\phi_0)\right]$. The fact that $E_0\left[\dot{V}_1(\phi_0)\right] = 0$ is obvious. □

Lemma 7 *Under assumptions C**??**–C**??** and C**??**–C**??**, and let ϕ_k^* be any possibly P_0–strongly consistent estimator sequence of ϕ_0, then $k^{-1}\,\ddot{\ell}_k(\phi_k^*) \underset{k\to\infty}{\longrightarrow} -\mathcal{I}_0$ P_0–probability.*

Proof. First we show by the strong law of large numbers that $(k^{-1}\,\ddot{\ell}_k(\phi_0))_{k\geq 1}$ converges P_0–almost surely to $E_0\left[\ddot{V}_1(\phi_0)\right]$. For this purpose, it is enough to prove that $\ddot{V}_1(\phi_0)$ belongs to $L^1(P_0)$. The proof is omitted since it follows the lines of proof of Lemma **??**.

Then we have to show that $(k^{-1}\ddot{\ell}_k(\phi_k^*))_{k\geq 1}$ and $(k^{-1}\ddot{\ell}_k(\phi_0))_{k\geq 1}$ are asymptotically equivalent in P_0–probability. For all $\eta > 0$ and for all $0 < \xi < \varepsilon$, we have

$$\begin{aligned} P_0\left(\left|\frac{1}{k}\ddot{\ell}_k(\phi_k^*) - \frac{1}{k}\ddot{\ell}_k(\phi_0)\right| > \eta\right) &\leq P_0\left(\frac{1}{k}\sum_{j=1}^{k}\sup_{\phi\in\bar{B}(\phi_0,\xi)}\left|\ddot{V}_j(\phi) - \ddot{V}_j(\phi_0)\right| > \eta\right) \\ &\quad + P_0(\phi_k^* \notin \bar{B}(\phi_0,\xi)). \end{aligned}$$

The second term of the right hand side goes to zero as k goes to infinity by strong consistency of ϕ_k^*. For the first term of the right hand side we notice

$$\varrho(\xi; Y_1^{N+\tilde{N}}) = \sup_{\phi\in B(\phi_0,\xi)}\left|\ddot{V}_j(\phi) - \ddot{V}_j(\phi_0)\right| \underset{\xi\to 0}{\longrightarrow} 0 \quad a.e.$$

In addition there exists a P_0–integrable function h, such that $\ddot{V}_j(\phi) \leq h(Y_1^{N+\tilde{N}})$ on $B(\phi_0,\varepsilon)$, which implies $\varrho(\xi; Y_1^{N+\tilde{N}}) \leq 2h(Y_1^{N+\tilde{N}})$. Now, using the Lebesgue's continuity theorem, it follows

$$E_0\left[\varrho(\xi; Y_1^{N+\tilde{N}})\right] \underset{\xi\to 0}{\longrightarrow} 0. \tag{16}$$

Finally using Tchebychev inequality we have

$$\begin{aligned} P_0\left(\frac{1}{k}\sum_{j=1}^{k}\varrho(\xi; Y_{\tau_j}^{\tau_{j+1}-1}) \geq \varepsilon\right) &\leq \frac{1}{k[\varepsilon - E_0[\varrho(\xi; Y_1^{N+\tilde{N}}))]}\sum_{j=1}^{k}E_0\left[\varrho(\xi; Y_{\tau_j}^{\tau_{j+1}-1})\right] \\ &= \frac{1}{\varepsilon - E_0\left[\varrho(\xi; Y_1^{N+\tilde{N}})\right]}E_0\left[\varrho(\xi; Y_1^{N+\tilde{N}})\right], \end{aligned}$$

which goes to zero, from (**??**), as ξ goes to 0.
Finally, it remains to show that

$$E_0\left[\ddot{V}_1(\phi_0)\right] = -E_0\left[\dot{V}_1(\phi_0)\,\dot{V}_1^T(\phi_0)\right] = -\mathcal{I}_0,$$

which follows from the fact that for $\phi = \phi_0$:

$$E_0\left[\frac{1}{p_\phi(\mathbf{Y}_1^N, \mathbf{Y}_{1+N}^{N+\tilde{N}})}\frac{\partial^2 p_\phi}{\partial\phi\partial\phi^T}(\mathbf{Y}_1^N, \mathbf{Y}_{1+N}^{N+\tilde{N}})\right] = 0.$$

□

Theorem 2 *Under assumptions C**??**–C**??**, and assuming that $\mathcal{I}_0$ is nonsingular, we get*

$$k^{1/2}(\hat{\phi}_k - \phi_0) \underset{k\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \mathcal{I}_0^{-1}).$$

For this purpose we notice that for all $0 < \xi < \varepsilon$, and without loss of generality we assume that PROOF. For k large enough $\hat{\phi}_k$ is an interior point of Φ, and $\|\hat{\phi}_k - \phi_0\| < \kappa$, and then by a Taylor expansion of $\dot{\ell}_\phi\ (\mathbf{Z}_{\tau_1}^{\tau_{k+1}-1})$ about ϕ_0 we get,

$$k^{1/2}(\hat{\phi}_k - \phi_0) = [-k^{-1}\ \ddot{\ell}_k\ (\phi_k^*)]^{-1} k^{1/2}\ \dot{\ell}_k\ (\phi_0),$$

where ϕ_k^* is a point of the line segment between ϕ_0 and $\hat{\phi}_k$. Therefore using Theorem **??**, Lemma **??** and **??** we obtain the asymptotic normality of the MLE. □

6 Concluding remarks

In this paper we have introduced a new missing data model based on HMM type observations. The main difference between PHMMs and HMMs is that in the first one partial information on the latent Markov chain is given. This partial information is reduced here into the fact that the latent Markov chain is visible when it reaches a specified state. This framework allows us to deal with i.i.d. pieces of trajectories and then to establish strong consistency and asymptotic normality of the MLE. We point out now that a natural extension of this work would be the study of the same kind of models when the latent Markov chain is observed in a subset (not reduced to one state) of its state space. In that case the pieces of trajectories described previously are no longer i.i.d. making the study of the MLE much more tricky. In addition, asymptotic results are obtained under weaker assumptions than those found in standard HMM literature, especially the PHMMs can include periodic underlying Markov chains. We show on particular models how to prove identifiability using some basic linear algebra arguments. For the numerical computation of the MLE two ways are possible. The first one could be based on standard (stochastic or deterministic) likelihood maximization techniques, using recursive formula (see Rabiner, 1989). The second one could be based on an adaptation of the EM (Expectation Maximization) algorithm. Finally the PHMM are alternative models for both reliability analysis of degradation data involving explanatory variables and specific longitudinal survival analysis models.

References

Bagdonavičius, V., and Nikulin M. (2001). Estimation in Degradation Models with Explanatory Variables, *Lifetime Data Analysis,* **7**, 85–103.

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.

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

Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41**, 164–171.

Bickel, P.J., and Ritov, Y. (1996). Inference in Hidden Markov models I: LAN in the stationary case. *Bernoulli*, **2**, 199–228.

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.

Brémaud, P. (1998). *Markov chains, Gibbs Fields, Monte Carlo simulation, and queues.* Springer.

Chan, K.S., and Ledolter, J. (1995). Monte Carlo EM estimation in time series models involving counts. *J. Am. Statist. Assoc.* **90**, No. 429, 242–252.

Chen, Y., and Singpurwalla, N. (1997). Unification of software reliability models by self-exciting point processes, *Advances in Applied Probability*, **29**, 337–352.

Choi, H., and Baraniuk, R. (2001). Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Trans. Image Process.* **10**, No. 9, 1309–1321.

Churchill, G.A. (1989). Stochastic models for Heterogeneous DNA sequences. *Bull. Math.*

Biol. **51**, 79–94.

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

DeJong, P., and Shepard, N. (1995). The simulation smoother for time series models. *Biometrika*, **82**, 339–350.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via EM algorithm. Discussion. *J. Royal Statist. Soc., Ser. B*, **39**, 1–38.

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

Durand, J-B., and Gaudouin, O. (2003). Software reliability modelling and prediction with hidden Markov chains, *Rapport de recherche INRIA*, **4747**.

Cappé, O., Douc, R., Moulines, E., and Robert, C. (2002). On the convergence of the Monte Carlo maximum likelihood method for latent variable models. *Scand. J. Satistit.* **29**, 615–635.

Durbin, J., and Koopman, S.J. (1997). Monte Carlo maximum likelihood estimation for non Gaussian state space models. *Biometrika,* **84**, 669–684.

Forchhammer, S., and Rissanen, J. (1996). Partially Hidden Markov Models. *IEEE, Trans. Inform. Theory* **42**, 1253–1256.

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

Guihenneuc-Jouyaux, C., Richardson, S., and Longini, I. M. (2000). Modelling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. *Biometrics*, **56**, 733–741.

Jackson, C. H., and Sharples, L.D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine*, **21**, No. 1, 113–128.

Khale, W., and Wendt, H. (2000). Statistical analysis of damage processes, In: *Recent Advances in Reliability Theory*, Birkäuser, Boston (Ed. Limnios, N. and Nikulin, M.), 199–212.

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

Kim, S., Shepard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Econometrics studies*, **65**, 361–364.

Leroux, B.G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127–143.

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

Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications.* NSF-CBMS Regional Conference Series in Probability and statistics.

MacDonald, I.L., and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman & Hall, London.

Neuts, M.F., (1994). Matrix-geometric solutions in stochastic models: an algorithmic approach. *Dover publications, New York.*

Prakasa Rao, B.L.S (1992). *Identifiability in Stochastic Models, Characterization of Probability Distributions.* Academic, Boston.

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.

Rydén, T. (1997). On recursive estimation for hidden Markov models. *Stochastic. Proc. Appl.* **66**, 79–96.

Singpurwalla, N. (1995). Survival in dynamic environments. *Statistical Science*, **10**, 86–103.

Teicher, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Satist.* **38**, 1300–1302.

A Identifiability of models M1, M2, and M3

Identifiability of M1. To prove injectivity in I2, we just need to consider trajectories of length 1 and 2, from a to a. In fact we are going to prove that the trajectories probabilities of length 1 and 2 (from a to a) induced by two 3×3 stochastic matrices α and α_0, coincide if and only if $\alpha = \alpha_0$ (up to permutation of the states).

Since $P_0(\tilde{N} = 1) > 0$ and $P_0(\tilde{N} = 2) > 0$ we then obtain the system

$$(E) \quad \begin{cases} \sum_{j=1}^{a} \alpha(i,j) = 1 & \text{for} \quad i \in E \quad (1) \\ \alpha(a,i)\alpha(i,a) = \alpha_0(a,i)\alpha_0(i,a) & \text{for} \quad i \in E_a \quad (2) \\ \alpha(a,i)\alpha(i,j)\alpha(j,a) = \alpha_0(a,i)\alpha_0(i,j)\alpha_0(j,a) & \text{for} \quad (i,j) \in E_a^2 \quad (3) \end{cases}$$

We notice that above equations (2) and (3), for $i = j$, allows to identify the parameters $\alpha(i,i)$ for $i \in E_a$. Thus the first diagonal of α is identified. It remains now to identify the parameters $\alpha(i,j)$ for $i \neq j$. As a consequence we switch from system (E) into the equivalent following system (E1) :

$$(E1) \quad \begin{cases} \sum_{j \in E \setminus \{i\}} \alpha(i,j) = \sum_{j \in E \setminus \{i\}} \alpha_0(i,j) & \text{for} \quad i \in E \quad (1) \\ \alpha(a,i)\alpha(i,a) = \alpha_0(a,i)\alpha_0(i,a) & \text{for} \quad i \in E_a \quad (2) \\ \alpha(a,i)\alpha(i,j)\alpha(j,a) = \alpha_0(a,i)\alpha_0(i,j)\alpha_0(j,a) & \text{for} \quad i \neq j \in E_a \quad (3) \end{cases}$$

For $a = 3$, we write α as :

$$\alpha = \begin{pmatrix} \alpha(1,1) & x_1 & x_2 \\ x_3 & \alpha(2,2) & x_4 \\ x_5 & x_6 & \alpha(3,3) \end{pmatrix}$$

Let us denote by x_i^0 the corresponding values for the matrix $\alpha(\phi_0)$, and execute the change of variable $y_i = \log(x_i)$ and $y_i^0 = \log(x_i^0)$ for $i = 1, \ldots, 6$. The system (E1) is now equivalent to :

$$\begin{cases} \left.\begin{array}{lcl} x_1 + x_2 & = & x_1^0 + x_2^0 \\ x_3 + x_4 & = & x_3^0 + x_4^0 \\ x_5 + x_6 & = & x_5^0 + x_6^0 \end{array}\right\} (E2) \\ \left.\begin{array}{lcl} x_5 x_2 & = & x_5^0 x_2^0 \\ x_6 x_4 & = & x_6^0 x_4^0 \\ x_5 x_1 x_4 & = & x_5^0 x_1^0 x_4^0 \\ x_6 x_3 x_2 & = & x_6^0 x_3^0 x_2^0 \end{array}\right\} (E3) \end{cases}$$

Taking (E3) through the logarithm function we show that the system can be written $BY = BY^0$ where:

$$B = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix},$$

with $Y = (y_1, y_2, y_3, y_4, y_5, y_6)^T$ and $Y^0 = (y_1^0, y_2^0, y_3^0, y_4^0, y_5^0, y_6^0)^T$.

We show that Ker$B$ is generated by the vectors Y_1 et Y_2 where:

$$Y_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad Y_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

We deduce that the solutions of $BY = BY^0$ take the form:

$$Y = Y^0 + \beta_1 Y_1 + \beta_2 Y_2.$$

Consider $\eta_i = \exp(\beta_i)$ for $i = 1, 2$, and take the previous equality componentwise through the exponential function:

$$\begin{cases} x_1 &= x_1^0 \eta_2/\eta_1 \\ x_2 &= x_2^0/\eta_1 \\ x_3 &= x_3^0 \eta_1/\eta_2 \\ x_4 &= x_4^0/\eta_2 \\ x_5 &= x_5^0 \eta_1 \\ x_6 &= x_6^0 \eta_2 \end{cases}$$

Using $(E2)$ the previous system becomes:

$$\begin{cases} x_1^0 \eta_2/\eta_1 + x_2^0/\eta_1 &= x_1^0 + x_2^0 \\ x_3^0 \eta_1/\eta_2 + x_4^0/\eta_2 &= x_3^0 + x_4^0 \\ x_5^0 \eta_1 + x_6^0 \eta_2 &= x_5^0 + x_6^0 \end{cases}$$

Multiplying the first equation of the system by η_1, and the second one by η_2, we obtain the equivalent system:

$$\begin{cases} -\eta_1(x_1^0 + x_2^0) + \eta_2 x_1^0 &= -x_2^0 \\ \eta_1 x_3^0 - \eta_2(x_3^0 + x_4^0) &= -x_4^0 \\ \eta_1 x_5^0 + \eta_2 x_6^0 &= x_5^0 + x_6^0 \end{cases}$$

The previous system admits the solution $(\eta_1, \eta_2) = (1, 1)$, and this solution is unique since the determinant associated to the two first equations is:

$$\begin{vmatrix} -(x_1^0 + x_2^0) & x_1^0 \\ x_3^0 & -(x_3^0 + x_4^0) \end{vmatrix} = x_1^0 x_4^0 + x_2^0 x_3^0 + x_2^0 x_4^0 > 0.$$

Which conclude the identifiability of M1.

Identifiability of M2. We have

$$\alpha = \begin{pmatrix} 0 & \alpha(1,2) & 0 & \alpha(1,4) \\ \alpha(2,1) & 0 & \alpha(2,3) & 0 \\ 0 & \alpha(3,2) & 0 & \alpha(3,4) \\ \alpha(4,1) & 0 & \alpha(4,3) & 0 \end{pmatrix} = \begin{pmatrix} 0 & x_1 & 0 & x_2 \\ x_3 & 0 & x_4 & 0 \\ 0 & x_5 & 0 & x_6 \\ x_7 & 0 & x_8 & 0 \end{pmatrix},$$

By $\mathcal{K}_{\tilde{n}}^{(3)}(\phi, \phi_0) = 0$ for $\tilde{n} = 2, 4$ we obtain:

$$\begin{cases} \alpha(4,1)\alpha(1,4) & = & \alpha_0(4,1)\alpha_0(1,4) \\ \alpha(4,3)\alpha(3,4) & = & \alpha_0(4,3)\alpha_0(3,4) \\ \alpha(4,3)\alpha(3,2)\alpha(2,3)\alpha(3,4) & = & \alpha_0(4,3)\alpha_0(3,2)\alpha_0(2,3)\alpha_0(3,4) \\ \alpha(4,1)\alpha(1,2)\alpha(2,1)\alpha(1,4) & = & \alpha_0(4,1)\alpha_0(1,2)\alpha_0(2,1)\alpha_0(1,4) \\ \alpha(4,3)\alpha(3,2)\alpha(2,1)\alpha(1,4) & = & \alpha_0(4,3)\alpha_0(3,2)\alpha_0(2,1)\alpha_0(1,4) \\ \alpha(4,1)\alpha(1,2)\alpha(2,3)\alpha(3,4) & = & \alpha_0(4,1)\alpha_0(1,2)\alpha_0(2,3)\alpha_0(3,4) \end{cases}$$

or equivalently

$$\begin{cases} \alpha(4,1)\alpha(1,4) & = & \alpha_0(4,1)\alpha_0(1,4) \\ \alpha(4,3)\alpha(3,4) & = & \alpha_0(4,3)\alpha_0(3,4) \\ \alpha(3,2)\alpha(2,3) & = & \alpha_0(3,2)\alpha_0(2,3) \\ \alpha(1,2)\alpha(2,1) & = & \alpha_0(1,2)\alpha_0(2,1) \\ \alpha(4,3)\alpha(3,2)\alpha(2,1)\alpha(1,4) & = & \alpha_0(4,3)\alpha_0(3,2)\alpha_0(2,1)\alpha_0(1,4) \\ \alpha(4,1)\alpha(1,2)\alpha(2,3)\alpha(3,4) & = & \alpha_0(4,1)\alpha_0(1,2)\alpha_0(2,3)\alpha_0(3,4) \end{cases}$$

Then, using the same notations as for model M1, the above system leads to the following one

$$\begin{cases} x_2 x_7 & = & x_2^0 x_7^0 \\ x_6 x_8 & = & x_6^0 x_8^0 \\ x_1 x_3 & = & x_1^0 x_3^0 \\ x_4 x_5 & = & x_4^0 x_5^0 \\ x_2 x_3 x_5 x_8 & = & x_2^0 x_3^0 x_5^0 x_8^0 \\ x_1 x_4 x_6 x_7 & = & x_1^0 x_4^0 x_6^0 x_7^0 \end{cases}$$

or equivalently, denoting $y_i = \log(x_i)$ and $y_i^0 = \log(x_i^0)$ $(i = 1, \ldots, 6)$, to the system $BY = BY^0$ where $Y = (y_1, \ldots, y_6)^T$, $Y^0 = (y_1^0, \ldots, y_6^0)^T$, and

$$B = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Then the general solution Y of $BY = BY^0$ is given by

$$Y = Y^0 + \beta_1 \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ -1 \end{pmatrix} \quad \text{and then} \quad X = \begin{pmatrix} x_1^0\eta_1 \\ x_2^0\eta_2 \\ x_3^0/\eta_1 \\ x_4^0/\eta_1 \\ x_5^0\eta_1 \\ x_6^0\eta_2 \\ x_7^0/\eta_2 \\ x_8^0/\eta_2 \end{pmatrix}.$$

where $\eta_i = \exp(\beta_i)$ $(i = 1, 2)$. Now, by using the stochasticity of α we get

$$\begin{cases} x_1^0\eta_1 + x_2^0\eta_2 & = & 1 \\ x_3^0/\eta_1 + x_4^0/\eta_1 & = & 1 \\ x_5^0\eta_1 + x_6^0\eta_2 & = & 1 \\ x_7^0/\eta_2 + x_8^0/\eta_2 & = & 1 \end{cases}$$

therefore $\eta_1 = \eta_2 = 1$ and Assumption I2 is satisfied for M2.

Identifiability of M3. It is easy to check that $P_\phi(\tilde{n}) = 0$ for all $\tilde{n} < a - 1$.

Since $\mathcal{K}^{(1)}(\phi, \phi_0) = 0$ identifies $\alpha(a, a)$, by stochasticity of α we identify $\alpha(a, 1)$. Now, since $P_\phi(\tilde{N} = a - 1) > 0$ we have by $\mathcal{K}_{a-1}^{(3)}(\phi, \phi_0) = 0$

$$\alpha(1, 2)\alpha(2, 3) \ldots \alpha(a - 1, a) = \alpha_0(1, 2)\alpha_0(2, 3) \ldots \alpha_0(a - 1, a),$$

and $P_\phi(\tilde{N} = a) > 0$ gives

$$\alpha(i, i) \prod_{i=2}^{a} \alpha(i - 1, i) = \alpha_0(i, i) \prod_{i=2}^{a} \alpha_0(i - 1, i), \quad \text{for } i = 1, \ldots, a - 1.$$

Then we identify $\alpha(i, i)$ for $i = 1, \ldots, a-1$ and by stochasticity of α we identify all the terms of α.