

Asymptotics of the maximum likelihood estimator for general hidden Markov models

RANDAL DOUC¹ and CATHERINE MATIAS²

¹*Dpt. TSI/CNRS URA 820, École Nationale Supérieure des Télécommunications, Paris, France. E-mail: douc@tsi.enst.fr*

²*CNRS UMR C 8628. Laboratoire de Probabilités, Statistique et Modélisation. Bât. 425. Université Paris Sud, 91405 Orsay Cedex, France. E-mail: Catherine.Matias@math.u-psud.fr*

In this paper, we consider the consistency and asymptotic normality of the maximum likelihood estimator for a possibly non-stationary hidden Markov model where the hidden state space is a separable and compact space not necessarily finite, and both the transition kernel of the hidden chain and the conditional distribution of the observations depend on a parameter θ . For identifiable models, consistency and asymptotic normality of the maximum likelihood estimator are shown to follow from exponential memorylessness properties of the state prediction filter and geometric ergodicity of suitably extended Markov chains.

Keywords: asymptotic normality; consistency; geometric ergodicity; hidden Markov models; identifiability; maximum likelihood estimation

1. Introduction

Hidden Markov models (HMMs) form a wide class of discrete-time stochastic processes, used in different areas such as speech recognition (Juang and Rabiner 1991), neurophysiology (Fredkin and Rice 1987), biology (Churchill 1989), econometrics (Chib *et al.* 1998) and time-series analysis (De Jong and Shephard 1995; Chan and Ledolter 1995; see also MacDonald and Zucchini 1997 and the references therein).

Most work on maximum likelihood estimation in such models has focused on iterative numerical methods, suitable for approximating the maximum likelihood estimator. For finite hidden state-space models, the pioneering contribution is due to Baum *et al.* (1970) who presented an early and non-trivial application of the expectation–maximization principle (Dempster *et al.* 1977), known as the ‘forward–backward’ procedure. The more challenging issue of hidden Markov models with continuous state space was much studied throughout the 1990s, mostly using simulation-based approaches made possible by recent advances in Markov chain Monte Carlo methods (Chib *et al.* 1998; Durbin and Koopman 1997; De Jong and Shephard 1995; Chan and Ledolter 1995).

By contrast, the statistical issues regarding the asymptotic properties of the maximum likelihood estimator for hidden Markov models have until recently been largely ignored.

Baum and Petrie (1966) showed the consistency and asymptotic normality of the maximum likelihood estimator in the particular case where both the observed and the latent variables take only finitely many values. These results were recently extended in a series of papers by Leroux (1992), Bickel and Ritov (1996), Bickel *et al.* (1998) and Bakry *et al.* (1997). Bickel *et al.* generalized the method followed by Baum and Petrie (1966) to the case where the hidden Markov process X_n takes a finite number of values, the observations belonging to a general topological space. Around the same time, LeGland and Mevel (2000a; 2000b) independently developed a different technique to prove the consistency and asymptotic normality of the maximum likelihood estimator (Mevel 1997) for hidden Markov models with finite hidden state space. The work of LeGland and Mevel (2000a; 2000b) is based on the remark that the likelihood can be expressed as an additive function of an extended Markov chain. The key to the proof of LeGland and Mevel (2000b) consists in showing that, under appropriate conditions, this extended chain is geometrically ergodic, in a sense given below. The proof applies even when the hidden Markov chain is non-stationary, that is, the initial distribution of the chain does not correspond to its invariant distribution, while Baum and Petrie (1966), Leroux (1992) and Bickel *et al.* (1998) assume stationarity (see, however, Bakry *et al.* 1997, for an extension to the non-stationary case, based on a coupling technique).

In this paper, we study maximum likelihood estimation when the hidden Markov chain takes values in a topological space, assumed to be compact. The main result is the consistency of the maximum likelihood estimator, under the usual condition of identifiability of the parameters. We also identify the limit of the likelihood process as the Kullback–Leibler divergence. In addition, we prove the asymptotic normality of the maximum likelihood estimator in both the stationary and the non-stationary case. To the best of our knowledge, the only contribution dealing with the maximum likelihood for hidden Markov models taking values in a general topological space (for the definition of Markov chains in such spaces, see Meyn and Tweedie 1993) is Jensen and Petersen (1999). The authors show the asymptotic normality (but not the consistency) of the maximum likelihood estimator when the hidden Markov chain is stationary, the technique being based on an extension of Bickel *et al.* (1998). The technique used in our contribution to prove the consistency and the asymptotic normality of the maximum likelihood estimator for general hidden Markov models is completely different and is essentially based on the technique developed by LeGland and Mevel (2000a; 2000b) in a finite state space. A distinctive advantage of this technique is that it makes it possible to study in a common framework the stationary and the non-stationary case. The main difficulty encountered when trying to use the ‘extended chain method’ for hidden Markov models with a continuous state space is that some components of the extended chain lie in an infinite-dimensional functional space. This is in sharp contrast to the case of a finite hidden state space in which the predictive density can always be identified with an element of \mathbb{R}^M , where M is the number of different possible values for X_n . Instead of the Riesz representation theorem (Mevel 1997), we investigate another approach based on the asymptotic tightness of a sequence of measures to prove the existence of an invariant measure for the extended chain (Sunyach 1975). It is then possible to extend the result of Leroux (1992) under the basic assumption that the model is identifiable (in the sense that different parameter values lead to different

distributions for the stationarized version of the observed process; see the definition below), a point covered by neither Mevel (1997) nor Jensen and Petersen (1999). We stress that the technique developed in Leroux (1992) to prove consistency, which is based on the Kingman subadditive theorem, cannot be immediately adapted to continuous state space. For technical reasons detailed below, we restrict our attention to cases where the hidden state space can be assumed to be compact.

The rest of the paper is organized as follows. In section 2, we first prove the geometric ergodicity of the extended chain. As a corollary, pointwise convergence of the normalized log-likelihood is obtained and its limit is identified with the generalized Kullback–Leibler divergence. The final part of Section 2 deals with the consistency of the maximum likelihood estimator. Section 3 is devoted to the asymptotic normality of the maximum likelihood estimator, using the technique of Section 2 applied to a further extended Markov chain, encompassing the gradient and the Hessian of the prediction filter.

2. Consistency of the maximum likelihood estimator

Let $\{X_n\}_{n=0}^\infty$ be a Markov chain on a compact set $\mathbf{K} \subset \mathcal{X}$, where \mathcal{X} is a separable state space, equipped with a metrizable topology, and $\mathcal{B}(\mathbf{K})$ is the associated Borel σ -field. We denote by $\{Q_\theta(x, \mathcal{A}), x \in \mathbf{K}, \mathcal{A} \in \mathcal{B}(\mathbf{K})\}$ the Markov transition kernel (see Meyn and Tweedie 1993) of the chain. We also let $\{Y_n\}_{n=0}^\infty$ be a sequence of random variables taking values in a space \mathcal{Y} which is separable with respect to a topology that is metrizable by a complete metric. Conditional on $\{X_n\}_{n=0}^\infty$, the Y_n are independent, with a conditional density $g_\theta(y|X_n)$ dependent on X_n only. Here the density is with respect to a σ -finite measure ν on the Borel σ -field $\mathcal{B}(\mathcal{Y})$. \mathcal{X} and \mathcal{Y} are usually subsets of \mathbb{R}^s and \mathbb{R}^t respectively, but they may also be higher-dimensional spaces. Moreover, both Q_θ and g_θ depend on a parameter θ in Θ , where Θ is a compact subset of \mathbb{R}^p . Θ is equipped with a norm denoted by $\|\cdot\|$. The true parameter value will be denoted θ^* and is assumed to be in the interior of Θ .

Assume that each transition kernel Q_θ has a density q_θ with respect to the same σ -finite dominating measure γ on \mathcal{X} . For notational simplicity, it is assumed that the initial distribution of $\{X_k\}$ under θ^* has a density with respect to γ denoted by π^* . We stress that π^* is not necessarily the invariant distribution of the chain. Denote by P^* the probability distribution of $\{Y_n\}_{n \geq 0}$ induced by the parameter (θ^*, π^*) , and let E^* be the associated expectation. The convergence results that will be proved throughout this paper hold under P^* . In the following, for $m \leq n$, denote by Y_n^m the family of random variables (Y_n, \dots, Y_m) . The $L_p(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ -norm will be denoted $\|\cdot\|_p$, and C, C', \dots will denote unspecified finite constants which may take different values upon each appearance. Moreover, for any measurable function f on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, write $\text{ess sup}(f) = \inf\{M \geq 0, \gamma(\{M < |f|\}) = 0\}$ and, if f is non-negative, $\text{ess inf}(f) = \sup\{M \geq 0, \gamma(\{M > f\}) = 0\}$ (with obvious conventions if those sets are empty).

As outlined in Section 1, the central idea consists in writing the log-likelihood of the observations as an additive function of a Markov chain, comprising the observations $\{Y_n\}$, the current state X_n of the initial Markov chain and the predictive density of the state, often referred to as the filter in the HMM literature.

We define a family of probability density functions (pdfs) $\{f_{\theta,n}^\zeta\}_{n \geq 0}$ over $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, by the recurrence relations

$$f_{\theta,0}^\zeta(v) = \zeta(v),$$

$$f_{\theta,n+1}^\zeta(v) \triangleq \Phi_1(Y_n, f_{\theta,n}^\zeta; \theta) \triangleq \frac{\int_{\mathbf{K}} q_\theta(u, v) g_\theta(Y_n|u) f_{\theta,n}^\zeta(u) d\gamma(u)}{\int_{\mathbf{K}} g_\theta(Y_n|u) f_{\theta,n}^\zeta(u) d\gamma(u)}, \tag{1}$$

for all $v \in \mathbf{K}$, where ζ is an initial pdf. Note that $f_{\theta,n+1}^\zeta$ is a deterministic function of $(Y_n, f_{\theta,n}^\zeta)$. If $\zeta = \pi^*$, it is easily checked that $f_{\theta^*,n}^{\pi^*}$ is the predictive density distribution, that is, the density of the conditional distribution of X_n given Y_{n-1}^0 . Otherwise, $f_{\theta,n}^\zeta$ is usually called the prediction filter.

We consider the following family of contrast functions:

$$l_n(\theta, \zeta) = \sum_{m=0}^{n-1} \log \left(\int_{\mathbf{K}} g_\theta(Y_m|x) f_{\theta,m}^\zeta(x) d\gamma(x) \right). \tag{2}$$

When ζ is fixed and equal to π^* , then $l_n(\theta, \pi^*)$ is the log-likelihood function for the parameter θ . We show below that the choice of the initial predictive density function ζ does not affect the limiting value of $(1/n)l_n(\theta, \zeta)$. This parameter can thus be chosen by the user. Moreover, the parameter θ^* is the only one that can be consistently estimated. As usual, for ergodic non-stationary chains, the initial distribution cannot be consistently estimated. We define the state space $\mathbf{E} \triangleq \mathbf{K} \times \mathcal{Y} \times \mathbf{S}^2$, where $\mathbf{S}^+ \triangleq \{f \in L_1(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma); f \geq 0, \|f\|_1 = 1\}$. In the following, \mathbf{S}^+ is equipped with the topology induced by the L_1 -norm, and $\mathcal{B}(\mathbf{S}^+)$ denotes the corresponding Borel σ -field. We denote by $\mathcal{B}(\mathbf{E})$ the Borel σ -field on the space \mathbf{E} itself, induced by the product topology. On the state space \mathbf{E} , define the extended Markov chain $Z_n \triangleq (X_n, Y_n, F_n, F_n^*)$ by an initial law λ , and the following transition kernel: for all (x, y, f, f^*) in \mathbf{E} and $\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\mathcal{F}^*}$ in $\mathcal{B}(\mathbf{K}) \times \mathcal{B}(\mathcal{Y}) \times \mathcal{B}(\mathbf{S}^+) \times \mathcal{B}(\mathbf{S}^+)$,

$$\Pi_\theta((x, y, f, f^*); \mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\mathcal{F}^*}) \tag{3}$$

$$= P(X_{n+1} \in \mathcal{A}_{\mathcal{X}}, Y_{n+1} \in \mathcal{A}_{\mathcal{Y}}, F_{n+1} \in \mathcal{A}_{\mathcal{F}}, F_{n+1}^* \in \mathcal{A}_{\mathcal{F}^*} | X_n = x, Y_n = y, F_n = f, F_n^* = f^*)$$

$$= \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} q_{\theta^*}(x, x') g_{\theta^*}(y'|x') d\gamma(x') d\nu(y') \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\Phi_1(y, f; \theta)) \mathbb{1}_{\mathcal{A}_{\mathcal{F}^*}}(\Phi_1(y, f^*; \theta^*)),$$

where $\Phi_1(y, f; \theta)$ is defined in (1).

In what follows, we will often consider the particular initial distribution $\lambda(\zeta)$ defined for each pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ by

$$\lambda(\zeta)(\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\mathcal{F}^*}) = \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} \pi^*(x) g_{\theta^*}(y|x) d\gamma(x) d\nu(y) \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\zeta) \mathbb{1}_{\mathcal{A}_{\mathcal{F}^*}}(\pi^*). \tag{4}$$

This initial distribution generates a Markov chain $\{Z_n\}$ where the first coordinate corresponds

to the Markov chain $\{X_n\}$, the second to the observations $\{Y_n\}$, the third to the prediction filter $\{f_{\theta,n}^\zeta\}$ and the fourth to the predictive density of X_n given Y_{n-1}^0 , that is, $\{f_{\theta^*,n}^{\pi^*}\}$.

Finally, we denote by $P_{\theta,\lambda}$ the probability distribution of the chain $\{Z_n\}$ induced by the initial distribution λ and the transition kernel Π_θ ($P_{\theta,\lambda}$ is classically defined on $\bigvee_{n=1}^\infty \mathcal{B}(\mathbf{E}_n)$, where $(\mathbf{E}_n)_{n \geq 0}$ is a sequence of copies of \mathbf{E}).

Using this definition, and assuming that $\{Z_n\}$ has the initial distribution $\lambda(\zeta)$, the normalized log-likelihood may be expressed as

$$\frac{1}{n} l_n(\theta, \zeta) = \frac{1}{n} \sum_{m=0}^{n-1} h_\theta(Z_m) \quad P_{\theta,\lambda(\zeta)} \text{ almost surely,} \tag{5}$$

where

$$h_\theta(x, y, f, f^*) \triangleq \log \left(\int_u g_\theta(y|u) f(u) d\gamma(u) \right). \tag{6}$$

Equation (5) shows that the normalized log-likelihood of the observations for the parameter θ can be written as an additive functional of Z_n . Thus, following ideas of LeGland and Mevel (2000a; 2000b), geometric ergodicity of the extended Markov chain will be the key ingredient in proving the consistency of the maximum likelihood estimator.

It is clear from (2) that the log-likelihood is indeed a function of the first three components of Z_n , which themselves form a Markovian process. Extending the chain Z_n so as to include the prediction density corresponding to the actual value θ^* of the parameter is, however, needed to obtain the consistency result of Section 2.4 below (for a closely related idea, see Leroux 1992).

The following definitions and assumptions are stated for future reference. For y in \mathcal{Y} , define

$$\delta(y) \triangleq \sup_{\theta \in \Theta} \delta_\theta(y) \triangleq \sup_{\theta \in \Theta} \left[\frac{\text{ess sup}_x g_\theta(y|x)}{\text{ess inf}_x g_\theta(y|x)} \right], \tag{7}$$

$$\epsilon \triangleq \inf_{\theta \in \Theta} \epsilon_\theta \triangleq \inf_{\theta \in \Theta} \frac{\text{ess inf}_{x,x'} q_\theta(x, x')}{\text{ess sup}_{x,x'} q_\theta(x, x')}. \tag{8}$$

Assumption 1. $0 < \epsilon < 1$.

For $s > 0$, define $\Delta_s \triangleq \text{ess sup}_x \int \sup_\theta [\delta_\theta(y)]^s g_{\theta^*}(y|x) d\nu(y)$.

Assumption 2. Δ_1 is finite.

For $s > 0$, define $\Gamma_s \triangleq \text{ess sup}_x \int \sup_\theta [k_\theta(y)]^s g_{\theta^*}(y|x) d\nu(y)$, where $k_\theta(y) \triangleq \text{ess sup}_x |\log g_\theta(y|x)|$.

Assumption 3. There exists some $s > 1$ such that Γ_s is finite.

If Assumption 1 holds, then $\inf_\theta \text{ess inf}_{x,x'} q_\theta(x, x') > 0$ (note that $\text{ess sup}_{x,x'} q_\theta(x, x') \geq \gamma(\mathbf{K})^{-1}$). Doeblin's condition is then satisfied for the Markov chain associated

with Q_θ (see Meyn and Tweedie 1993, p. 391). Moreover, the kernel Q_θ is strongly aperiodic (see Meyn and Tweedie 1993, p. 118). Thus, by Theorems 16.0.2 and 16.2.4 in Meyn and Tweedie (1993, p. 384), there exists some real number $0 < \rho_0 < 1$ (uniform in θ), such that, for all θ in Θ , there exists a probability measure on $\mathcal{B}(\mathbf{K})$, having density α_θ with respect to γ , such that, for all x in \mathbf{K} ,

$$\|Q_\theta^n(x, \cdot) - \alpha_\theta(\cdot)d\gamma(\cdot)\|_{TV} \leq \rho_0^n, \tag{9}$$

where $\|\cdot\|_{TV}$ is the total variation norm, and Q_θ^n is the kernel Q_θ iterated n times. Then $\alpha_\theta(\cdot)d\gamma(\cdot)$ is the unique invariant probability measure of the chain. Thus, for all θ in Θ , the Markov chain associated with the kernel Q_θ is uniformly ergodic.

Remark 1. Note that we may relax $\epsilon \in]0, 1[$ (Assumption 1) and derive the same results under the following:

Assumption 1'. There exists m in \mathbb{N}^* such that for all θ in Θ , Q_θ is aperiodic and

$$0 < \inf_\theta \frac{\text{ess inf}_{x,x'} q_\theta^m(x, x')}{\text{ess sup}_{x,x'} q_\theta^m(x, x')} < 1,$$

where q_θ^m denotes the density of the kernel Q_θ iterated m times.

In that case, all the properties concerning the Markov chain associated with Q_θ are derived from properties concerning the m -skeleton (see the definition in Meyn and Tweedie 1993, p. 68).

2.1. Exponential memorylessness of the prediction filter

Ergodic properties for $\{Y_n\}$ are inherited from those of $\{X_n\}$ (see, for instance, Leroux 1992; Celeux *et al.* 1993). The point of this subsection is to obtain those of the extended chain $Z_n = (X_n, Y_n, F_n, F_n^*)$ from the exponential memorylessness of the prediction filter. Iterating the recurrence relation (1) $n - m + 1$ times backwards ($m \leq n$) yields $f_{\theta, n+1}^\zeta = \Phi_{n-m+1}(Y_n^m, f_{\theta, m}^\zeta; \theta)$, where

$$\Phi_{n-m+1}(Y_n^m, f; \theta)(v) \triangleq$$

$$\frac{\int_{u_n} q_\theta(u_n, v) g_\theta(Y_n|u_n) \cdots \int_{u_m} q_\theta(u_m, u_{m+1}) g_\theta(Y_m|u_m) f(u_m) d\gamma(u_n) \cdots d\gamma(u_m)}{\int_{u_n} g_\theta(Y_n|u_n) \cdots \int_{u_m} q_\theta(u_m, u_{m+1}) g_\theta(Y_m|u_m) f(u_m) d\gamma(u_n) \cdots d\gamma(u_m)}.$$

Using this expression, we may show the following exponential memorylessness inequality, which is a non-trivial adaptation from Mevel (1997, Proposition 2.2, part A).

Proposition 1. Suppose that Assumption 1 holds. Then for all f, f' in \mathbf{S}^+ ,

$$\|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \leq 2\epsilon_\theta^{-1} \delta_\theta(Y_m)(1 - \epsilon_\theta)^{n-m+1} \|f - f'\|_1. \quad (10)$$

The proof is given in Appendix A. The upper-bound of the right-hand side only depends on the initial conditions (Y_m, f, f') . Note that Assumption 1 implies that ϵ_θ is bounded from below by a strictly positive ϵ . Hence, $1 - \epsilon_\theta$ is bounded from above by $1 - \epsilon$ and the geometrical rate of the right-hand side of (10) can be bounded uniformly for θ in Θ , provided that $\delta(Y_m)$ is finite.

2.2. Geometric ergodicity of the extended Markov chain

In this section, it is shown that the exponential memorylessness of the prediction filter implies the geometric ergodicity (in a sense to be defined) of the extended chain $\{Z_n\}$.

Definition 1. *Lip(E) is the set of real-valued measurable functions h on $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$ such that, for all (x, y) in $\mathbf{K} \times \mathcal{Y}$, there exist $\text{lip}(h, x, y)$ and $\text{k}(h, x, y)$ such that, for all f_1, f_1^*, f_2, f_2^* in \mathbf{S}^+ ,*

$$\begin{aligned} |h(x, y, f_1, f_1^*) - h(x, y, f_2, f_2^*)| &\leq \text{lip}(h, x, y)(\|f_1 - f_2\|_1 + \|f_1^* - f_2^*\|_1), \\ |h(x, y, f_1, f_1^*)| &\leq \text{k}(h, x, y) \end{aligned}$$

and

$$\begin{aligned} \text{lip}(h) &\triangleq \text{ess sup}_x \int \text{lip}(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty, \\ \text{k}(h) &\triangleq \text{ess sup}_x \int \text{k}(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty. \end{aligned}$$

These functions are of Lipschitz type with respect to the two last components and are bounded independently of the two last components by a function which satisfies an integrability condition. We will prove that, for all h in $\text{Lip}(\mathbf{E})$ verifying some moment condition (see (15)), a law of large numbers holds for $\{h(Z_n)\}_{n \geq 0}$.

Proposition 2. *Suppose that Assumptions 1 and 2 hold, and define*

$$\rho \triangleq \max(\rho_0^{1/2}, (1 - \epsilon)^{1/2}), \quad (11)$$

where ϵ and ρ_0 are as defined in (8) and (9). There exists a constant $C > 0$ such that, for all θ in Θ , for all h in $\text{Lip}(\mathbf{E})$, for all z, z' in \mathbf{E} , and for all $n \geq 1$,

$$|\Pi_\theta^n h(z) - \Pi_\theta^n h(z')| \leq C[\text{lip}(h) + \text{k}(h)]\rho^n,$$

where Π_θ^n denotes the n th iterate of Π_θ and $\Pi_\theta^n h(z) = \int \Pi_\theta^n(z, dw)h(w)$.

The proof of this proposition is given in Appendix B. As a consequence, the extended

Markov chain satisfies a geometric ergodicity property, stated in the following corollaries, also proved in Appendix B.

Corollary 1. *Suppose that Assumptions 1 and 2 hold. Then there exists a constant $C > 0$ such that, for all θ in Θ and for all h in $\text{Lip}(\mathbf{E})$, there exists $\Lambda_\theta(h) < \infty$ such that, for all z in \mathbf{E} and for all $n \geq 1$,*

$$|\Pi_\theta^n h(z) - \Lambda_\theta(h)| \leq C[\text{lip}(h) + k(h)] \frac{\rho^n}{1 - \rho}. \tag{12}$$

This inequality may be used in the following way: for any initial probability measure λ on $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$, and for any function h in $\text{Lip}(\mathbf{E})$,

$$\Lambda_\theta(h) = \lim_{n \rightarrow \infty} \mathbb{E}_{\theta, \lambda}(h(Z_n)). \tag{13}$$

In addition, for all h in $\text{Lip}(\mathbf{E})$, there exists a unique solution V of the Poisson equation

$$[I - \Pi_\theta]V(z) = h(z) - \Lambda_\theta(h).$$

Corollary 2. *Suppose that Assumption 1 and 2 hold. The kernel Π_θ admits a unique invariant probability measure λ_θ on $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$, and, for any function h in $\text{Lip}(\mathbf{E})$,*

$$\Lambda_\theta(h) = \int h(z)\lambda_\theta(dz). \tag{14}$$

The type of convergence in Corollary 1 does not coincide with the usual definition of uniform geometric ergodicity (see Meyn and Tweedie 1993, Theorem 16.0.2, p. 384) since our result applies here only to certain Lipschitz functions on \mathbf{E} , whereas in the standard literature about Markov chains the ergodicity is expressed in terms of exponential decrease to zero of the total variation norm, implying that (12) applies to all measurable bounded functions.

Note that it is easy to show that $\Lambda_\theta(h)$ is a linear form. Nevertheless, it is not straightforward to prove that there exists a measure λ_θ on \mathbf{E} representing $\Lambda_\theta(h)$. This is because the Riesz representation theorem does not apply. To prove the existence of an invariant probability measure, we use a technique based on asymptotic tightness of an appropriately defined sequence of probability measures, exploiting the basic ideas of Sunyach (1975).

Remark 2. We will often make use of the following property. Let \mathbf{E}_n be a sequence of copies of the space \mathbf{E} , let \mathcal{A} belong to $\bigvee_{n \geq 0} \mathcal{B}(\mathbf{E}_n)$ and let Υ a real-valued function on $\prod_{n \geq 0} \mathbf{E}_n$, measurable with respect to $\bigvee_{n \geq 0} \mathcal{B}(\mathbf{E}_n)$. Note that, due to the initial distribution $\lambda(\zeta)$, (F_n, F_n^*) is $P_{\theta, \lambda(\zeta)}$ almost surely equal to $(f_{\theta, n}^\zeta, f_{\theta^*, n}^{\pi^*})$ for all n in \mathbb{N} . This implies that

$$\Upsilon[\{Z_n\}_{n \geq 0}] = \Upsilon[\{(X_n, Y_n, f_{\theta, n}^\zeta, f_{\theta^*, n}^{\pi^*})\}_{n \geq 0}] \quad P_{\theta, \lambda(\zeta)}\text{-a.s.}$$

Thus,

$$P_{\theta, \lambda(\zeta)}(\Upsilon(\{Z_n\}_{n \geq 0}) \in \mathcal{A}) = P^*[\Upsilon(\{(X_n, Y_n, f_{\theta, n}^\zeta, f_{\theta^*, n}^{\pi^*})\}_{n \geq 0}) \in \mathcal{A}]$$

since the marginal distribution of $P_{\theta, \lambda(\zeta)}$ on events only depending on $\{(X_n, Y_n)\}_{n \geq 0}$ is equal to P^* .

Proposition 3. *Suppose that Assumptions 1 and 2 hold. For any θ in Θ , for any pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, and for any function h in $\text{Lip}(\mathbf{E})$ such that there exists $s > 1$ satisfying*

$$\text{ess sup}_x \int k^s(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty, \tag{15}$$

the following convergence results hold:

- (i) $\lim_{n \rightarrow \infty} (1/n) \sum_{m=0}^{n-1} h(Z_m) = \Lambda_{\theta}(h) \quad P_{\theta, \lambda(\zeta)}\text{-a.s.};$
- (ii) $\lim_{n \rightarrow \infty} (1/n) \sum_{m=0}^{n-1} h(X_m, Y_m, f_{\theta, m}^{\zeta}, f_{\theta^*, m}^{\pi^*}) = \Lambda_{\theta}(h) \quad P^*\text{-a.s.}$

The proof of this proposition (see Appendix C) is based on the ergodicity of $\{Z_n\}$ and, more precisely, on the existence of a solution to the Poisson equation stated in Corollary 1 using a classical martingale decomposition method (the proof is analogous to that of Meyn and Tweedie 1993, Theorem 17.4.3). Note that the statement (ii) is a direct consequence of (i) using Remark 2.

2.3. Pointwise convergence

The following proposition ensures that, under some regularity conditions, the normalized log-likelihood of the observations converges almost everywhere to a finite function of θ , and identifies the limit function. It is a straightforward application of Proposition 3.

Proposition 4. *Suppose that Assumptions 1–3 hold. Then, for any θ in Θ , there exists a finite value $l(\theta)$ such that, for any initial pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} l_n(\theta, \zeta) = l(\theta) \quad P^*\text{-a.s.}$$

Moreover,

$$l(\theta) = \int \log \left[\int g_{\theta}(y|u) f(u) d\gamma(u) \right] \lambda_{\theta}(dx, dy, df, df^*). \tag{16}$$

Note that the limiting value $l(\theta)$ does not depend on the initial distribution $\lambda(\zeta)$. This means that the limiting value of $l_n(\theta, \zeta)/n$ does not depend on the initial distribution of the chain $\{X_n\}$ (which does not necessarily coincide with the stationary distribution of the chain, in contrast with Bickel *et al.* 1998 and Jensen and Petersen 1999) or on the initial value of the prediction filter ζ , which may be chosen arbitrarily. The limiting value $l(\theta)$ is the integral of h_{θ} with respect to the invariant measure of $\{Z_n\}$.

2.4. Consistency of the maximum likelihood estimator

In order to prove the consistency of the maximum likelihood estimator, we need to check that the limit of the normalized log-likelihood is maximized only at θ^* , the true value of the parameter; that is, $l(\theta) \leq l(\theta^*)$ with equality if and only if $\theta = \theta^*$. Define

$$g_{\theta, n+1}(y_n^0 | X_0) \triangleq g_{\theta}(y_0 | X_0) \int q_{\theta}(X_0, u_1) g_{\theta}(y_1 | u_1) \cdots q_{\theta}(u_{n-1}, u_n) g_{\theta}(y_n | u_n) d\gamma(u_1) \cdots d\gamma(u_n), \quad (17)$$

the density under the parameter θ of $\{Y_n^0\}$ conditionally on X_0 with respect to the measure $\nu^{\otimes(n+1)}$.

We make the following assumption:

Assumption 4. *The equality*

$$\begin{aligned} & \int_{\mathbf{K} \times \mathcal{A}_0 \times \cdots \times \mathcal{A}_p} g_{\theta, p+1}(y_0, \dots, y_p | u_0) d\alpha_{\theta}(u_0) d\nu(y_0) \cdots d\nu(y_p) \\ &= \int_{\mathbf{K} \times \mathcal{A}_0 \times \cdots \times \mathcal{A}_p} g_{\theta', p+1}(y_0, \dots, y_p | u_0) d\alpha_{\theta'}(u_0) d\nu(y_0) \cdots d\nu(y_p) \end{aligned} \quad (18)$$

holds for all non-negative p and all Borel sets \mathcal{A}_i in $\mathcal{B}(\mathcal{Y})$, $0 \leq i \leq p$, if and only if $\theta = \theta'$.

(Recall that α_{β} is the invariant measure for the kernel q_{β} .) This condition is equivalent to the following:

Assumption 4'. *The equality*

$$\begin{aligned} & \int q_{\theta}(u, u_1) \cdots q_{\theta}(u_{p-1}, u_p) g_{\theta}(y_0 | u) \cdots g_{\theta}(y_p | u_p) d\alpha_{\theta}(u) d\gamma(u_1) \cdots d\gamma(u_p) \\ &= \int q_{\theta'}(u, u_1) \cdots q_{\theta'}(u_{p-1}, u_p) g_{\theta'}(y_0 | u) \cdots g_{\theta'}(y_p | u_p) d\alpha_{\theta'}(u) d\gamma(u_1) \cdots d\gamma(u_p) \end{aligned}$$

almost everywhere with respect to the measure $\nu^{\otimes(p+1)}$, for all non-negative p , if and only if $\theta = \theta'$.

Note that when the state space \mathbf{K} is finite, Assumption 4 is implied by Condition 2 in Leroux (1992). In fact, Assumption 4 is the minimal assumption guaranteeing identifiability for the model.

Remark 3. In many models of interest, the parameter itself is identifiable only up to a permutation of states. It is most often possible to reparametrize the model to avoid this type of degeneracy. When it is not possible, one should define a slightly extended notion of identifiability, based on an equivalence relation, for example $\theta \equiv \theta'$ if and only if θ is equal

to θ' up to a permutation of its components. The results presented below can be straightforwardly adapted in such a situation.

Theorem 1. *Under Assumptions 1–4, $l(\theta) \leq l(\theta^*)$ and $l(\theta) = l(\theta^*)$ if and only if $\theta = \theta^*$.*

The key point in the proof is that the difference $l(\theta^*) - l(\theta)$ can be written as the expected difference between two functions under the same measure. This is possible since the fourth coordinate of Z_n is the prediction filter associated with the true value of the parameter θ^* . This is the main motivation for extending (X_n, Y_n, F_n) to (X_n, Y_n, F_n, F_n^*) .

Proof. The first part of the proof follows an idea from Leroux (1992, Lemma 6), but the conclusion has to be adapted to cope with general state space. We first express $l(\theta)$ in a more tractable way.

Using Corollary 1,

$$2l(\theta) = \lim_n \mathbb{E}_{\theta, \lambda(\zeta)}(h_\theta(Z_n) + h_\theta(Z_{n+1})).$$

By direct application of the recurrence relation between $f_{\theta, n}^\zeta$ and $f_{\theta, n+1}^\zeta$, we have

$$\begin{aligned} h_\theta(Z_n) + h_\theta(Z_{n+1}) &= \log \left[\int g_{\theta, 2}(Y_n, Y_{n+1} | u) f_{\theta, n}^\zeta(u) d\gamma(u) \right] && P_{\theta, \lambda(\zeta)}\text{-a.s.} \\ &\triangleq H_\theta(Y_n, Y_{n+1}, f_{\theta, n}^\zeta), \end{aligned}$$

where $g_{\theta, 2}$ is defined by (17). Then, using Remark 2,

$$2l(\theta) = \lim_n \mathbb{E}^*(\mathbb{E}^*(H_\theta(Y_n, Y_{n+1}, f_{\theta, n}^\zeta) | Y_{n-1}^0)). \tag{19}$$

Let $G_\theta(x, y, f, f^*) \triangleq \int H_\theta(y_0, y_1, f) g_{\theta^*, 2}(y_0, y_1 | u) f^*(u) d\nu(y_0) d\nu(y_1) d\gamma(u)$. It is easy to check that G_θ is in $\text{Lip}(\mathbb{E})$. Thus, by Corollary 1,

$$\Lambda_\theta(G_\theta) = \lim_n \mathbb{E}_{\theta, \lambda(\zeta)}(G_\theta(Z_n)). \tag{20}$$

Using the equality

$$\begin{aligned} \mathbb{E}^*[H_\theta(Y_n, Y_{n+1}, f_{\theta, n}^\zeta) | Y_{n-1}^0] &= \mathbb{E}^*(\mathbb{E}^*[H_\theta(Y_n, Y_{n+1}, f_{\theta, n}^\zeta) | X_n, Y_{n-1}^0] | Y_{n-1}^0) \\ &= G_\theta(X_n, Y_n, f_{\theta, n}^\zeta, f_{\theta^*, n}^{\pi^*}), \end{aligned}$$

combined with (19) and (20), yields that

$$\begin{aligned} 2l(\theta) &= \Lambda_\theta(G_\theta) \\ &= \int \log \left[\int g_{\theta, 2}(y_0, y_1 | u_0) f(u_0) d\gamma(u_0) \right] \\ &\quad \times g_{\theta^*, 2}(y_0, y_1 | u) f^*(u) d\nu(y_0) d\nu(y_1) d\gamma(u) \lambda_\theta(\mathbf{K}, \mathcal{Y}, \mathbf{d}f, \mathbf{d}f^*). \end{aligned}$$

Define the function $h^*(x, y, f, f^*) \triangleq \log(\int g_{\theta^*}(y|u)f^*(u)d\gamma(u))$. Replacing the function h_θ by h^* , it is easily checked in the same way that

$$2l(\theta^*) = \int \log \left[\int g_{\theta^*,2}(y_0, y_1|u_0)f^*(u_0)d\gamma(u_0) \right] \\ \times g_{\theta^*,2}(y_0, y_1|u)f^*(u)d\nu(y_0)d\nu(y_1)d\gamma(u)\lambda_\theta(\mathbf{K}, \mathcal{Y}, \mathbf{S}^+, df^*).$$

Hence,

$$2(l(\theta^*) - l(\theta)) = \int \left[\int g_{\theta^*,2}(y_0, y_1|u)f^*(u)d\gamma(u) \right] \\ \times \log \left(\frac{\int g_{\theta^*,2}(y_0, y_1|u)f^*(u)d\gamma(u)}{\int g_{\theta,2}(y_0, y_1|u)f(u)d\gamma(u)} \right) d\nu(y_0)d\nu(y_1)\lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*)$$

which is the integral of a Kullback–Leibler divergence and is thus non-negative. We now show that the equality $l(\theta) = l(\theta^*)$ implies that $\theta = \theta^*$ under the assumption of identifiability (Assumption 4). If $l(\theta) = l(\theta^*)$ the Kullback–Leibler divergence is null, and thus

$$\int g_{\theta^*,2}(y_0, y_1|u)f^*(u)d\gamma(u) = \int g_{\theta,2}(y_0, y_1|u)f(u)d\gamma(u) \tag{21}$$

almost everywhere with respect to the measure $d\nu(y_0)d\nu(y_1)\lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*)$. This technique can easily be generalized by computing the quantities $(m + 1)l_n(\theta, \zeta)$ for each integer m , such that $l(\theta) = l(\theta^*)$ implies almost everywhere with respect to the measure $\lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*)$, that, for all $m \geq 1$,

$$\int g_{\theta^*,m+1}(y_0, \dots, y_m|u)f^*(u)d\gamma(u) = \int g_{\theta,m+1}(y_0, \dots, y_m|u)f(u)d\gamma(u) \\ \nu(dy_0) \otimes \dots \otimes \nu(dy_m)\text{-a.e.} \tag{22}$$

Now, let us prove that Assumption 4 holds. Let (f_0, f_0^*) be in $\mathbf{S}^+ \times \mathbf{S}^+$ such that property (22) is satisfied. Let $p \leq m$ and h be any bounded, continuous function on \mathcal{Y}^{p+1} . Multiplying both sides of the equation by $h(y_{m-p}, \dots, y_m)$, and then taking the integral with respect to $d\nu(y_0) \dots d\nu(y_m)$, yields

$$\int F_\theta(u_{m-p})q_\theta^{m-p}(u_0, u_{m-p})f_0(u_0)d\gamma(u_0)d\gamma(u_{m-p}) \\ = \int F_{\theta^*}(u_{m-p})q_{\theta^*}^{m-p}(u_0, u_{m-p})f_0^*(u_0)d\gamma(u_0)d\gamma(u_{m-p}), \tag{23}$$

where, for any β ,

$$F_\beta(u) \triangleq \int h(y_{m-p}, \dots, y_m)g_{\beta,p+1}(y_{m-p}, \dots, y_m|u)d\nu(y_{m-p}) \dots d\nu(y_m), \tag{24}$$

which is continuous and bounded by $\text{ess sup}_{u \in \mathcal{Y}^{p+1}} |h(u)|$ on \mathbf{K} . Now, using (9), we have, for any u in \mathbf{K} ,

$$\left| \int F_\theta(u') q_\theta^{m-p}(u, u') d\gamma(u') - \int F_\theta(u') d\alpha_\theta(u') \right| \leq \rho_0^{m-p} \times \operatorname{ess\,sup}_{u \in \mathcal{Y}^{p+1}} |h(u)|.$$

This implies, by dominated convergence,

$$\lim_{m \rightarrow \infty} \int F_\theta(u') q_\theta^{m-p}(u, u') f_0(u) d\gamma(u) d\gamma(u') = \int F_\theta(u') d\alpha_\theta(u').$$

This result also holds with the parameter θ^* instead of θ . Now, combining with (23) yields

$$\int F_\theta(u') d\alpha_\theta(u') = \int F_{\theta^*}(u') d\alpha_{\theta^*}(u').$$

Replacing F_θ and F_{θ^*} by their definition in (24) and knowing that h is any bounded continuous function in \mathcal{Y}^{p+1} ,

$$\int g_{\theta, p+1}(y_0, \dots, y_p | u_0) d\alpha_\theta(u_0) = \int g_{\theta^*, p+1}(y_0, \dots, y_p | u_0) d\alpha_{\theta^*}(u_0)$$

almost everywhere with respect to the measure $d\nu(y_0) \cdots d\nu(y_p)$, for any p in \mathbb{N} . Thus $\theta = \theta^*$ by Assumption 4. This completes the proof. \square

Let $\hat{\theta}_n(\zeta) \triangleq \arg \max_\theta l_n(\theta, \zeta)$ be the maximum likelihood estimator for a fixed initial pdf ζ (the existence of which is guaranteed under assumptions stated below). It has been proved in Proposition 4 and Theorem 1 that the normalized log-likelihood for a fixed θ converges P^* -a.s. to $l(\theta)$, which is maximum at the point θ^* . Under additional assumptions on the regularity of some functions with respect to θ , uniformity (in a sense to be defined) in the convergence of the normalized log-likelihood may be shown, and thus, consistency can be stated, using Wald's (1949) consistency proof for the M -estimator.

Define for all y in \mathcal{Y} and for $\eta > 0$, the expression $\omega^g(y, \eta) \triangleq \sup_{\|\theta - \theta'\| \leq \eta} \operatorname{ess\,sup}_x |g_\theta(y|x) - g_{\theta'}(y|x)|$. Let $\eta > 0$ and $\theta, \theta' \in \Theta$ such that $\|\theta - \theta'\| < \eta$. Then, for all pdfs ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$,

$$\begin{aligned} & l_n(\theta, \zeta) - l_n(\theta', \zeta) \\ &= \sum_{m=0}^{n-1} \log \left(1 + \frac{\int [g_\theta(Y_m|u) f_{\theta, m}^\zeta(u) - g_{\theta'}(Y_m|u) f_{\theta', m}^\zeta(u)] d\gamma(u)}{\int g_{\theta'}(Y_m|u) f_{\theta', m}^\zeta(u) d\gamma(u)} \right) \\ &\leq \sum_{m=0}^{n-1} \left(\frac{\int |g_\theta(Y_m|u) - g_{\theta'}(Y_m|u)| f_{\theta, m}^\zeta(u) d\gamma(u)}{\int g_{\theta'}(Y_m|u) f_{\theta', m}^\zeta(u) d\gamma(u)} + \frac{\int g_{\theta'}(Y_m|u) |f_{\theta, m}^\zeta(u) - f_{\theta', m}^\zeta(u)| d\gamma(u)}{\int g_{\theta'}(Y_m|u) f_{\theta', m}^\zeta(u) d\gamma(u)} \right), \\ & |l_n(\theta, \zeta) - l_n(\theta', \zeta)| \leq 2 \sum_{m=0}^{n-1} \left(\frac{\omega^g(Y_m, \eta)}{\inf_\theta \operatorname{ess\,inf}_x g_\theta(Y_m|x)} + \delta(Y_m) \|f_{\theta, m}^\zeta - f_{\theta', m}^\zeta\|_1 \right). \end{aligned} \quad (25)$$

Define for all y in \mathcal{Y} , for all s, t in \mathbb{N} and for all $\eta > 0$,

$$\delta'(y, \eta) \triangleq \left[\frac{\omega^g(y, \eta)}{\inf_{\theta} \text{ess inf}_x g_{\theta}(y|x)} \right], \quad (26)$$

$$\Delta'_s(\eta) \triangleq \text{ess sup}_x \int [\delta'(y, \eta)]^s g_{\theta^*}(y|x) d\nu(y), \quad (27)$$

$$(\Delta\Delta')_{s,t}(\eta) \triangleq \text{ess sup}_x \int [\delta(y)]^s [\delta'(y, \eta)]^t g_{\theta^*}(y|x) d\nu(y), \quad (28)$$

$$\omega^q(\eta) \triangleq \sup_{\|\theta - \theta'\| \leq \eta} \int \text{ess sup}_u |q_{\theta}(u, v) - q_{\theta'}(u, v)| d\gamma(v). \quad (29)$$

Assumption 5. For all y in \mathcal{Y} ,

$$\lim_{\eta \rightarrow 0} \omega^q(\eta) = 0,$$

$$\lim_{\eta \rightarrow 0} \omega^g(y, \eta) = 0.$$

Assumption 6. Δ_2 is finite and, for some positive η_0 , $\Delta'_1(\eta_0)$ and $(\Delta\Delta')_{1,1}(\eta_0)$ are finite.

We prove in Appendix C the following lemma:

Lemma 1. Under Assumption 1,

$$\sup_{\zeta \in \mathbb{S}^+} \|f_{\theta, m}^{\zeta} - f_{\theta', m}^{\zeta}\|_1 \leq 2\epsilon^{-1} \sum_{k=0}^{m-1} \delta(Y_{k+1}) \rho^{m-k} [\omega^q(\eta) + 2\delta'(Y_k, \eta)] + [\omega^q(\eta) + 2\delta'(Y_m, \eta)]. \quad (30)$$

Combining (25) with this lemma, we prove under Assumption 5 that $\theta \mapsto l_n(\theta, \zeta)$ is continuous. The continuity of $\theta \mapsto l(\theta)$ is then straightforward: the family of functions $(\Pi_{\theta}^n h_{\theta}(z))_{n \geq 0}$ converges uniformly in θ to $l(\theta)$ as we have

$$|\Pi_{\theta}^n h_{\theta}(z) - \Lambda_{\theta}(h_{\theta})| \leq C[\Delta_1 + \Gamma_1] \frac{\rho^n}{1 - \rho}.$$

Moreover,

$$\begin{aligned} \Pi_{\theta}^n h_{\theta}(z) &= \int \log \left(\int g_{\theta}(y_n|u) f_{\theta, n}^{\zeta}(u) d\gamma(u) \right) q_{\theta^*}(x, x_1) \cdots q_{\theta^*}(x_{n-1}, x_n) \\ &\quad \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_1) \cdots d\gamma(x_n) d\nu(y_1) \cdots d\nu(y_n) \end{aligned}$$

The functions $\theta \mapsto \log(\int g_{\theta}(y_n|u) f_{\theta, n}^{\zeta}(u) d\gamma(u))$ are continuous (by the same argument as for $l_n(\theta, \zeta)$) and uniformly integrable as Γ_s is finite for some $s > 1$. Then we obtain that $\theta \mapsto \Pi_{\theta}^n h_{\theta}(z)$ is continuous, and so is $\theta \mapsto l(\theta)$.

Now write $\omega_n^l(\eta) \triangleq \sup\{(1/n)|l_n(\theta, \zeta) - l_n(\theta', \zeta)|, \|\theta - \theta'\| < \eta\}$. Combining (25), (30) and Assumptions 5 and 6, we obtain that

$$\lim_k \limsup_n E^* \left[\omega_n^l \left(\frac{1}{k} \right) \right] = 0.$$

Thus, using that

$$P^* \left[\omega_n^l \left(\frac{1}{k} \right) \geq \epsilon_k \right] \leq \frac{1}{\epsilon_k} E^* \left[\omega_n^l \left(\frac{1}{k} \right) \right], \quad (31)$$

it is easy to choose ϵ_k decreasing to 0 such that

$$\lim_k \limsup_n P^* \left[\omega_n^l \left(\frac{1}{k} \right) \geq \epsilon_k \right] = 0.$$

Then a straightforward adaptation of Dacunha-Castelle and Duflo (1986, Theorem 3.2.8, p. 93), shows that the maximum likelihood estimator is consistent.

Theorem 2. *Suppose that Assumptions 1–6 hold. Then for all initial pdf ζ ,*

$$\hat{\theta}_n(\zeta) \xrightarrow[n \rightarrow \infty]{} \theta^* \quad \text{in } P^*\text{-probability.}$$

2.5. Examples

Let us give some illustrations.

Example 1 *Markov chain in additive noise.*

Define $Y_t = h(X_t) + B_t$, where $\{B_t\}$ is a sequence of independent and identically distributed centred Gaussian random variables with variance σ^2 , and $\{X_t\}$ is a Markov chain on a compact set in \mathbb{R}^d , independent of $\{B_t\}$, and with transition Q_α . The parameter here is $\theta = (\alpha, \sigma^2)$. Let ϕ_t be the log-Laplace function of the marginal distributions of $\{Y_t\}$,

$$\phi_t(u) = \log(E(e^{uY_t})) = \log E(e^{uh(X_t)}) + \log(e^{u^2\sigma^2/2}),$$

so that

$$\frac{2}{u^2} \phi_t(u) = \sigma^2 + \frac{2}{u^2} \log(E(e^{uh(X_t)})) \xrightarrow[u \rightarrow \infty]{} \sigma^2.$$

σ^2 is thus identifiable. If different parameters α lead to different stationary distributions for the process $\{h(X_t)\}$, then Assumption 4 holds. Such models are used in the following applications:

- Underwater acoustics (see Quinn 1995), where $h(x) = A \cos(2\pi x)$, and A is a constant.
- Speech processing, count data analysis, and so on (see Pagano 1974), where $h(x) = x$ and $\{X_t\}$ is an autoregressive process of first order, with bounded innovations, or more generally an ARMA process. In this last case, just note that Theorems 1 and 2 may be

generalized when Assumption 1 holds only for a power of the transition Q_α – that is, when Assumption 1' holds.

Example 2 *Markov chain in multiplicative noise.*

Define $Y_t = B_t \cdot h(X_t)$, where $\{B_t\}$ is a process of independent and identically distributed, centred Gaussian random variables with variance σ^2 , $\{X_t\}$ is a Markov chain on a compact subset of \mathbb{R} , independent of $\{B_t\}$, and h is any measurable function. Such models are used in a number of different areas (see Chib *et al.* 1998).

3. Asymptotic normality of the maximum likelihood estimator

The asymptotic normality of the maximum likelihood estimator involves a second-order Taylor expansion of the log-likelihood. We shall use similar arguments to those for the extended chain Z_n , but we will now consider Markov chains involving both the gradient and the Hessian of the prediction filter. These chains have an exponential lack of memory of the initial condition, and we prove that they are uniformly ergodic. Denote by $\partial_k(\cdot)$ the first-order partial derivative with respect to the component θ_k of θ , and by ∇, ∇^2 the Jacobian and Hessian operators with respect to the parameter θ . Define, for all y in \mathcal{Y} ,

$$\delta'(y) \triangleq \sup_{\theta, \theta'} \max_{1 \leq k \leq p} \left[\frac{\text{ess sup}_x |\partial_k g_\theta(y|x)|}{\text{ess inf}_x g_{\theta'}(y|x)} \right] \tag{32}$$

and, for all s, t in \mathbb{N} ,

$$\Delta'_s \triangleq \text{ess sup}_x \int [\delta'(y)]^s g_{\theta^*}(y|x) d\nu(y), \tag{33}$$

$$(\Delta\Delta')_{s,t} \triangleq \text{ess sup}_x \int [\delta(y)]^s [\delta'(y)]^t g_{\theta^*}(y|x) d\nu(y). \tag{34}$$

Assumption 7. (i) For any (x, x', y) in $\mathbf{K} \times \mathbf{K} \times \mathcal{Y}$, the functions $\theta \mapsto g_\theta(y|x)$ and $\theta \mapsto q_\theta(x, x')$ are differentiable.

(ii) $\sup_\theta \text{ess sup}_{u,v} q_\theta(u, v) < \infty$, $\max_{1 \leq k \leq p} \sup_\theta \text{ess sup}_{u,v} |\partial_k q_\theta(u, v)| < \infty$. For some $s > 1$, $\Delta'_{2s}, \Delta_4, (\Delta\Delta')_{4,1}$ and $(\Delta\Delta')_{3,2}$ are finite.

The first step consists in differentiating the recurrence relation between $f_{\theta,n}^\zeta$ and $f_{\theta,n+1}^\zeta$ (which is possible under Assumption 7). Then $\{\nabla f_{\theta,n}^\zeta\}$ appears as a functional autoregressive chain. For all $1 \leq k \leq p$,

$$\partial_k f_{\theta,n+1}^\zeta(v) = \int_u a_\theta(Y_n, f_{\theta,n}^\zeta)(u, v) \partial_k f_{\theta,n}^\zeta(u) d\gamma(u) + U_{\theta,k}(Y_n, f_{\theta,n}^\zeta)(v), \tag{35}$$

where

$$a_\theta(y, f)(u, v) \triangleq \left(q_\theta(u, v) - \frac{\int q_\theta(s, v) g_\theta(y|s) f(s) d\gamma(s)}{\int g_\theta(y|s) f(s) d\gamma(s)} \right) \frac{g_\theta(y|u)}{\int g_\theta(y|s) f(s) d\gamma(s)}$$

and

$$U_{\theta, k}(y, f)(v) \triangleq \frac{\int \partial_k q_\theta(u, v) g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} + \int \left(q_\theta(u, v) - \frac{\int q_\theta(s, v) g_\theta(y|s) f(s) d\gamma(s)}{\int g_\theta(y|s) f(s) d\gamma(s)} \right) \times \frac{\partial_k g_\theta(y|u) f(u)}{\int g_\theta(y|s) f(s) d\gamma(s)} d\gamma(u).$$

Denote by U_θ the vector function $(U_{\theta, k})_{1 \leq k \leq p}$.

We consider the space $\mathbf{E}' = \mathbf{K} \times \mathcal{Y} \times \mathbf{S}^+ \times \Sigma$, where

$$\Sigma \triangleq \left\{ \sigma = (\sigma_k)_{1 \leq k \leq p}; \sigma_k \in L_1(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma); \int_u \sigma_k(u) d\gamma(u) = 0 \right\}.$$

The space Σ is equipped with the topology induced by the L_1 -norm, and denote by $\mathcal{B}(\mathbf{E}')$ the Borel σ -field induced by the product topology on \mathbf{E}' . For all (y, f, σ) in $\mathcal{Y} \times \mathbf{S}^+ \times \Sigma$, θ in Θ , and v in \mathbf{K} , write

$$\Psi_1(y, f, \sigma; \theta)(v) \triangleq \int a_\theta(y, f)(u, v) \sigma(u) d\gamma(u) + U_\theta(y, f)(v).$$

(Note that $\Psi_1 = (\Psi_1^k)_{1 \leq k \leq p}$ is a vector in \mathbb{R}^p .)

We define an extended Markov chain $\tilde{Z}_n = (X_n, Y_n, F_n, G_n)$ on $(\mathbf{E}', \mathcal{B}(\mathbf{E}'))$ by its initial law $\tilde{\lambda}$, and the transition kernel $\tilde{\Pi}_\theta$ given, for all (x, y, f, σ) in \mathbf{E}' and $\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm}$ in $\mathcal{B}(\mathbf{K}) \times \mathcal{B}(\mathcal{Y}) \times \mathcal{B}(\mathbf{S}^+) \times \mathcal{B}(\Sigma)$, by

$$\begin{aligned} & \tilde{\Pi}_\theta((x, y, f, \sigma); \mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm}) \\ &= P(X_{n+1} \in \mathcal{A}_{\mathcal{X}}, Y_{n+1} \in \mathcal{A}_{\mathcal{Y}}, F_{n+1} \in \mathcal{A}_{\mathcal{F}}, G_{n+1} \in \mathcal{A}_{\pm} | X_n = x, Y_n = y, F_n = f, G_n = \sigma) \\ &= \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} q_{\theta^*}(x, x') g_{\theta^*}(y' | x') d\gamma(x') d\nu(y') \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\Phi_1(y, f; \theta)) \mathbb{1}_{\mathcal{A}_{\pm}}(\Psi_1(y, f, \sigma; \theta)). \end{aligned}$$

In what follows, we will often consider the particular initial distribution $\tilde{\lambda}(\xi)$ defined for each pdf ξ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ by

$$\tilde{\lambda}(\xi)(\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm}) = \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} \pi^*(x) g_{\theta^*}(y|x) d\gamma(x) d\nu(y) \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\xi) \mathbb{1}_{\mathcal{A}_{\pm}}(0). \quad (36)$$

We denote by $\tilde{P}_{\theta, \tilde{\lambda}(\xi)}$ the law of $\{\tilde{Z}_n\}$ induced by $\tilde{\lambda}(\xi)$, and $\tilde{\Pi}_\theta$, and by $\tilde{E}_{\theta, \tilde{\lambda}(\xi)}$ the corresponding expectation. The exponential ergodicity of the extended chain $\{\tilde{Z}_n\}$ (similar to the result obtained in Proposition 1) and convergence of normalized sums of some functions of $\{\tilde{Z}_n\}$, are stated in Appendix D.

These properties, combined with the results of Propositions 6 and 7, ensure the convergence of the gradient of the normalized log-likelihood.

Assumption 8. For any (x, x', y) in $\mathbf{K} \times \mathbf{K} \times \mathcal{Y}$, the functions $\theta \mapsto g_\theta(y|x)$ and $\theta \mapsto q_\theta(x, x')$ are twice differentiable. We also assume that

$$\sup_{\theta} \max_{1 \leq k, l \leq p} \operatorname{ess\,sup}_{u, v} |\partial_{k, l}^2 q_\theta(u, v)| < \infty.$$

Define for all y in \mathcal{Y} , and for all s, t, r in \mathbb{N} ,

$$\delta''(y) \triangleq \sup_{\theta, \theta'} \max_{1 \leq k, l \leq p} \left[\frac{\operatorname{ess\,sup}_x |\partial_{k, l}^2 g_\theta(y|x)|}{\operatorname{ess\,inf}_x g_{\theta'}(y|x)} \right],$$

$$\Delta_s'' \triangleq \operatorname{ess\,sup}_x \int [\delta''(y)]^s g_{\theta^*}(y|x) d\nu(y),$$

$$(\Delta\Delta'')_{s, t} \triangleq \operatorname{ess\,sup}_x \int [\delta(y)]^s [\delta''(y)]^t g_{\theta^*}(y|x) d\nu(y),$$

$$(\Delta\Delta'\Delta'')_{s, t, r} \triangleq \operatorname{ess\,sup}_x \int [\delta(y)]^s [\delta'(y)]^t [\delta''(y)]^r g_{\theta^*}(y|x) d\nu(y).$$

Assumption 9. There exists some $s > 1$ such that Δ_s'' is finite. Moreover, the quantities Δ_{11} , $(\Delta\Delta')_{11, 1}$, $(\Delta\Delta')_{10, 2}$, $(\Delta\Delta')_{9, 3}$, $(\Delta\Delta')_{8, 4}$, $(\Delta\Delta'')_{7, 1}$, $(\Delta\Delta'\Delta'')_{7, 1, 1}$ and $(\Delta\Delta'\Delta'')_{6, 2, 1}$ are supposed to be finite.

Proposition 5. Under Assumptions 1 and 7–9 and the assumption that the maximum likelihood estimator is consistent (see Theorem 2), there exists a matrix J_{θ^*} such that

$$n^{-1/2} \nabla l_n(\theta^*, \zeta) \rightarrow N(0, J_{\theta^*}) \text{ weakly under } P^*.$$

Moreover, $J_{\theta^*} = \lim_{n \rightarrow \infty} (1/n) E^*(\nabla l_n(\theta^*, \zeta)' \cdot \nabla l_n(\theta^*, \zeta))$.

This result extends Theorem 3.1 in Jensen and Petersen (1999) to the non-stationary case, and its proof is given in Section D.1. The second-order Taylor expansion of the log-likelihood also involves its Hessian matrix. Differentiating (35), we obtain $\{\nabla^2 f_{\theta, n}^\zeta\}$ as a functional autoregressive chain:

$$\nabla^2 f_{\theta, n+1}^\zeta \triangleq \alpha_1(Y_n, f_{\theta, n}^\zeta, \nabla f_{\theta, n}^\zeta, \nabla^2 f_{\theta, n}^\zeta; \theta).$$

As we have done with Z_n and \tilde{Z}_n , we now consider $\check{Z}_n = (X_n, Y_n, F_n, G_n, H_n)$, an extended Markov chain on a space \mathbf{E}'' defined by $\mathbf{E}'' = \mathbf{K} \times \mathcal{Y} \times \mathbf{S}^+ \times \Sigma \times \Sigma_2$, where

$$\Sigma_2 = \{\tau = (\tau_{k, l}); \tau_{k, l} \in L_1(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma) \text{ and } \forall v \in \mathbf{K}, \tau(v) \text{ is symmetric and non-negative}\}.$$

(The matrix $M = (M_{i, j})_{1 \leq i, j \leq n}$ is said to be non-negative if for all $(a_i)_{1 \leq i \leq n}$ in \mathbb{R} , we have $\sum_{1 \leq i, j \leq n} a_i M_{i, j} a_j \geq 0$, with equality if and only if $a_i = 0$ for all $1 \leq i \leq n$.) The space Σ_2 is equipped with the topology induced by the L_1 -norm and $\mathcal{B}(\mathbf{E}'')$ will denote the Borel σ -field on \mathbf{E}'' induced by the product topology. The Markov chain \check{Z}_n is defined by its initial law $\check{\lambda}$

and its transition kernel $\check{\Pi}_\theta$: for all (x, y, f, σ, τ) in \mathbf{E}'' and $\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm} \times \mathcal{A}_{\pm'}$ in $\mathcal{B}(\mathbf{K}) \times \mathcal{B}(\mathcal{Y}) \times \mathcal{B}(\mathbf{S}^+) \times \mathcal{B}(\boldsymbol{\Sigma}) \times \mathcal{B}(\boldsymbol{\Sigma}_2)$,

$$\begin{aligned} & \check{\Pi}_\theta[(x, y, f, \sigma, \tau); \mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm} \times \mathcal{A}_{\pm'}] \\ &= P[(X_{n+1}, Y_{n+1}, F_{n+1}, G_{n+1}, H_{n+1}) \in \mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm} \times \mathcal{A}_{\pm'} \\ & \quad | (X_n, Y_n, F_n, G_n, H_n) = (x, y, f, \sigma, \tau)] \\ &= \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} q_{\theta^*}(x, x') g_{\theta^*}(y'|x') d\gamma(x') d\nu(y') \\ & \quad \times \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\Phi_1(y, f; \theta)) \mathbb{1}_{\mathcal{A}_{\pm}}(\Psi_1(y, f, \sigma; \theta)) \mathbb{1}_{\mathcal{A}_{\pm'}}(\alpha_1(y, f, \sigma, \tau; \theta)). \end{aligned}$$

We will consider the particular initial distribution $\check{\lambda}(\zeta)$ defined for each pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ by

$$\begin{aligned} & \check{\lambda}(\zeta)(\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}} \times \mathcal{A}_{\mathcal{F}} \times \mathcal{A}_{\pm} \times \mathcal{A}_{\pm'}) \\ &= \int_{\mathcal{A}_{\mathcal{X}} \times \mathcal{A}_{\mathcal{Y}}} \pi^*(x) g_{\theta^*}(y|x) d\gamma(x) d\nu(y) \times \mathbb{1}_{\mathcal{A}_{\mathcal{F}}}(\zeta) \mathbb{1}_{\mathcal{A}_{\pm}}(0) \mathbb{1}_{\mathcal{A}_{\pm'}}(0). \end{aligned}$$

We denote by $\check{P}_{\theta, \check{\lambda}(\zeta)}$ the law of $\{\check{Z}_n\}_n$ induced by $\check{\lambda}(\zeta)$ and $\check{\Pi}_\theta$, and by $\check{E}_{\theta, \check{\lambda}(\zeta)}$ the corresponding expectation. In Section D.3, we prove the exponential lack of memory of the initial condition for this chain, its ergodicity, and the convergence of normalized sums of some functions of this chain. Then we are able to prove the convergence of the Hessian of the normalized log-likelihood.

Proposition 6. *Under Assumptions 1 and 7–9, there exists a matrix I_θ such that*

$$\frac{1}{n} \nabla^2 l_n(\theta, \zeta) \rightarrow -I_\theta \quad P^* \text{-a.s.}$$

Moreover, I_{θ^*} is the Fisher information matrix, that is,

$$I_{\theta^*} = -\lim_{n \rightarrow \infty} \frac{1}{n} E^*(\nabla^2 l_n(\theta^*, \zeta)) = \lim_{n \rightarrow \infty} \frac{1}{n} E^*(\nabla l_n(\theta^*, \zeta)^T \nabla l_n(\theta^*, \zeta)) = J_{\theta^*},$$

where superscript T denotes matrix transposition.

Write, for all $\eta > 0$ and y in \mathcal{Y} ,

$$\begin{aligned}\omega^{q'}(\eta) &\triangleq \sup_{\|\theta - \theta^*\| < \eta} \max_{1 \leq k \leq p} \int \text{ess sup}_u |\partial_k q_\theta(u, v) - \partial_k q_{\theta^*}(u, v)| d\gamma(v), \\ \omega^{q''}(\eta) &\triangleq \sup_{\|\theta - \theta^*\| < \eta} \max_{1 \leq k, l \leq p} \int \text{ess sup}_u |\partial_{k,l}^2 q_\theta(u, v) - \partial_{k,l}^2 q_{\theta^*}(u, v)| d\gamma(v), \\ \omega^{g''}(y, \eta) &\triangleq \sup_{\|\theta - \theta^*\| < \eta} \max_{1 \leq k, l \leq p} \text{ess sup}_u |\partial_{k,l}^2 g_\theta(y|u) - \partial_{k,l}^2 g_{\theta^*}(y|u)|, \\ \Delta'''(\eta) &\triangleq \int \frac{\omega^{g''}(y, \eta)}{\inf_\theta \text{ess sup}_x g_\theta(y|x)} g_{\theta^*}(y|x) d\nu(y).\end{aligned}$$

Assumption 10. $\lim_{\eta \rightarrow 0} \omega^{q'}(\eta) = \lim_{\eta \rightarrow 0} \omega^{q''}(\eta) = 0$. Furthermore, there exists some $\eta_0 > 0$ such that $\Delta'''(\eta_0)$ is finite.

We finally prove the following proposition.

Proposition 7. Under Assumptions 1, 5–6 and 9–10, we have, for any k, l in $\{1, \dots, p\}$,

$$\lim_{\eta \rightarrow 0} \mathbb{E}^* \left(\sup_{\|\theta - \theta^*\| \leq \eta} \frac{1}{n} |\partial_{k,l}^2 l_n(\theta, \zeta) - \partial_{k,l}^2 l_n(\theta^*, \zeta)| \right) = 0.$$

Remark 4. This proposition ensures that $\nabla[\lim_{n \rightarrow \infty} (1/n) \nabla l_n(\theta, \zeta)]_{\theta = \theta^*} = -I_{\theta^*}$ and that $\nabla[\Lambda_\theta(h_\theta)]_{\theta = \theta^*} = 0$.

These properties are sufficient to obtain the asymptotic normality of the maximum likelihood estimator using the following assumption.

Assumption 11. I_{θ^*} is an invertible matrix.

Theorem 3. Under Assumptions 7–11 and the assumption that $\hat{\theta}_n(\zeta)$ is consistent,

$$n^{1/2}(\hat{\theta}_n - \theta^*) \rightarrow N(0, I_{\theta^*}^{-1}) \text{ weakly under } P^*.$$

Remark 5. Combining Propositions 6 and 7 with Theorem 3, we obtain a confidence interval for θ^* . Moreover, we also have the local asymptotic normality conditions for our model, and the maximum likelihood estimator is locally asymptotically minimax (see Bickel *et al.* 1993).

Appendix A. Exponential memorylessness of the prediction filter

Proof of Proposition 1. Only the main steps of the proof will be described here. It is adapted from Mevel (1997) and uses results from Seneta (1981). The original idea comes from work

by Arapostathis and Marcus (1990). Proofs which are not easily deduced from the finite state-space case are described in full.

Step 1. Let p be a transition density function on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \gamma)$. Define

$$\tau_1(p) \triangleq \frac{1}{2} \operatorname{ess\,sup}_{x, x'} \int_v |p(x, v) - p(x', v)| d\gamma(v)$$

and, for all u, v in \mathbf{K} , for all f measurable and non-negative, write, for $n \geq m$,

$$\begin{aligned} A_{n,m}^\theta(u, v) &\triangleq \int_{u_n} q_\theta(u_n, v) g_\theta(Y_n | u_n) \cdots \int_{u_{m+1}} q_\theta(u_{m+1}, u_{m+2}) g_\theta(Y_{m+1} | u_{m+1}) \\ &\quad \times q_\theta(u, u_{m+1}) g_\theta(Y_m | u) d\gamma(u_{m+1}) \cdots d\gamma(u_n), \\ N(A_{n,m}^\theta)(u) &\triangleq \int_v A_{n,m}^\theta(u, v) d\gamma(v), \\ N(A_{n,m}^\theta f) &\triangleq \int_u N(A_{n,m}^\theta)(u) f(u) d\gamma(u) = \int_{u,v} A_{n,m}^\theta(u, v) f(u) d\gamma(u) d\gamma(v). \end{aligned}$$

Lemma 2.

$$\begin{aligned} \|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \\ \leq \tau_1 \left(\frac{A_{n,m}^\theta}{N(A_{n,m}^\theta)} \right) \left\| \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f)} f - \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f')} f' \right\|_1. \end{aligned}$$

Proof. With the previous definitions, we may write

$$\begin{aligned} \|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \\ = \int_v \left| \int_u \frac{A_{n,m}^\theta(u, v)}{N(A_{n,m}^\theta)(u)} \left[\frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f)} f(u) - \frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f')} f'(u) \right] d\gamma(u) \right| d\gamma(v). \end{aligned}$$

We then use the following general result. Let Ψ be a measurable function such that $\int_u \Psi(u) d\gamma(u) = 0$. Then we have $\int \Psi^+ = \int \Psi^- = \frac{1}{2} \int |\Psi|$, where $\Psi^+ = \max(\Psi, 0)$, $\Psi^- = \max(-\Psi, 0)$. We may also write

$$\Psi(x) = \int a(s, x) d\gamma(s) - \int a(x, t) d\gamma(t), \quad \text{where } a(s, t) = \frac{\Psi^-(s) \Psi^+(t)}{\int \Psi^+}.$$

Note that $\int \int a(s, t) d\gamma(s) d\gamma(t) = \frac{1}{2} \int |\Psi(x)| d\gamma(x)$. Applying this result to the function

$$\Psi(u) = \frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f)} f(u) - \frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f')} f'(u)$$

(with mean zero), we obtain

$$\begin{aligned}
& \int_v \left| \int_u \frac{A_{n,m}^\theta(u, v)}{N(A_{n,m}^\theta)(u)} \Psi(u) d\gamma(u) \right| d\gamma(v) \\
&= \int_v \left| \int_u \frac{A_{n,m}^\theta(u, v)}{N(A_{n,m}^\theta)(u)} \left(\int_s a(s, u) d\gamma(s) - \int_t a(u, t) d\gamma(t) \right) d\gamma(u) \right| d\gamma(v) \\
&= \int_v \left| \int_{x,y} \frac{A_{n,m}^\theta(x, v) a(y, x)}{N(A_{n,m}^\theta)(x)} d\gamma(x) d\gamma(y) - \int_{x,y} \frac{A_{n,m}^\theta(y, v) a(y, x)}{N(A_{n,m}^\theta)(y)} d\gamma(x) d\gamma(y) \right| d\gamma(v) \\
&\leq \int_{v,x,y} \left| \frac{A_{n,m}^\theta(x, v)}{N(A_{n,m}^\theta)(x)} - \frac{A_{n,m}^\theta(y, v)}{N(A_{n,m}^\theta)(y)} \right| a(y, x) d\gamma(x) d\gamma(y) d\gamma(v) \\
&\leq \frac{1}{2} \left(\operatorname{ess\,sup}_{x,y} \int_v \left| \frac{A_{n,m}^\theta(x, v)}{N(A_{n,m}^\theta)(x)} - \frac{A_{n,m}^\theta(y, v)}{N(A_{n,m}^\theta)(y)} \right| d\gamma(v) \right) \\
&\quad \times \int_u \left| \frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f)} f(u) - \frac{N(A_{n,m}^\theta)(u)}{N(A_{n,m}^\theta f')} f'(u) \right| d\gamma(u) \\
&= \tau_1 \left(\frac{A_{n,m}^\theta}{N(A_{n,m}^\theta)} \right) \left\| \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f)} f - \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f')} f' \right\|_1,
\end{aligned}$$

which concludes the proof of Lemma 2. \square

Step 2. This consists in bounding from above the last term of the right-hand side of the previous inequality.

Lemma 3.

$$\left\| \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f)} f - \frac{N(A_{n,m}^\theta)}{N(A_{n,m}^\theta f')} f' \right\|_1 \leq 2\omega^\theta \|f - f'\|_1, \quad \text{where } \omega^\theta = \sup \frac{N(A_{n,m}^\theta f)}{N(A_{n,m}^\theta g)}.$$

The supremum is taken over the set of measurable and non-negative functions f and g such that $\int f d\gamma = \int g d\gamma = 1$.

Step 3. We prove the following result:

Lemma 4. $\omega^\theta \leq \epsilon_\theta^{-1} \delta_\theta(Y_m)$.

The proofs of Lemmas 3 and 4 are straightforward adaptations from Mevel (1997, Propositions A.1 and A.2), and are omitted here for brevity.

Then we conclude that

$$\|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \leq 2\epsilon_\theta^{-1} \delta_\theta(Y_m) \times \tau_1 \left(\frac{A_{n,m}^\theta}{N(A_{n,m}^\theta)} \right) \times \|f - f'\|_1.$$

Let us now examine the behaviour of $\tau_1(A_{n,m}^\theta/N(A_{n,m}^\theta))$. We wish to proceed by analogy between $A_{n,m}^\theta$ and a product of random matrices (which appears in finite state-space models): this technique will bring out the term $(1 - \epsilon_\theta)^{n-m+1}$. We obtain an upper bound for τ_1 by using another coefficient, denoted by τ_B (Birkhoff's coefficient), which does not depend on the normalization $N(A_{n,m}^\theta)$.

Step 4. All these results are generalizations of those obtained by Seneta (1981, Theorems 3.12 and 3.13, pp. 106–110) and the proofs will be omitted. Let us define Birkhoff's coefficient by

$$\tau_B(p) = \sup \frac{d(p^*f, p^*g)}{d(f, g)},$$

where

$$d(f, g) = \operatorname{ess\,sup}_{x,x'} \log \left[\frac{f(x)g(x')}{g(x)f(x')} \right],$$

$$p^*f(v) = \int_u p^*(v, u)f(u)d\gamma(u) = \int_u p(u, v)f(u)d\gamma(u).$$

The first supremum is taken over the set of measurable and non-negative functions f and g such that $f \neq \lambda g$ ($\lambda \in \mathbb{R}$), and p^* is the transpose density function associated with p (i.e. $p^*(x, y) = p(y, x)$). With this definition, τ_B does not depend on the normalization, and is, moreover, submultiplicative. We now give another expression for τ_B , and then the upper bound of τ_1 by τ_B .

Lemma 5.

$$\tau_B(p) = \frac{1 - \alpha(p)^{1/2}}{1 + \alpha(p)^{1/2}} \geq \tau_1(p), \quad \text{where } \alpha(p) = \operatorname{ess\,inf}_{x,y,z,t} \frac{p(x, z)p(y, t)}{p(x, t)p(y, z)} = \alpha(p^*).$$

We now apply these results with $p = A_{n,m}^\theta/N(A_{n,m}^\theta)$. We have

$$\tau_1 \left(\frac{A_{n,m}^\theta}{N(A_{n,m}^\theta)} \right) \leq \tau_B \left(\frac{A_{n,m}^\theta}{N(A_{n,m}^\theta)} \right).$$

It follows from the definition that $\alpha(A_{n,m}^\theta/N(A_{n,m}^\theta)) = \alpha(A_{n,m}^\theta)$, and then $\tau_B(A_{n,m}^\theta/N(A_{n,m}^\theta)) = \tau_B(A_{n,m}^\theta)$ by application of Lemma 5. Since τ_B is submultiplicative,

$$\tau_B(A_{n,m}^\theta) \leq \prod_{k=m}^n \tau_B[q_\theta(\cdot, \cdot)g_\theta(Y_k|\cdot)].$$

Finally, again using Lemma 5, $\tau_B[q_\theta(\cdot, \cdot)g_\theta(Y_k|\cdot)] = \tau_B(q_\theta)$, and, under Assumption 1,

$$\alpha(q_\theta) = \operatorname{ess\,inf}_{x,y,z,t} \frac{q_\theta(x,z)q_\theta(y,t)}{q_\theta(x,t)q_\theta(y,z)} \geq \epsilon_\theta^2.$$

Thus $\tau_B(q_\theta) \leq 1 - \epsilon_\theta$, which finally implies that $\tau_B(A_{n,m}^\theta) \leq (1 - \epsilon_\theta)^{n-m+1}$. \square

Appendix B. Geometric ergodicity of the extended chain

Proof of Proposition 2. Let h be in $\operatorname{Lip}(\mathbf{E})$, and $z = (x, y, f, f^*)$, $z' = (x', y', f', (f^*)')$ be in \mathbf{E} . In order to obtain an upper bound for $|\Pi_\theta^n h(z) - \Pi_\theta^n h(z')|$, use the decomposition

$$\begin{aligned} \Pi_\theta^n h(z) - \Pi_\theta^n h(z') &= [\Pi_\theta^n h(x, y, f, f^*) - \Pi_\theta^n h(x, y', f', (f^*)')] \\ &\quad + [\Pi_\theta^n h(x, y', f', (f^*)') - \Pi_\theta^n h(x', y', f', (f^*)')]. \end{aligned}$$

Denote by A and B the first and second terms on the right-hand side of this equation, respectively. A and B will be treated separately.

For $i \leq j$, write $y_j^i = (y_j, \dots, y_i)$ and, with an abuse of notation, $d\gamma(x_j^i) \triangleq d\gamma(x_i) \dots d\gamma(x_j)$ and $d\nu(y_j^i) \triangleq d\nu(y_i) \dots d\gamma(y_j)$. Write also $\{y_j^i, y\} \triangleq (y_j, \dots, y_i, y)$. Then

$$\begin{aligned} A &= \int_{x_1 \dots y_n} [h(x_n, y_n, \Phi_n(\{y_{n-1}^1, y\}, f; \theta), \Phi_n(\{y_{n-1}^1, y\}, f^*; \theta^*)) \\ &\quad - h(x_n, y_n, \Phi_n(\{y_{n-1}^1, y'\}, f'; \theta), \Phi_n(\{y_{n-1}^1, y'\}, (f^*)'; \theta^*))] \\ &\quad \times q_{\theta^*}(x, x_1) \cdots q_{\theta^*}(x_{n-1}, x_n) g_{\theta^*}(y_1 | x_1) \cdots g_{\theta^*}(y_n | x_n) d\gamma(x_n^1) d\nu(y_n^1). \end{aligned}$$

Then, using $\operatorname{Lip}(\mathbf{E})$ as in Definition 1, and the fact that, for all θ in Θ , $\Phi_n(y_{n-1}^1, y, f; \theta) = \Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta)$, we have

$$\begin{aligned} |A| &\leq \int_{x_1 \dots y_n} \operatorname{lip}(h, x_n, y_n) [\|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', f'; \theta); \theta)\|_1 \\ &\quad + \|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f^*; \theta^*); \theta^*) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', (f^*)'; \theta^*); \theta^*)\|_1] \\ &\quad \times q_{\theta^*}(x, x_1) \cdots q_{\theta^*}(x_{n-1}, x_n) g_{\theta^*}(y_1 | x_1) \cdots g_{\theta^*}(y_n | x_n) d\gamma(x_n^1) d\nu(y_n^1). \end{aligned}$$

Exponential lack of memory of the prediction filter yields

$$\begin{aligned} &\|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', f'; \theta); \theta)\|_1 \\ &\leq 2\epsilon_\theta^{-1} \delta_\theta(y_1) (1 - \epsilon_\theta)^{n-1} \|\Phi_1(y, f; \theta) - \Phi_1(y', f'; \theta)\|_1 \\ &\leq 4\epsilon_\theta^{-1} \delta_\theta(y_1) (1 - \epsilon_\theta)^{n-1}. \end{aligned}$$

The same inequality holds at the point θ^* . Integrating the above expression and using Assumption 1 yields

$$\begin{aligned}
|A| &\leq 4\text{lip}(h)[\epsilon_{\theta}^{-1}\Delta_1(1-\epsilon_{\theta})^{n-1} + \epsilon_{\theta^*}^{-1}\Delta_1(1-\epsilon_{\theta^*})^{n-1}] \\
&\leq 8\text{lip}(h)\epsilon^{-1}\Delta_1(1-\epsilon)^{n-1}.
\end{aligned} \tag{37}$$

Remark 6. We may prove along the same lines that if h is in $\text{Lip}(\mathbf{E})$ then, for any integer n , $\Pi_{\theta}^n h$ also belongs to $\text{Lip}(\mathbf{E})$, with $\text{lip}(\Pi_{\theta}^n h, x, y) \leq \text{lip}(h)\delta_{\theta}(y)$ and $\mathbf{k}(\Pi_{\theta}^n h, x, y) \leq \mathbf{k}(h)$.

We now turn our attention to B , given by

$$\begin{aligned}
B &= \int_{x_1 \cdots y_n} h(x_n, y_n, \Phi_n(\{y_{n-1}^1, y'\}, f'; \theta), \Phi_n(\{y_{n-1}^1, y'\}, (f^*)'; \theta^*)) \\
&\quad \times [q_{\theta^*}(x, x_1) - q_{\theta^*}(x', x_1)]q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
&\quad \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1).
\end{aligned}$$

Let $m = \lceil n/2 \rceil$ and F, F^* be fixed functions in \mathbf{S}^+ . Note that

$$\begin{aligned}
&h(x_n, y_n, \Phi_n(\{y_{n-1}^1, y'\}, f'; \theta), \Phi_n(\{y_{n-1}^1, y'\}, (f^*)'; \theta^*)) \\
&= h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, \Phi_m(\{y_{m-1}^1, y'\}, f'; \theta); \theta), \\
&\quad \Phi_{n-m}(y_{n-1}^m, \Phi_m(\{y_{m-1}^1, y'\}, (f^*)'; \theta^*); \theta^*)) \\
&\quad - h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, F; \theta), \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*)) \\
&\quad + h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, F; \theta), \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*)).
\end{aligned}$$

Insert this form into B and use the fact that h is in $\text{Lip}(\mathbf{E})$ for the first part and take the integral with respect to y_{m-1}^1 and x_{m-1}^1 in the second; this yields

$$\begin{aligned}
|B| &\leq \int_{x_1 \cdots y_n} \text{lip}(h, x_n, y_n) \times (|\Phi_{n-m}(y_{n-1}^m, \Phi_m(\{y_{m-1}^1, y'\}, f'; \theta); \theta) - \Phi_{n-m}(y_{n-1}^m, F; \theta)| \\
&\quad + |\Phi_{n-m}(y_{n-1}^m, \Phi_m(\{y_{m-1}^1, y'\}, (f^*)'; \theta^*); \theta^*) - \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*)|) \\
&\quad \times |q_{\theta^*}(x, x_1) - q_{\theta^*}(x', x_1)|q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
&\quad \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1) \\
&\quad + \int_{x_m \cdots x_n} \int_{y_m \cdots y_n} |h(x_n, y_n, \Phi_n(y_{n-1}^m, F; \theta), \Phi_n(y_{n-1}^m, F^*; \theta^*))| \\
&\quad \times |q_{\theta^*}^m(x, x_m) - q_{\theta^*}^m(x', x_m)|q_{\theta^*}^m(x_m, x_{m+1}) \cdots q_{\theta^*}^m(x_{n-1}, x_n) \\
&\quad \times g_{\theta^*}^m(y_m|x_m) \cdots g_{\theta^*}^m(y_n|x_n) d\gamma(x_n^m) d\nu(y_n^m).
\end{aligned}$$

Using Proposition 1 and inequality (9) yields

$$\begin{aligned}
 |B| &\leq \int_{x_1 \cdots y_n} \text{lip}(h, x_n, y_n) \times 2\epsilon^{-1}(1 - \epsilon)^{n-m} \\
 &\quad \times (\|\Phi_m(\{y_{m-1}^1, y'\}, f'; \theta) - F\|_1 \delta_\theta(y_m) + \|\Phi_m(\{y_{m-1}^1, y'\}, (f^*)'; \theta^*) - F^*\|_1 \delta_{\theta^*}(y_m)) \\
 &\quad \times (q_{\theta^*}(x, x_1) + q_{\theta^*}(x', x_1))q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n)g_{\theta^*}(y_1|x_1) \cdots \\
 &\quad \quad g_{\theta^*}(y_n|x_n)d\gamma(x_n^1)d\nu(y_n^1) \\
 &\quad + \int_{x_m \cdots x_n} \int_{y_m \cdots y_n} k(h, x_n, y_n)\rho_0^m q_{\theta^*}(x_m, x_{m+1}) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
 &\quad \times g_{\theta^*}(y_m|x_m) \cdots g_{\theta^*}(y_n|x_n)d\gamma(x_n^m)d\nu(y_n^m),
 \end{aligned}$$

that is,

$$|B| \leq 16 \times \text{lip}(h)\epsilon^{-1}\Delta_1(1 - \epsilon)^{n-m} + k(h)\rho_0^m. \tag{38}$$

Inequalities (37) and (38) complete the proof. □

Proof of Corollaries 1 and 2. The existence of $\Lambda_\theta(h)$ follows from Proposition 2, using the same ideas as in Sunyach (1975, Theorem 1). Then we can easily check that the function defined on \mathbf{E} by

$$V(z) = \sum_{n \geq 0} [\Pi_\theta^n h(z) - \Lambda_\theta(h)] \tag{39}$$

is a solution to the Poisson equation. As the space \mathbf{E} is not locally compact, we cannot use the Riesz representation theorem to show that $\Lambda_\theta(h)$ is the expectation of h under a suitable measure. Let us show that the family of probability measures $(\Pi_\theta^n(z, \cdot))_{n \geq 0}$ is asymptotically tight (for definition see, van der Vaart and Wellner 1996, p. 20), using the same basic ideas as in Sunyach (1975, Theorem 1).

Let $d(\cdot, \cdot)$ be a distance in \mathbf{E} . For \mathbf{D} a compact set of \mathbf{E} and $\delta > 0$, define $\varphi_{\delta, \mathbf{D}}(z) \triangleq \inf(1, (1/\delta)d(z, \mathbf{D}))$. It is straightforward to show that $\varphi_{\delta, \mathbf{D}}(z)$ is in $\text{Lip}(\mathbf{E})$, with $\text{lip}(\varphi_{\delta, \mathbf{D}}) \leq \delta^{-1}$ and $k(\varphi_{\delta, \mathbf{D}}) \leq 1$. Moreover, for any set \mathcal{A} , write $\mathcal{A}_\delta \triangleq \{z \in \mathbf{E}; d(z, \mathcal{A}) \leq \delta\}$ for the δ -enlargement of \mathcal{A} .

Let $\epsilon, \delta > 0$ and z in \mathbf{E} . There exists p sufficiently large such that $C[\delta^{-1} + 1] \times \rho^p(1 - \rho)^{-1} < \epsilon/2$ (where C is the constant appearing in (12)). Now, choose a compact \mathbf{D}^ϵ such that $\Pi_\theta^n[z, (\mathbf{D}^\epsilon)^c] < \epsilon/2$ for $k = 1, \dots, p$, which is possible since any probability measure in a separable and complete space is tight (the superscript c denotes the complement in the space \mathbf{E}). Note that we have $\mathbb{1}_{((D^\epsilon)_\delta)^c} \leq \varphi_{\delta, \mathbf{D}^\epsilon} \leq \mathbb{1}_{(D^\epsilon)^c}$, so we obtain, for any $n > p$,

$$\begin{aligned}
 \Pi_\theta^n[z, ((\mathbf{D}^\epsilon)_\delta)^c] &\leq \Pi_\theta^n(\varphi_{\delta, \mathbf{D}^\epsilon})(z) \\
 &\leq \Pi_\theta^p(\varphi_{\delta, \mathbf{D}^\epsilon})(z) + C \left[\frac{1}{\delta} + 1 \right] \frac{\rho^p}{1 - \rho} \\
 &\leq \Pi_\theta^p[z, (\mathbf{D}^\epsilon)^c] + \frac{\epsilon}{2} \\
 &\leq \epsilon.
 \end{aligned} \tag{40}$$

So $(\Pi_\theta^n(z, \cdot))_{n \geq 0}$ is asymptotically tight. Hence, by Prohorov's theorem (van der Vaart and Wellner 1996, p. 21), there exists a probability measure λ_θ and a subsequence $(n_i)_{i \geq 0}$ such that $\Pi_\theta^{n_i}(z, \cdot)$ converges weakly to λ_θ . This proves that for any continuous and bounded function h in $\text{Lip}(\mathbf{E})$, $\Lambda_\theta(h) = \int h(z) d\lambda_\theta(z)$. Using the result of Corollary 1, we obtain that, for any continuous and bounded function h in $\text{Lip}(\mathbf{E})$,

$$\lambda(\zeta) \Pi_\theta^n h \stackrel{\Delta}{=} \int \Pi_\theta^n h(z) \lambda(\zeta)(dz) = E_{\theta, \lambda(\zeta)}(h(Z_n)) \xrightarrow{n \rightarrow \infty} \int h(z) d\lambda_\theta(z) = \Lambda_\theta(h). \tag{41}$$

Then, using the fact that, for any h in $\text{Lip}(\mathbf{E})$, $\Pi_\theta h$ is still in $\text{Lip}(\mathbf{E})$ (see Remark 6), λ_θ is the unique invariant probability measure associated with the kernel Π_θ .

Now let h be a function in $\text{Lip}(\mathbf{E})$ (h is not necessarily continuous and bounded). Define

$$G_h(x, y, f, f^*) \stackrel{\Delta}{=} \int h(x_0, y_0, f, f^*) g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \tag{42}$$

It can be easily checked that G_h is a bounded and continuous function in $\text{Lip}(\mathbf{E})$. Applying the previous result to the function G_h , we obtain

$$\begin{aligned}
 \lim_n \lambda(\zeta) \Pi_\theta^n G_h &= \int G_h(z) d\lambda_\theta(z) \\
 &= \int h(x_0, y_0, f, f^*) g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*).
 \end{aligned} \tag{43}$$

Using Remark 2,

$$\begin{aligned}
 \lambda(\zeta) \Pi_\theta^n h &= E_{\theta, \lambda(\zeta)}(h(Z_n)) \\
 &= E^*[E^*(h(X_n, Y_n, f_{\theta, n}^\zeta, f_{\theta^*, n}^{\tau^*}) | Y_{n-1}^0)] \\
 &= E^*(G_h(X_n, Y_n, f_{\theta, n}^\zeta, f_{\theta^*, n}^{\tau^*})) \\
 &= E_{\theta, \lambda(\zeta)}(G_h(Z_n)) \\
 &= \lambda(\zeta) \Pi_\theta^n G_h.
 \end{aligned} \tag{44}$$

But we already know by (41) that

$$\lim_n \lambda(\zeta) \Pi_\theta^n h = \Lambda_\theta(h). \tag{45}$$

Combining with (43), (44) and (45),

$$\Lambda_\theta(h) = \int h(x_0, y_0, f, f^*) g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*).$$

But we have

$$g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*) = \lambda_\theta(dx_0, dy_0, df, df^*),$$

as these measures are entirely identified by their integrals on Lipschitz and bounded functions on \mathbf{E} . Then we obtain, for all h in $\text{Lip}(\mathbf{E})$,

$$\Lambda_\theta(h) = \int h(z) d\lambda_\theta(z).$$

□

Appendix C. Consistency of the maximum likelihood estimator

Proof of Proposition 3. (i) Denote by V by the solution of the Poisson equation associated to the function h (see Corollary 1) and by \mathcal{F}_n the σ -algebra generated by $\{Z_0, \dots, Z_n\}$. It is easy to show, using inequality (12) for $n \geq 1$, that

$$|V(z)| \leq C\rho \frac{\text{lip}(h) + k(h)}{(1 - \rho)^2} + |\Lambda_\theta(h)| + k(h, x, y), \quad (46)$$

and thus there exist $s > 1$ (cf. Assumption 3) such that $\Pi_\theta |V|^s$ is bounded, which implies that $E_\theta(|V(Z_{n+1}) - \Pi_\theta V(Z_n)|^s | \mathcal{F}_n)$ is bounded independently of n . The proof follows from the classical identity (Meyn and Tweedie 1993, 17.4.3)

$$\begin{aligned} & \frac{1}{n} \sum_{m=0}^{n-1} h(Z_m) \\ &= \Lambda_\theta(h) + \frac{1}{n} \sum_{m=1}^n [V(Z_m) - \Pi_\theta V(Z_{m-1})] + \frac{1}{n} [\Pi_\theta V(Z_0) - \Pi_\theta V(Z_n) + h(Z_0) - h(Z_n)] \end{aligned}$$

by applying the corollary of Chow's theorem (Hall and Heyde 1980, p. 36).

(ii) This is a direct consequence of (i) by applying Remark 2. □

Proof of Proposition 4. It is easy to show that

$$\begin{aligned} |h_\theta(x, y, f_1, f_2) - h_\theta(x, y, f'_1, f'_2)| &= \left| \log \left(1 + \frac{\int_u g_\theta(y|u)(f_1 - f'_1)(u) d\gamma(u)}{\int_u g_\theta(y|u) f'_1(u) d\gamma(u)} \right) \right| \\ &\leq \delta_\theta(y) \|f_1 - f'_1\|_1 \end{aligned} \quad (47)$$

and

$$|h_\theta(x, y, f_1, f_2)| \leq k_\theta(y), \quad (48)$$

which proves that h is in $\text{Lip}(\mathbf{E})$ and verifies equation (15), using Assumptions 2 and 3. \square

Proof of Lemma 1. Let $\eta > 0$ and $\|\theta - \theta'\| \leq \eta$. By Proposition 1,

$$\begin{aligned} & \|\Phi_{n+1}(Y_n^0, \xi; \theta) - \Phi_{n+1}(Y_n^0, \xi; \theta')\|_1 \\ &= \|\Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta'); \theta')\|_1 \\ &\leq \|\Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta'); \theta)\|_1 \\ &\quad + \|\Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta'); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \xi; \theta'); \theta')\|_1 \\ &\leq 2c_\theta^{-1} \delta_\theta(Y_1) \rho^n \|\Phi_1(Y_0, \xi; \theta) - \Phi_1(Y_0, \xi; \theta')\|_1 + \sup_f \|\Phi_n(Y_n^1, f; \theta) - \Phi_n(Y_n^1, f; \theta')\|_1. \end{aligned} \quad (49)$$

It is straightforward to show that

$$\begin{aligned} & \|\Phi_1(Y_0, \xi; \theta) - \Phi_1(Y_0, \xi; \theta')\|_1 \\ &\leq \left\| \frac{\int (q_\theta(u, \cdot) - q_{\theta'}(u, \cdot)) g_\theta(Y_0|u) \xi(u) d\gamma(u)}{\int g_\theta(Y_0|u) \xi(u) d\gamma(u)} \right\|_1 \\ &\quad + \left\| \frac{\int q_{\theta'}(u, \cdot) g_\theta(Y_0|u) \xi(u) d\gamma(u)}{\int g_\theta(Y_0|u) \xi(u) d\gamma(u)} - \frac{\int q_{\theta'}(u, \cdot) g_{\theta'}(Y_0|u) \xi(u) d\gamma(u)}{\int g_{\theta'}(Y_0|u) \xi(u) d\gamma(u)} \right\|_1. \end{aligned}$$

The first term on the right-hand side of this expression is bounded from above by $\omega^q(\eta)$. The second term on the right is bounded from above by $2\delta'(Y_0, \eta)$, by breaking the difference into two terms including $\int q_{\theta'}(u, \cdot) g_\theta(Y_0|u) \xi(u) d\gamma(u) / \int g_{\theta'}(Y_0|u) \xi(u) d\gamma(u)$. Combining these two bounds with (49),

$$\begin{aligned} & \sup_\xi \|\Phi_{n+1}(Y_n^0, \xi; \theta) - \Phi_{n+1}(Y_n^0, \xi; \theta')\|_1 \\ &\leq 2c_\theta^{-1} \delta_\theta(Y_1) (\omega^q(\eta) + 2\delta'(Y_0, \eta)) \rho^n \|\theta - \theta'\| + \sup_\xi \|\Phi_n(Y_n^1, \xi; \theta) - \Phi_n(Y_n^1, \xi; \theta')\|_1. \end{aligned}$$

A straightforward recurrence completes the proof. \square

Appendix D. Asymptotic normality of the maximum likelihood estimator

D.1. Exponential memorylessness of the gradient of the prediction filter and consequences

Iterating equation (35) at order m , $m \leq n$, yields, for all v in \mathbf{K} ,

$$\nabla f_{\theta, n+1}^{\zeta}(v) = \Psi_{n-m+1}(Y_n^m, f_{\theta, m}^{\zeta}, \nabla f_{\theta, m}^{\zeta}; \theta)(v),$$

where Ψ_n is the vector function $(\Psi_n^k)_{1 \leq k \leq p}$ defined by

$$\begin{aligned} & \Psi_{n-m+1}^k(Y_n^m, f, \sigma; \theta)(v) \\ &= \int_{u_n} a_{\theta}(Y_n, f_n)(u_n, v) \cdots \int_{u_m} a_{\theta}(Y_m, f)(u_m, u_{m+1}) \sigma_k(u_m) d\gamma(u_m) \cdots d\gamma(u_n) \\ &+ \sum_{l=m}^{n-1} \int_{u_n} a_{\theta}(Y_n, f_n)(u_n, v) \cdots \int_{u_{l+1}} a_{\theta}(Y_{l+1}, f_{l+1})(u_{l+1}, u_{l+2}) \\ &\times U_{\theta, k}(Y_l, f_l)(u_{l+1}) d\gamma(u_{l+1}) \cdots d\gamma(u_n) + U_{\theta, k}(Y_n, f_n)(v), \end{aligned}$$

where, for all $l \leq m+1$, we use the notation $f_l \triangleq \Phi_{l-m}(Y_{l-1}^m, f; \theta)$.

In this subsection, the exponential memorylessness for Ψ of the initial conditions is proved. It is straightforward to show that, for any k in $\{1, \dots, p\}$, $U_{\theta, k}$ verifies a Lipschitz condition. For all y in \mathcal{Y} , for all f, f' in \mathbf{S}^+ ,

$$\|U_{\theta, k}(y, f) - U_{\theta, k}(y, f')\|_1 \leq \text{lip}(U_{\theta, k}, y) \|f - f'\|_1,$$

$$\|U_{\theta, k}(y, f)\|_1 \leq k(U_{\theta, k}, y),$$

where

$$\begin{aligned} \text{lip}(U_{\theta, k}, y) &= \delta_{\theta}(y)(1 + \delta_{\theta}(y)) \left(\max_{1 \leq k \leq p} \int_{x'} \text{ess sup}_x |\partial_k q_{\theta}(x, x')| d\gamma(x') \right) + \delta'_{\theta}(y)(1 + \delta_{\theta}(y)) \\ &+ 2\delta_{\theta}(y)\delta'_{\theta}(y)(1 + \delta_{\theta}(y)^2), \end{aligned}$$

$$k(U_{\theta, k}, y) = \delta_{\theta}(y) \left(\max_{1 \leq k \leq p} \int_{x'} \text{ess sup}_x |\partial_k q_{\theta}(x, x')| d\gamma(x') \right) + \delta'_{\theta}(y)(1 + \delta_{\theta}(y)).$$

Note that $\text{lip}(U_{\theta, k}) \triangleq \text{ess sup}_x \int_y \text{lip}(U_{\theta, k}, y) g_{\theta^*}(y|x) d\nu(y) < \infty$ provided that Δ_2 and $(\Delta\Delta')_{3,1}$ are finite, and $k(U_{\theta, k}) \triangleq \text{ess sup}_x \int_y k(U_{\theta, k}, y) g_{\theta^*}(y|x) d\nu(y) < \infty$ provided that Δ_1 and $(\Delta\Delta')_{11}$ are finite.

Proposition 8. *Under Assumptions 1 and 7, and for all f, f' in \mathbf{S}^+ , σ, σ' in Σ , θ in Θ , and all $1 \leq k \leq p$,*

$$\begin{aligned}
 & \|\Psi_{n-m+1}^k(Y_n^m, f, \sigma; \theta) - \Psi_{n-m+1}^k(Y_n^m, f', \sigma'; \theta)\|_1 \\
 & \leq 6\delta_\theta(Y_m)^3 \epsilon_\theta^{-1} (1 - \epsilon_\theta)^{n-m+1} [\|\sigma_k - \sigma'_k\|_1 + \|f - f'\|_1 (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1)] \\
 & \quad + 10\delta_\theta(Y_m) \epsilon_\theta^{-1} (1 - \epsilon_\theta)^{n-m+1} \|f - f'\|_1 \\
 & \quad \left[\text{lip}(U_{\theta,k}, Y_n) + \sum_{l=m}^{n-1} \text{lip}(U_{\theta,k}, Y_l) \delta_\theta(Y_{l+1})^2 + \sum_{l=m}^{n-1} \delta_\theta(Y_{l+1})^3 \mathbf{k}(U_{\theta,k}, Y_l) \right].
 \end{aligned}$$

The proof is a straightforward adaptation from Mevel (1997, Part I, Proposition 4.4) and is omitted here.

Now define a class of Lipschitz functions on \mathbf{E}' :

Definition 2. $\text{Lip}(\mathbf{E}')$ is the set of vector-valued functions $h = (h_k)_{1 \leq k \leq p}$ such that each h_k is a real-valued measurable function on \mathbf{E}' ; and for all (x, y) in $\mathbf{K} \times \mathcal{Y}$, there exist $\text{lip}'(h_k, x, y)$ and $\mathbf{k}'(h_k, x, y)$ such that, for all f, f' in \mathbf{S}^+ , for all σ, σ' in Σ ,

$$\begin{aligned}
 & |h_k(x, y, f, \sigma) - h_k(x, y, f', \sigma')| \\
 & \leq \text{lip}'(h_k, x, y) [\|\sigma_k - \sigma'_k\|_1 + \|f - f'\|_1 \times (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1)], \\
 & |h_k(x, y, f, \sigma) - h_k(x, y, f, \sigma')| \leq \mathbf{k}'(h_k, x, y) (1 + \|\sigma_k\|_1)
 \end{aligned}$$

and

$$\begin{aligned}
 \text{lip}'(h_k) &= \text{ess sup}_x \int \text{lip}'(h_k, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty, \\
 \mathbf{k}'(h_k) &= \text{ess sup}_x \int \mathbf{k}'(h_k, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty.
 \end{aligned}$$

The following analogue of Proposition 2 can be proved.

Proposition 9. Under Assumptions 1 and 7, there exist constants $C > 0$ and $\tilde{\rho} \in]\rho, 1[$ such that, for all z, z' in \mathbf{E}' , for all h in $\text{Lip}(\mathbf{E}')$, for all $n \geq 1$, and for all $1 \leq k \leq p$, we have

$$\begin{aligned}
 & |\tilde{\Pi}_\theta^n h_k(z) - \tilde{\Pi}_\theta^n h_k(z')| \leq \\
 & C(\mathbf{k}'(h_k) + \text{lip}'(h_k)) \tilde{\rho}^n (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1) (\delta_\theta(y)^2 + \delta_\theta(y')^2 + \mathbf{k}(U_{\theta,k}, y) + \mathbf{k}(U_{\theta,k}, y')).
 \end{aligned}$$

Corollary 3. Under Assumptions 1 and 7, there exist constants $C > 0$ and $\tilde{\rho} \in]\rho, 1[$, such that, for all h in $\text{Lip}(\mathbf{E}')$, there exists a family of constants $(\tilde{\Lambda}_\theta(h_k))_{1 \leq k \leq p}$, such that for all z in \mathbf{E}' , for all $n \geq 1$, and for all $1 \leq k \leq p$,

$$|\tilde{\Pi}_\theta^n h_k(z) - \tilde{\Lambda}_\theta(h_k)| \leq C(\text{lip}'(h_k) + \mathbf{k}'(h_k)) \tilde{\rho}^n (1 + \|\sigma_k\|_1) (\delta_\theta(y)^2 + \mathbf{k}(U_{\theta,k}, y)).$$

For any initial probability measure $\tilde{\lambda}$ on $(\mathbf{E}', \mathcal{B}(\mathbf{E}'))$, and for any function h in $\text{Lip}(\mathbf{E}')$,

$$\tilde{\Lambda}_\theta(h) = \lim_{n \rightarrow \infty} \tilde{\mathbb{E}}_{\theta, \tilde{\lambda}}(h(\tilde{Z}_n)), \quad (50)$$

and, for all $1 \leq k \leq p$, there exists a solution in \mathbf{E}' to the Poisson equation $(I - \tilde{\Pi}_\theta)V(\cdot) = h_k(\cdot) - \tilde{\Lambda}_\theta(h_k)$.

The proofs are similar to the proofs of Proposition 2 and its corollary, and are omitted here (note that they assume that Δ_3 is finite).

We can now prove the convergence of the gradient of the normalized log-likelihood.

D.2. Proof of Proposition 5

Differentiating the expression for the log-likelihood, we obtain, for all k in $\{1, \dots, p\}$,

$$\begin{aligned} \frac{1}{n} \partial_k l_n(\theta, \zeta) &= \frac{1}{n} \sum_{m=0}^{n-1} \frac{\int \partial_k g_\theta(Y_m|u) f_{\theta, m}^\zeta(u) d\gamma(u) + \int g_\theta(Y_m|u) \partial_k f_{\theta, m}^\zeta(u) d\gamma(u)}{\int g_\theta(Y_m|u) f_{\theta, m}^\zeta(u) d\gamma(u)} \quad P^* \text{-a.s.} \\ &= \frac{1}{n} \sum_{m=0}^{n-1} j_{\theta, k}(\tilde{Z}_m) \quad P_{\theta, \lambda(\zeta)} \text{-a.s.}, \end{aligned}$$

where

$$j_{\theta, k}(x, y, f, \sigma) \triangleq \frac{\int \partial_k g_\theta(y|u) f(u) d\gamma(u) + \int g_\theta(y|u) \sigma_k(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)}.$$

We check (using simple calculus) that j_θ is in $\text{Lip}(\mathbf{E}')$, assuming that Δ_2 and $(\Delta\Delta')_{1,1}$ are finite. (In fact, $\text{lip}'(j_\theta, x, y) = \delta'_\theta(y)(1 + \delta_\theta(y)) + \delta_\theta(y)^2$ and $\text{k}'(j_\theta, x, y) = \delta_\theta(y) + \delta'_\theta(y)$). Assuming, moreover, that, for some $s > 1$, Δ_s and Δ'_s are finite (in order to obtain a condition like (15)), we prove, just as in Proposition 4, that, for all k in $\{1, \dots, p\}$,

$$\frac{1}{n} \partial_k l_n(\theta, \zeta) \rightarrow \tilde{\Lambda}_\theta(j_{\theta, k}) \quad P^* \text{-a.s.}$$

But the score function $\tilde{\Lambda}_\theta(j_\theta)$ vanishes at θ^* : Propositions 6 and 7 show that the convergence of $(1/n)\nabla^2 l_n(\theta, \zeta)$ happens P^* -a.s. and uniformly for θ in a neighbourhood of θ^* , for some subsequence; and then for the whole family, as it is still P^* -a.s. convergent. Thus, $(1/n)\nabla l_n(\theta, \zeta)$ is uniformly convergent, vanishes at $\hat{\theta}_n(\zeta)$, and $\hat{\theta}_n(\zeta)$ converges to θ^* , which proves the result.

Let V_k^* denote the solution to the Poisson equation $(I - \tilde{\Pi}_{\theta^*})V = j_{\theta^*, k} - \tilde{\Lambda}_{\theta^*}(j_{\theta^*, k})$ (which exists by Corollary 3). We have, $\tilde{P}_{\theta^*, \lambda(\zeta)}$ -a.s.

$$\begin{aligned}
 \frac{1}{n^{1/2}} \partial_k l_n(\theta^*, \zeta) &= \frac{1}{n^{1/2}} \sum_{m=0}^{n-1} j_{\theta^*,k}(\tilde{Z}_m) - \tilde{\Lambda}_{\theta^*}(j_{\theta^*,k}) \\
 &= \frac{1}{n^{1/2}} \sum_{m=1}^n V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_{m-1}) \\
 &\quad + \frac{1}{n^{1/2}} (\tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_0) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_n) + j_{\theta^*,k}(\tilde{Z}_0) - j_{\theta^*,k}(\tilde{Z}_n)).
 \end{aligned}$$

The second term converges to zero as the quantities $\tilde{\Pi}_{\theta^*} V_k^*(\cdot)$ and $j_{\theta^*,k}$ are bounded. (Recall that an inequality like (46) still holds. In fact, $V_k^*(z) \leq C(\delta_\theta(y) + \mathbf{k}(U_{\theta,k}, y)) + \mathbf{k}(j_{\theta^*,k}, x, y)$.) Thus, for all $1 \leq k \leq p$,

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n^{1/2}} \partial_k l_n(\theta^*, \zeta) - \frac{1}{n^{1/2}} \sum_{m=1}^n (V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_{m-1})) \right] = 0 \quad \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)}\text{-a.s.}$$

Now, since $(\tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_n))_n$ is bounded, the convergence holds in $L_2(\tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)})$:

$$\tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} \left[\frac{1}{n^{1/2}} \partial_k l_n(\theta^*, \zeta) - \frac{1}{n^{1/2}} \sum_{m=1}^n V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_{m-1}) \right]^2 \xrightarrow{n \rightarrow \infty} 0. \quad (51)$$

Let u be a fixed vector of \mathbb{R}^p ; $\sum_{m=1}^n u^\top B_m \triangleq \sum_{m=1}^n u^\top (V^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V^*(\tilde{Z}_{m-1}))$ is a real-valued martingale adapted to the Borel σ -field $\{\tilde{\mathcal{L}}_n = \sigma((\tilde{Z}_k)_{k \leq n})\}_{n \geq 0}$. We shall use Theorem 3.2 in Hall and Heyde (1980). We study

$$\begin{aligned}
 \sigma_n^2 &\triangleq \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} (u^\top \cdot B_m)^2 \\
 &= \sum_{1 \leq k, l \leq p} u_k u_l \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_{m-1})) (V_l^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_l^*(\tilde{Z}_{m-1})) | \tilde{\mathcal{L}}_{m-1}] \\
 &= \sum_{1 \leq k, l \leq p} u_k u_l \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(V_k^* V_l^*)(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*} V_k^*(\tilde{Z}_{m-1}) \tilde{\Pi}_{\theta^*} V_l^*(\tilde{Z}_{m-1})].
 \end{aligned}$$

Consider the function defined by

$$z \rightarrow (V_k^* V_l^*)(z) \triangleq \mathcal{V}_{k,l}(z).$$

We check that this function is in $\text{Lip}(\mathbf{E}')$ and verifies condition (15). First,

$$|\mathcal{V}_{k,l}(z)| \leq C(\delta_{\theta^*}(y) + \delta'_{\theta^*}(y))^2$$

(where C depends on Δ_2 and Δ'_2), with Δ_{2s} and Δ'_{2s} finite for some $s > 1$, by assumption. The Lipschitz criterion is verified by studying the quantity $\mathcal{V}_{k,l}(x, y, f, \sigma) - \mathcal{V}_{k,l}(x, y, f', \sigma')$, where $z = (x, y, f, \sigma)$ and $z' = (x, y, f', \sigma')$ are elements of the space \mathbf{E}' . We have

$$\begin{aligned} |V_k^* V_l^*(z) - V_k^* V_l^*(z')| &\leq |V_k^*(z)| \times |V_l^*(z) - V_l^*(z')| + |V_l^*(z)| \times |V_k^*(z) - V_k^*(z')| \\ &\leq C(\delta_{\theta^*}(y) + \delta'_{\theta^*}(y)) \max_{1 \leq l \leq p} |V_l^*(z) - V_l^*(z')|, \end{aligned}$$

and, by definition,

$$V_l^*(z) = \sum_{n \geq 0} (\tilde{\Pi}_{\theta^*}^n j_{\theta^*,l}(z) - \tilde{\Lambda}_{\theta^*}(j_{\theta^*,l})) = \sum_{n \geq 0} \tilde{\Pi}_{\theta^*}^n j_{\theta^*,l}(z).$$

But $j_{\theta^*,l}$ is in $\text{Lip}(\mathbf{E}')$ and, for $n \geq 1$, the function $\tilde{\Pi}_{\theta^*}^n j_{\theta^*,l}$ also belongs to $\text{Lip}(\mathbf{E}')$, with $\text{lip}'(\tilde{\Pi}_{\theta^*}^n j_{\theta^*,l}, x, y) \leq C(\delta_{\theta}(y)^3 + \text{lip}(U_{\theta,k}, y) + \mathbf{k}(U_{\theta,k}, y)) \tilde{\rho}^n$ (the same argument as in Remark 6 can be used to see this result). Then, assuming that Δ_4 , $(\Delta\Delta')_{4,1}$ and $(\Delta\Delta')_{3,2}$ are finite, we obtain that $\mathcal{Z}'_{k,l}$ belongs to $\text{Lip}(\mathbf{E}')$.

The same argument shows that $z \rightarrow \tilde{\Pi}_{\theta^*} V_k^*(z) \tilde{\Pi}_{\theta^*} V_l^*(z)$ belongs to $\text{Lip}(\mathbf{E}')$, and then

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n (V_k^* V_l^*)(\tilde{Z}_m) - (\tilde{\Pi}_{\theta^*} V_k^* \tilde{\Pi}_{\theta^*} V_l^*)(\tilde{Z}_{m-1}) &\xrightarrow[n \rightarrow \infty]{} \\ \tilde{\Lambda}_{\theta^*}(V_k^* V_l^*)(z) - \tilde{\Lambda}_{\theta^*}(\tilde{\Pi}_{\theta^*} V_k^* \times \tilde{\Pi}_{\theta^*} V_l^*) &\quad \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)}\text{-a.s.} \end{aligned}$$

that is,

$$\begin{aligned} \frac{1}{n} \sum_{1 \leq k, l \leq p} \sum_{m=1}^n u_k u_l [(V_k^* V_l^*)(\tilde{Z}_m) - (\tilde{\Pi}_{\theta^*} V_k^* \tilde{\Pi}_{\theta^*} V_l^*)(\tilde{Z}_{m-1})] \\ \xrightarrow[n \rightarrow \infty]{} \tilde{\Lambda}_{\theta^*}(V_k^* V_l^*) - \tilde{\Lambda}_{\theta^*}(\tilde{\Pi}_{\theta^*} V_k^* \times \tilde{\Pi}_{\theta^*} V_l^*) \quad \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)}\text{-a.s.} \end{aligned}$$

As this family is uniformly integrable under Assumption 7, the convergence also happens for its first moment, so that

$$\frac{1}{\sigma_n^2} \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(u^\top \cdot B_m)^2 | \tilde{\mathcal{Z}}_{m-1}] \xrightarrow[n \rightarrow \infty]{} 1 \text{ in } \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)}\text{-probability.}$$

We prove now that, for any positive ϵ ,

$$\frac{1}{n} \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} ((u^\top \cdot B_m)^2 \cdot \mathbb{1}_{|u^\top \cdot B_m| \geq \sigma_n \epsilon} | \tilde{\mathcal{Z}}_{m-1}) \xrightarrow[n \rightarrow \infty]{} 0 \text{ in } \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)}\text{-probability.}$$

For $\alpha > 0$,

$$\tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)} \left(\frac{1}{n} \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(u^\top \cdot B_m)^2 \cdot \mathbb{1}_{|u^\top \cdot B_m| \geq \sigma_n \epsilon} | \tilde{\mathcal{Z}}_{m-1}] \geq \alpha \right)$$

$$\begin{aligned}
&= \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} \left(\mathbb{1}_{\sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)}((u^\top \cdot B_m)^2 \mathbb{1}_{|u^\top \cdot B_m| \geq \sigma_n \epsilon} | \tilde{\mathcal{Z}}_{m-1}) \geq n\alpha} \right) \\
&\leq \frac{1}{n\alpha} \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(u^\top \cdot B_m)^2 \mathbb{1}_{|u^\top \cdot B_m| \geq \sigma_n \epsilon} | \tilde{\mathcal{Z}}_{m-1}] \\
&\leq \frac{1}{\alpha} \max_{1 \leq m \leq n} \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} [(u^\top \cdot B_m)^2 \mathbb{1}_{|u^\top \cdot B_m|^{2s-2} \geq (\sigma_n \epsilon)^{2s-2}}] \\
&\leq \frac{1}{\alpha(\epsilon \sigma_n)^{2s-2}} \max_{1 \leq m \leq n} \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} (|u^\top \cdot B_m|^{2s});
\end{aligned}$$

as $(\tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)}(|u^\top \cdot B_m|^{2s}))_m$ is bounded and $1/\sigma_n$ tends to zero, we obtain the result. We conclude that

$$\frac{1}{\sigma_n} \sum_{m=1}^n u^\top \cdot B_m \rightarrow N(0, 1) \text{ weakly under } \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)},$$

which means that

$$\frac{1}{n^{1/2}} \sum_{m=1}^n u^\top \cdot B_m \rightarrow N(0, u^\top \cdot J_{\theta^*} \cdot u) \text{ weakly under } \tilde{P}_{\theta^*, \tilde{\lambda}(\zeta)},$$

where $(J_{\theta^*})_{k,l} = \tilde{\Lambda}_{\theta^*}(V_k^* V_l^*) - \Lambda_{\theta^*}(\tilde{\Pi}_{\theta^*} V_k^* \times \tilde{\Pi}_{\theta^*} V_l^*)$. Moreover,

$$\frac{1}{n} \sum_{m=1}^n \tilde{\mathbb{E}}_{\theta^*, \tilde{\lambda}(\zeta)} (B_m^\top \cdot B_m) \xrightarrow{n \rightarrow \infty} J_{\theta^*}.$$

Then, equation (51) shows that

$$\begin{aligned}
n^{-1/2} \nabla l_n(\theta^*, \zeta) &\rightarrow N(0, J_{\theta^*}^*) \text{ weakly under } P^* \\
\text{and } \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}^* (\nabla l_n(\theta^*, \zeta)^\top \cdot \nabla l_n(\theta^*, \zeta)) &= J_{\theta^*}^*.
\end{aligned}$$

□

D.3. Exponential memorylessness of the Hessian of the prediction filter and consequences

The chain $\{\tilde{Z}_n\}$ lies in the space $\mathbf{E}'' = \mathbf{K} \times \mathcal{Y} \times \mathbf{S}^+ \times \Sigma \times \Sigma_2$ (where $\Sigma_2 = \{\tau = (\tau_{k,l})/\tau_{k,l}: \mathbf{K} \rightarrow \mathbb{R} \text{ and for all } v \text{ in } \mathbf{K}, \tau(v) \text{ is symmetric and positive}\}$). We first need to find a recurrence relation between $\nabla^2 f_{\theta, n+1}^\zeta$ and $\nabla^2 f_{\theta, n}^\zeta$. Using recurrence relation (35) between $\nabla f_{\theta, n+1}^\zeta$ and $\nabla f_{\theta, n}^\zeta$, we obtain that $\{\nabla^2 f_{\theta, n}^\zeta\}$ is a functional autoregressive process

$$\partial_{k,l}^2 f_{\theta, n+1}^\zeta(v) = \int_u a_\theta(Y_n, f_{\theta, n}^\zeta)(u, v) \partial_{k,l}^2 f_{\theta, n}^\zeta(u) d\gamma(u) + T_{\theta, k, l}(Y_n, f_{\theta, n}^\zeta, \nabla f_{\theta, n}^\zeta)(v), \quad (52)$$

where

$$\begin{aligned}
 T_{\theta,k,l}(y, f, \sigma)(v) &= \frac{\int \partial_{k,l}^2 q_{\theta}(u, v) g_{\theta}(y|u) f(u) d\gamma(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \\
 &+ \int \frac{\partial_l g_{\theta}(y|u) f(u) + g_{\theta}(y|u) \sigma_l(u)}{\int g_{\theta}(y|u) f(u) d\gamma(u)} \times \left(\partial_k q_{\theta}(u, v) - \frac{\int \partial_k q_{\theta}(w, v) g_{\theta}(y|w) f(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) d\gamma(u) \\
 &+ \int \frac{\partial_k g_{\theta}(y|u) f(u) + g_{\theta}(y|u) \sigma_k(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \times \left(\partial_l q_{\theta}(u, v) - \frac{\int \partial_l q_{\theta}(w, v) g_{\theta}(y|w) f(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) d\gamma(u) \\
 &+ \int \left[\frac{\partial_k g_{\theta}(y|u) \sigma_l(u) + \partial_l g_{\theta}(y|u) \sigma_k(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right. \\
 &- \left. \left(\frac{\int \partial_l g_{\theta}(y|w) f(w) d\gamma(w) + \int g_{\theta}(y|w) \sigma_l(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) \times \left(\frac{g_{\theta}(y|u) \sigma_k(u) + \partial_k g_{\theta}(y|u) f(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) \right. \\
 &- \left. \left(\frac{\int \partial_k g_{\theta}(y|w) f(w) d\gamma(w) + \int g_{\theta}(y|w) \sigma_k(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) \times \left(\frac{g_{\theta}(y|u) \sigma_l(u) + \partial_l g_{\theta}(y|u) f(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right) \right] \\
 &\times \left[q_{\theta}(u, v) - \frac{\int q_{\theta}(w, v) g_{\theta}(y|w) f(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right] d\gamma(u) \\
 &+ \int \frac{\partial_{k,l}^2 g_{\theta}(y|u) f(u)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \left[q_{\theta}(u, v) - \frac{\int q_{\theta}(w, v) g_{\theta}(y|w) f(w) d\gamma(w)}{\int g_{\theta}(y|w) f(w) d\gamma(w)} \right] d\gamma(u).
 \end{aligned}$$

Relation (52) can be iterated to obtain $\nabla^2 f_{\theta,n+1}^{\zeta} = \alpha_{n-m+1}(Y_n^m, f_{\theta,m}^{\zeta}, \nabla f_{\theta,m}^{\zeta}, \nabla^2 f_{\theta,m}^{\zeta}; \theta)$. Just as in the preceding sections, we prove an exponential lack of memory of the initial conditions for $\alpha_{n-m}(\cdot; \theta)$.

Proposition 10. *Under Assumptions 1 and 9, there exists $C > 0$ such that*

$$\begin{aligned}
 &\sup_{x_m, \dots, x_n} \int \|\alpha_{n-m+1}^{k,l}(y_n^m, f, \sigma, \tau; \theta) - \alpha_{n-m+1}^{k,l}(y_n^m, f', \sigma', \tau'; \theta)\|_1 \\
 &g_{\theta^*}(y_m|x_m) \dots g_{\theta^*}(y_n|x_n) d\nu(y_m) \dots d\nu(y_n) \\
 &\leq C(n-m)^2(1-\epsilon)^{n-m+1} [\|\tau_{k,l} - \tau'_{k,l}\|_1 + [\|\sigma_k - \sigma'_k\|_1 + \|\sigma_l - \sigma'_l\|_1] \\
 &\quad \times (1 + \|\sigma_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_k\|_1 + \|\sigma'_l\|_1) \\
 &\quad + \|f - f'\|_1 [\|\tau_{k,l}\|_1 + \|\tau'_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_k\|_1 + \|\sigma'_l\|_1)^2]].
 \end{aligned}$$

Then we define a set of Lipschitz functions on \mathbf{E}'' .

Definition 3. $\text{Lip}(\mathbf{E}'')$ is the set of vector-valued functions $h = (h_{k,l})_{1 \leq k, l \leq p}$, with each $h_{k,l}$ being a measurable function on \mathbf{E}'' such that, for all k, l in $\{1 \dots p\}$, for all (x, y) in $\mathbf{K} \times \mathcal{Y}$, there exist $\text{lip}''(h, x, y)$ and $\mathbf{k}''(h, x, y)$ such that, for any f, f' in \mathbf{S}^+ , σ, σ' in $\mathbf{\Sigma}$, and τ, τ' in $\mathbf{\Sigma}_2$,

$$\begin{aligned} & |h_{k,l}(x, y, f, \sigma, \tau) - h_{k,l}(x, y, f', \sigma', \tau')| \\ & \leq \text{lip}''(h, x, y)[\|\tau_{k,l} - \tau'_{k,l}\|_1 + (\|\sigma_k - \sigma'_k\|_1 + \|\sigma_l - \sigma'_l\|_1) \\ & \quad \times (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_l\|_1) \\ & \quad + \|f - f'\|_1[\|\tau_{k,l}\|_1 + \|\tau'_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_l\|_1)^2]], \\ & |h_{k,l}(x, y, f, \sigma, \tau)| \leq \mathbf{k}''(h, x, y)[\|\tau_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma_l\|_1)^2] \end{aligned}$$

and

$$\begin{aligned} \text{lip}''(h) &= \text{ess sup}_x \int \text{lip}''(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty, \\ \mathbf{k}''(h) &= \text{ess sup}_x \int \mathbf{k}''(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty. \end{aligned}$$

The transition kernel $\check{\Pi}$ satisfies a contraction condition for these functions; as a result, we can prove that the chain is ergodic.

Proposition 11. Under Assumptions 1 and 9, there exist finite constants $\check{\rho} \in]\rho, 1[$ and $C > 0$, and for all z, z' in \mathbf{E}'' , h in $\text{Lip}(\mathbf{E}'')$ and $n \geq 1$,

$$\begin{aligned} |\check{\Pi}_\theta^n h_{k,l}(z) - \check{\Pi}_\theta^n h_{k,l}(z')| &\leq C(\mathbf{k}''(h) + \text{lip}''(h))\check{\rho}^n[\|\alpha_1^{k,l}(y, f, \sigma, \tau; \theta)\|_1 \\ &\quad + \|\alpha_1^{k,l}(y', f', \sigma', \tau'; \theta)\|_1 + [1 + \|\Psi_1^k(y, f, \sigma; \theta)\|_1 + \|\Psi_1^l(y, f, \sigma; \theta)\|_1 \\ &\quad + \|\Psi_1^k(y', f', \sigma'; \theta)\|_1 + \|\Psi_1^l(y', f', \sigma'; \theta)\|_1]^2]. \end{aligned}$$

Corollary 4. Under Assumptions 1 and 9, the Markov chain \check{Z}_n is geometrically ergodic. There exist finite constants $\check{\rho} \in]\rho, 1[$ and $C > 0$ such that, for all h in $\text{Lip}(\mathbf{E}'')$, there exists a finite constant $\check{\Lambda}_\theta(h)$, such that, for all z in \mathbf{E}'' and for all $n \geq 1$,

$$\begin{aligned} |\check{\Pi}_\theta^n h_{k,l}(z) - \check{\Lambda}_\theta(h)| &\leq C \times (\text{lip}''(h) + \mathbf{k}''(h))\check{\rho}^n \\ &\quad [\|\alpha_1^{k,l}(y, f, \sigma, \tau; \theta)\|_1 + (1 + \|\Psi_1^k(y, f, \sigma; \theta)\|_1 + \|\Psi_1^l(y, f, \sigma; \theta)\|_1)^2]. \end{aligned} \quad (53)$$

For any initial probability measure $\check{\lambda}$ on $(\mathbf{E}'', \mathcal{B}(\mathbf{E}''))$ and for any function h in $\text{Lip}(\mathbf{E}'')$,

$$\check{\Lambda}_\theta(h) = \lim_{n \rightarrow \infty} \check{\mathbb{E}}_{\theta, \check{\lambda}}(h(\check{Z}_n)), \quad (54)$$

and there exists a solution in \mathbf{E}'' to the Poisson equation

$$(I - \check{\Pi}_\theta)V = h - \check{\Lambda}_\theta(h).$$

Proofs of Proposition 11 and Corollary 4 follow the same lines as for Proposition 9 and Corollary 3.

Proof of Proposition 6. Differentiating the expression for the gradient of the log-likelihood, we obtain, for all k, l in $\{1, \dots, p\}$,

$$\frac{1}{n} \partial_{k,l}^2 l_n(\theta, \zeta) = \frac{1}{n} \sum_{m=0}^{n-1} i_{\theta,kl}(\check{Z}_m) \quad \tilde{P}_{\theta, \check{\lambda}(\zeta)}\text{-a.s.}, \quad (55)$$

where

$$\begin{aligned} i_{\theta,kl}(x, y, f, \sigma, \tau) &\triangleq \frac{\int g_\theta(y|u) \tau^{kl}(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} + \frac{\int \partial_{kl}^2 g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} \\ &+ \frac{\int \partial_k g_\theta(y|u) \sigma_l(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} + \frac{\int \partial_l g_\theta(y|u) \sigma_k(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} \\ &- \left(\frac{\int \partial_l g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} + \frac{\int g_\theta(y|u) \sigma_l(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} \right) \\ &\times \left(\frac{\int \partial_k g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} + \frac{\int g_\theta(y|u) \sigma_k(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} \right). \end{aligned}$$

By a long but simple exercise in calculus, we can check that $i_{\theta,kl}$ is in $\text{Lip}(\mathbf{E}^n)$ and verify (15), assuming that Δ_{2s} , Δ'_{2t} , $\Delta''_r(\Delta\Delta'')_{1,1}$ and $(\Delta\Delta')_{2,2}$ are finite for some $r, s, t > 1$. Using similar arguments as for Proposition 3, we obtain

$$\frac{1}{n} \nabla^2 l_n(\theta^*, \zeta) \rightarrow \check{\Lambda}_{\theta^*}(i_{\theta^*}) \triangleq -I_{\theta^*}.$$

Then

$$I_{\theta^*} = -\lim_{n \rightarrow \infty} \mathbb{E}^* \left(\frac{1}{n} \nabla^2 l_n(\theta^*, \zeta) \right).$$

As permutation between differentiation and integration is valid in our case,

$$J_{\theta^*} = \lim_{n \rightarrow \infty} \mathbb{E}^* \left(\frac{1}{n} \nabla l_n(\theta^*, \zeta)^T \cdot \nabla l_n(\theta^*, \zeta) \right) = -\lim_{n \rightarrow \infty} \mathbb{E}^* \left(\frac{1}{n} \nabla^2 l_n(\theta^*, \zeta) \right) = I_{\theta^*}.$$

□

Proof of Proposition 7. This proof is the same as for the quantity $l_n(\theta, \zeta) - l_n(\theta', \zeta)$, generalizing Lemma 1 to the gradient and the Hessian of the prediction filter. □

Acknowledgements

The authors would like to thank Elisabeth Gassiat and Eric Moulines for helpful and fruitful discussions. The authors are also indebted to Olivier Capp for reading and correcting earlier versions of this manuscript.

References

- Arapostathis, A. and Marcus, S.I. (1990) Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Math. Control Signals Systems*, **3**, 1–29.
- Bakry, D., Milhaud, X. and Vandekerckhove, P. (1997) Statistique de chaînes de Markov cachées à espace d'états fini. Le cas non stationnaire. *C. R. Acad. Sci. Paris Sér. 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 occurring 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: Local asymptotic normality in the stationary case. *Bernoulli*, **2**, 199–228.
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD and London: Johns Hopkins University Press.
- Bickel, P.J., Ritov, Y. and Rydén, T. (1998) Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, **26**, 1614–1635.
- Celeux, G., Diebolt, J. and Robert, C. (1993) Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statist. Probab. Lett.*, **16**, 77–83.
- Chan, K.S. and Ledolter, J. (1995) Monte Carlo EM estimation in time series models involving counts. *J. Amer. Statist. Assoc.*, **90**, 242–252.
- Chib, S., Kim, S. and Shephard, N. (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.*, **65**, 361–394.
- Churchill, G.A. (1989) Stochastic models for heterogeneous DNA sequences. *Bull. Math. Biol.*, **51**, 79–94.
- Dacunha-Castelle, D. and Duflo, M. (1986) *Probability and Statistics II*. Berlin: Springer-Verlag.
- De Jong, P. and Shephard, 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 the EM algorithm. Discussion. *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38.
- Durbin, J. and Koopman, S.J. (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, 669–684.
- Fredkin, D.R. and Rice, J.A. (1987) Correlation functions on a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci.*, **87**, 161–172.
- Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Applications*. New York: Academic Press.
- Jensen, J.L. and Petersen, N.V. (1999) Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27**, 514–535.
- Juang, B.H. and Rabiner, L.R. (1991) Hidden Markov models for speech recognition. *Technometrics*, **33**, 251–272.

- LeGland, F. and Mevel, L. (2000a) Basic properties of the projective product, with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems*, **13**, 41–62.
- LeGland, F. and Mevel, L. (2000b) Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems*, **13**, 63–93.
- Leroux, B.G. (1992) Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.*, **40**, 127–143.
- MacDonald, I. and Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series*. London: Chapman & Hall.
- Mevel, L. (1997) Statistique asymptotique pour les modèles de Markov cachés. Doctoral thesis, Université de Rennes I.
- Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*. Berlin: Springer-Verlag.
- Pagano, M. (1974) Estimation of models of autoregressive signal plus white noise. *Ann. Statist.*, **2**, 99–108.
- Quinn, B.G. (1995) Statistical problems in the analysis of underwater sound. In P.M. Robinson and M. Rosenblatt (eds), *Athens Conference on Applied Probability and Time Series, Analysis*, Volume II: *Time Series Analysis in Memory of E.J. Hannan*, Lecture Notes in Statist. 115. New York: Springer-Verlag.
- Seneta, E. (1981) *Non-negative Matrices and Markov Chains*, 2nd edition. New York: Springer-Verlag.
- Sunyach, C. (1975) Une classe de chaînes récurrentes sur un espace métrique complet. *Ann. Inst. H. Poincaré Probab. Statist.*, **11**, 325–343.
- van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
- Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **20**, 595–602.

Received September 1999 and revised February 2001