Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Stochastic Processes and their Applications 116 (2006) 222–243

stochastic processes and their applications

www.elsevier.com/locate/spa

# Leroux's method for general hidden Markov models

Valentine Genon-Catalot[a,*], Catherine Laredo[b]

[a]*Université René Descartes Paris 5, U.F.R. de Mathématiques et Informatique, Laboratoire MAP5 (CNRS-UMR 8145), 45, rue des Saints-Pères, 75270 Paris Cedex 06, France*

[b]*Institut National de la Recherche Agronomique, Laboratoire Mathématiques Informatique et Applications (M.I.A.) et Laboratoire de Probabilités et Modèles Aléatoires (CNRS-UMR 7599), 78350 Jouy-en-Josas, France*

Received 28 June 2004; received in revised form 17 May 2005; accepted 19 October 2005

Available online 28 November 2005

## Abstract

The method introduced by Leroux [Maximum likelihood estimation for hidden Markov models, Stochastic Process Appl. 40 (1992) 127–143] to study the exact likelihood of hidden Markov models is extended to the case where the state variable evolves in an open interval of the real line. Under rather minimal assumptions, we obtain the convergence of the normalized log-likelihood function to a limit that we identify at the true value of the parameter. The method is illustrated in full details on the Kalman filter model.



*Keywords:* Markov chain; Hidden Markov models; Discrete time filtering; Parametric inference; Likelihood; Conditional likelihood; Subadditive ergodic theorem

## 1. Introduction

Hidden Markov models (HMMs) form a class of stochastic models which are of classical use in numerous fields of applications. In these models, the process of interest is a Markov chain $(U_n)$ with state space $\mathcal{U}$, which is not observed. Given the whole sequence of state variables $(U_n)$, the observed random variables $(Z_n)$ are conditionally independent and the conditional distribution of $Z_i$ depends only on the corresponding state variable $U_i$. Due to this description, HMMs are also called state space models. They are often concretely

*Corresponding author. Tel.: +33 1 42 86 21 14; fax: +33 1 44 55 35 35.

*E-mail addresses:* genon@math-info.univ-paris5.fr (V. Genon-Catalot), Catherine.Laredo@jouy.inra.fr (C. Laredo).

0304-4149/$ - see front matter 

doi:10.1016/j.spa.2005.10.005

obtained as follows. Suppose that $(\varepsilon_n)$ is a sequence of independent and identically distributed random variables (a noise), independent of the unobserved Markov chain $(U_n)$, and let the observed process be given by

$$Z_n = G(U_n, \varepsilon_n), \tag{1}$$

where $G$ is a known function. (For instance, $Z_n = h(U_n) + \varepsilon_n$ is classical). These models raise two kinds of problems which are addressed in two different areas of research and have a wide range of applications.

- Problem (1): Estimation of the unobserved variable $U_n$ (resp. $U_{n+1}$) from past observations $Z_n, \ldots, Z_1$. This is the problem of filtering (resp. prediction) in discrete time.
- Problem (2): Statistical inference based on $(Z_1, \ldots, Z_n)$ generally with the aim of estimating unknown parameters in the distribution of $(U_n)$.

In the literature devoted to problem (1), it is generally assumed that the spate space $\mathcal{U}$ of $(U_n)$ is a subset of an Euclidian space. In papers dealing with problem (2), it is more often assumed that the hidden chain $(U_n)$ has a finite state space $\mathcal{U} = \{u_1, \ldots, u_m\}$ and one wants to estimate its transition probabilities. For general references, see e.g. [14]. More recently, HMMs have been the object of a growing interest because they appear in the field of finance and econometry. Indeed, in stochastic volatility models (see e.g. [11]), the observed price process of a stock or asset, $S_n$, is such that $\log(S_{n+1}/S_n) = Z_n = h(U_n)\varepsilon_n$, where $(U_n)$ is a Markov chain and $(\varepsilon_n)$ a Gaussian white noise. The Markov chain is generally obtained as a discretisation of a continuous time Markov process and evolves in an open subset of an Euclidian space (see e.g. [16,1,7,8,17]).

In this paper, we are interested in problem (2), when the state variable $(U_n)$ evolves in an open interval $\mathcal{U} = (l, r)$ of the real line, with $-\infty \leqslant l < r \leqslant +\infty$. Moreover, we assume below that the hidden chain $(U_n, n \in \mathbb{Z})$ is strictly stationary and ergodic, and that the conditional distribution of $Z_n$ given $U_n = u$ does not depend on $n$ (for instance, in (1), it is the distribution of $G(u, \varepsilon_1)$). Under these assumptions, it is well known that the joint process $((U_n, Z_n), n \in \mathbb{Z})$ is also strictly stationary and ergodic (see e.g. [15,7]). We assume that we observe $Z_1, \ldots, Z_n$ extracted from the ergodic sequence $(Z_n, n \in \mathbb{Z})$. In this set-up, our aim is to study parametric inference based on the exact likelihood of $Z_1, \ldots, Z_n$.

Before giving details on the content of our paper, let us present the results and open problems in this domain. In a seminal paper, Leroux [15], assuming that $\mathcal{U}$ is a finite set, proves the convergence of the normalized log-likelihood of $(Z_1, \ldots, Z_n)$ and the consistency of the exact maximum likelihood estimator (MLE). The impressive feature of Leroux's paper is that his results are obtained under minimal assumptions. Relying on the consistency result proved by Leroux, Bickel et al. [2] prove the asymptotic normality of the exact MLE. Then, these results are extended to the case where $\mathcal{U}$ is a compact set by Jensen and Petersen [12] and more completely by Douc and Matias [4]. In this context, more general hidden Markov models such as switching autoregressive models are investigated by Douc et al. [5].

For a general state space of $(U_n)$, the asymptotic behaviour of the exact likelihood of $(Z_1, \ldots, Z_n)$ is still open, and consequently, the asymptotic behaviour of the exact MLE is not known. However, there is a well-known model which makes exception and is completely solved, namely the Kalman filter. In its simplest form, it may be described as follows. Let $(U_n)$ be a one-dimensional Gaussian AR(1)-process

$$U_n = aU_{n-1} + \eta_n, \tag{2}$$

with $|a|<1$ and $(\eta_n, n \in \mathbb{Z})$ a sequence of independent and identically distributed random variables with Gaussian distribution $\mathcal{N}(0, \beta^2)$. Suppose that the observed process is given by

$$Z_n = U_n + \varepsilon_n, \tag{3}$$

with $(\varepsilon_n, n \in \mathbb{Z})$ i.i.d. $\mathcal{N}(0, \gamma^2)$. It is easily seen that $(Z_n, n \in \mathbb{Z})$ is a Gaussian ARMA(1, 1) process. Therefore, by the theory of ARMA Gaussian likelihood functions, it is well known that the exact MLE of $(a, \beta^2, \gamma^2)$ is consistent and asymptotically Gaussian.

Below, we prove, for a general HMM, with $\mathcal{U}$ an open interval of $\mathbb{R}$, the convergence of the normalized log-likelihood to a limit that we identify at the true value of the parameter. Our results are obtained under a set of assumptions that appear rather minimal and hold for the Kalman filter. As an auxiliary result, we give a new simpler proof of the convergence of the log-likelihood in the Kalman filter.

Now, we may outline the paper. We follow step by step Leroux's paper preserving its spirit in the sense of obtaining results under minimal assumptions, and we point out the analogies and the differences. In Section 2, we present our framework: the unobserved Markov chain $(U_n)$ has state space $\mathcal{U} = (l, r)$ an open interval of $\mathbb{R}$ $(-\infty \leqslant l < r \leqslant +\infty)$. Its transition operator $P_\theta$ depends on an unknown parameter $\theta$ and transition probabilities $P_\theta(u, \mathrm{d}v) = p(\theta, u, v)\,\mathrm{d}v$ have densities with respect to the Lebesgue measure of $\mathcal{U}$ (denoted by $\mathrm{d}v$) (Assumptions (A0)–(A1)). For simplicity, the conditional distribution of $Z_n$ given $U_n = u$, say $F_u(\mathrm{d}z)$, contains no additional unknown parameter. We assume that, when $u$ is considered as a parameter, $F_u(\mathrm{d}z) = f(z/u)\mu(\mathrm{d}z)$ defines a standard dominated regular family of distributions with $f(z/u) > 0$ and, for all $z$, ($\mu$-a.e.), $u \to f(z/u)$ continuous and bounded on $\mathcal{U}$ (Assumptions (B1)–(B2)). The exact likelihood of $(Z_1, \ldots, Z_n)$ may be obtained by several classical formulae that we recall. One way is to compute first the conditional density of $(Z_1, \ldots, Z_n)$ given $U_1 = u$, say $p_n(\theta, z_1, \ldots, z_n/u)$ and then integrate with respect to the distribution of $U_1$. More generally, for any probability density $g$ on $\mathcal{U}$, we define the functions

$$p_n^g(\theta, z_1, \ldots, z_n) = \int_{\mathcal{U}} g(u) p_n(\theta, z_1, \ldots, z_n/u)\,\mathrm{d}u, \tag{4}$$

and set $p_n^g(\theta) = p_n^g(\theta, Z_1, \ldots, Z_n)$. When $g$ is the exact density of $U_1$, $p_n^g(\theta)$ is the likelihood function, that we denote below by $p_n(\theta)$. Otherwise, we call $p_n^g(\theta)$ a contrast process. As usual we denote by $\theta_0$ the true value of the parameter. Sections 3–4 are devoted to proving that, for all positive and continuous densities $g$ on $\mathcal{U}$, $\frac{1}{n}\log p_n^g(\theta)$ converges, in $\mathbb{P}_{\theta_0}$-probability, to the same limit $H(\theta_0, \theta)$. This is obtained in two steps. First (Section 3), we set, as in [15]

$$q_n(\theta, z_1, \ldots, z_n) = \sup_{u \in \mathcal{U}} p_n(\theta, z_1, \ldots, z_n/u), \tag{5}$$

and we call $q_n(\theta) = q_n(\theta, Z_1, \ldots, Z_n)$ the Leroux contrast. Since $\mathcal{U}$ is neither finite nor compact, we need an adequate assumption to prove that $q_n(\theta)$ is well defined for all $n$: This is obtained by assuming that the transition operator $P_\theta$ of $(U_n)$ is Feller, a property shared by all standard Markov chains on Euclidian spaces (Assumption (A3) and Proposition 3.1). Then, under a weak moment Assumption (B3), we prove that $\frac{1}{n}\log q_n(\theta)$

converges, $\mathbb{P}_{\theta_0}$-a.s., to a limit $H(\theta_0, \theta) \in [-\infty, +\infty)$ (Theorem 3.1). In Section 4, we prove that $\frac{1}{n}\log q_n(\theta)$ and $\frac{1}{n}\log p_n^g(\theta)$ have the same limit in $\mathbb{P}_{\theta_0}$-probability, for all positive and continuous $g$ (Theorem 4.1). This requires, in our context, additional assumptions. The main new Assumption (B4) is that the sequence of random variables

$$\hat{u}_n(\theta) = \operatorname*{argsup}_{u \in \mathscr{U}} p_n(\theta, Z_1, \ldots, Z_n/u) \tag{6}$$

is $\mathbb{P}_{\theta_0}$-tight, for all $\theta$. By strengthening Assumption (B4), we obtain the convergence of $\mathbb{E}_{\theta_0}\frac{1}{n}\log p_n^g(\theta)$ to the limit $H(\theta_0, \theta)$ (Proposition 4.1).

Section 5 is devoted to identify the limit $H(\theta_0, \theta_0)$. This is done by obtaining the limit of $\mathbb{E}_{\theta_0}\frac{1}{n}\log p_n(\theta_0)$, with another approach. It requires a precise insight into the prediction algorithm which allows to compute recursively the successive conditional distributions of $U_n$ given $Z_{n-1}, \ldots, Z_1$ (Proposition 5.1). Then, we study the conditional distributions, under $\mathbb{P}_{\theta_0}$, of $U_n$ given the finite past $Z_{n-1}, \ldots, Z_{n-p}$ and the infinite past $\underline{Z}_{n-1} = (Z_{n-1}, Z_{n-2}, \ldots)$. We prove that the conditional distribution of $U_n$ given $\underline{Z}_{n-1}$ (under $\mathbb{P}_{\theta_0}$) has a continuous density $\tilde{g}(\theta_0, u/\underline{Z}_{n-1})$ with respect to the Lebesgue measure on $\mathscr{U}$. Moreover, the process $((U_n, Z_n, \tilde{g}(\theta_0, u/\underline{Z}_{n-1}\,\mathrm{d}u)), n \in \mathbb{Z})$ is a stationary version of the Markov process $((U_n, Z_n, \mathscr{L}_{\mathbb{P}_{\theta_0}}(U_n/Z_{n-1}, \ldots, Z_1), n \geqslant 1))$(Propositions 5.2–5.4). Finally, we use the previous results to prove that $H(\theta_0, \theta_0)$ is linked with the entropy of the conditional distribution under $\mathbb{P}_{\theta_0}$ of $Z_1$ given the infinite past $\underline{Z}_0$ (Theorem 5.1).

In Section 6, we study in full details the Kalman filter model (see (2)–(3)). We prove that it satisfies all our assumptions. The checking of Assumption (B4) is simple since the r.v. (6) is explicit. The computation of the limit $H(\theta_0, \theta)$ for all $\theta$ (not only for $\theta_0$) is also explicit and obtained by using the limit of $\frac{1}{n}\log p_n^g(\theta)$ for a well-chosen density $g$. In Section 7, other examples are given. Section 8 contains concluding remarks and discusses briefly the remaining open problems to achieve consistency.

## 2. General framework

### 2.1. Model assumptions

Let us first recall the definition of a hidden Markov model (HMM) $(Z_n, n \in \mathbb{Z})$, defined for $n \in \mathbb{Z}$, with hidden chain $U_n \in \mathscr{U}$ and observed process $Z_n \in \mathscr{Z}$. We assume that $\mathscr{U}$ and $\mathscr{Z}$ are Borel subsets of an Euclidian space equipped with their respective Borel $\sigma$-fields.

**Definition 2.1.** The process $(Z_n, n \in \mathbb{Z})$, is a HMM if

1. We are given a time homogeneous strictly stationary Markov chain $(U_n, n \in \mathbb{Z})$, with state space $\mathscr{U}$ which is unobserved.
2. Given the sequence $(U_n, n \in \mathbb{Z})$, the random variables $(Z_i)$ are independent and the conditional distribution of $Z_i$ only depends on $U_i$.
3. The conditional distribution of $Z_i$ given $U_i = u$ does not depend on $i$.

HMMs possess some generic properties that they inherit from the hidden chain (see e.g. [15] for a finite state space and [7] for a general state space).

**Proposition 2.1.** *The joint process* $((U_n, Z_n), n \in \mathbb{Z})$ *satisfying the conditions of Definition* 2.1 *is a strictly stationary time homogeneous Markov chain. Moreover, if* $(U_n, n \in \mathbb{Z})$ *is ergodic, so is* $((U_n, Z_n), n \in \mathbb{Z})$.

Let us now introduce our framework and assumptions on the model which are separated into two groups. Assumptions (A) concern the hidden chain and Assumptions (B) the conditional distribution together with the marginal distribution.

- (A0) $\mathscr{U} = (l, r)$ is an open interval of $\mathbb{R}$, with $-\infty \leqslant l < r \leqslant +\infty$.
- (A1) The transition operator $P_\theta$ of $(U_n)$ depends on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^p$, $p \geqslant 1$, and has transition densities with respect to the Lebesgue measure on $(\mathscr{U}, \mathscr{B}(\mathscr{U}))$ hereafter denoted by $\mathrm{d}u$: $\forall \theta \in \Theta,\ P_\theta(u, \mathrm{d}v) = p(\theta, u, v)\,\mathrm{d}v$.
- (A2) The transition operator of $(U_n)$ satisfies
  (i) $\forall \varphi \in C_b(\mathscr{U}), \quad P_\theta \varphi \in C_b(\mathscr{U})$, where $C_b(\mathscr{U})$ is the space of continuous and bounded functions on $\mathscr{U}$ ($P_\theta$ is Feller),
  (ii) if $\varphi > 0$ and continuous, $P_\theta \varphi > 0$.
- (A3) For all $\theta \in \Theta$, the transition operator $P_\theta$ admits a stationary distribution $\pi_\theta(\mathrm{d}u)$ having a density $g(\theta, u)$ with respect to $\mathrm{d}u$ and the chain with marginal distribution $\pi_\theta(\mathrm{d}u) = g(\theta, u)\,\mathrm{d}u$ is ergodic.
- (A4) For all $\theta$, $u \to g(\theta, u)$ is continuous and positive on $\mathscr{U}$.
- (A5) For all $\theta$,
  (i) $(u, v) \to p(\theta, u, v)$ is continuous.
  (ii) $P_\theta$ is reversible, i.e., for all $(u, v) \in \mathscr{U} \times \mathscr{U}$, $p(\theta, u, v)/g(\theta, v) = p(\theta, v, u)/g(\theta, u)$.
  (iii) For all compact subsets $K$ of $\mathscr{U}$, $\sup_{u \in K, v \in \mathscr{U}} (p(\theta, u, v)/g(\theta, v)) < +\infty$.
- (A6) $\int_{\mathscr{U} \times \mathscr{U}} \mathrm{d}u\,\mathrm{d}v\, g(\theta, u)\,(p^2(\theta, u, v)/g(\theta, v)) = \int p(\theta, v, u) p(\theta, u, v)\,\mathrm{d}u\,\mathrm{d}v < +\infty$.

Assumptions (A0)–(A5) are rather weak and standard. They hold for many classical models of Markov chains. We especially stress on the simplicity of (A2) which, together with (B2) below, allows the existence of Leroux's contrast. In particular, we do not need to bound the transition densities from below as it is done in general. Assumption (A6) is less standard. We just need it in Section 5.

- (B1) $\mathscr{Z} = \mathbb{R}$, the conditional distribution of $Z_i$ given $U_i = u$ is known and has a density $f(z/u)$ with respect to a dominating measure $\mu(\mathrm{d}z)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, the function $(u, z) \to f(z/u)$ is jointly measurable.
- (B2) For $\mu$ a.e. $z \in \mathbb{R}$, the function $u \to f(z/u)$ is continuous and bounded from above, and $\forall u \in \mathscr{U}$, $f(z/u) > 0$.
- (B3) Let $q_1(z) = \sup_{u \in \mathscr{U}} f(z/u)$. For all $\theta \in \Theta$, $\mathbb{E}_\theta(\log^+(q_1(Z_1)) < \infty$.

Assumptions (B) are not stringent and concern properties of a known family of distributions, the conditional laws of $Z_i$ given $U_i = u$, for $u \in \mathscr{U}$. They mean that these laws considered as a statistical model with respect to the parameter $u$, satisfy the usual properties of a dominated statistical experiment. Assumption (B3) is very weak as we shall see in the examples.

### *2.2. Likelihood and related contrast processes*

We now recall some classical formulae to derive the likelihood of HMMs and consider some associated contrast processes under Assumptions (A)–(B). We denote by

$\Omega = \mathscr{U}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}$ the canonical space endowed with the Borel $\sigma$-field $\mathscr{A} = \mathscr{B}(\Omega)$, $(U_n, Z_n)$ are the canonical coordinates on $\Omega$, and $\mathbb{P}_\theta$ is the distribution of $(U_n, Z_n)_{n\in\mathbb{Z}}$. For $n \in \mathbb{Z}$, the marginal distribution of $(U_n, Z_n)$ is

$$g(\theta, u)f(z/u)\,\mathrm{d}u\,\mu(\mathrm{d}z). \tag{7}$$

The transition probability of the Markov chain $(U_n, Z_n)$ is equal to

$$p(\theta, u, u')f(z'/u')\,\mathrm{d}u'\mu(\mathrm{d}z'). \tag{8}$$

On $(\Omega, \mathscr{A}, \mathbb{P}_\theta)$, the process $(Z_n)_{n\in\mathbb{Z}}$ is a HMM in the sense of Definition 2.1. We observe the sequence $(Z_1, \dots, Z_n)$ for $n \geqslant 1$, and study the problem of estimating the unknown parameter $\theta \in \Theta$ of the hidden chain $(U_n)$. We denote by $\theta_0$ the true value of the parameter. Now, for $u \in \mathscr{U}$, the conditional distribution of $(Z_1, \dots, Z_n)$ given $U_1 = u$, under $\mathbb{P}_\theta$, has a density such that, for $n = 1$,

$$p_1(\theta, z_1/u) = p_1(z_1/u) = f(z_1/u), \tag{9}$$

and for $n \geqslant 2$, setting $u_1 = u$ in the integral below,

$$p_n(\theta, z_1, \dots, z_n/u) = f(z_1/u) \int_{\mathscr{U}^{n-1}} \prod_{i=2}^{n} p(\theta, u_{i-1}, u_i)f(z_i/u_i)\,\mathrm{d}u_2 \dots \mathrm{d}u_n. \tag{10}$$

Under $\mathbb{P}_\theta$, $(Z_1, \dots, Z_n)$ has density (with respect to $\mu(\mathrm{d}z_1) \otimes \dots \otimes \mu(\mathrm{d}z_n)$)

$$p_n(\theta, z_1, \dots, z_n) = \int_{\mathscr{U}} g(\theta, u)p_n(\theta, z_1, \dots, z_n/u)\,\mathrm{d}u. \tag{11}$$

Now, let $g$ be a probability density w.r.t. $\mathrm{d}u$ on $\mathscr{U}$ and set

$$p_n^g(\theta, z_1, \dots, z_n) = \int_{\mathscr{U}} g(u)p_n(\theta, z_1, \dots, z_n/u)\,\mathrm{d}u. \tag{12}$$

Using these notations, the exact likelihood of $(Z_1, \dots, Z_n)$ is equal to

$$p_n(\theta) = p_n(\theta, Z_1, \dots, Z_n). \tag{13}$$

The likelihood of $(Z_1, \dots, Z_n)$ if $U_1$ had distribution $g(u)\,\mathrm{d}u$ is

$$p_n^g(\theta) = p_n^g(\theta, Z_1, \dots, Z_n). \tag{14}$$

We will study for all $\theta$ under $\mathbb{P}_{\theta_0}$ the exact likelihood $p_n(\theta)$ and the processes $p_n^g(\theta)$, that we shall call contrast processes.

Now, there is another expression for the exact likelihood $p_n(\theta)$ which relies on non-linear filtering theory. Let us denote by $p_i(\theta, z_i/z_{i-1}, \dots, z_1)$ the conditional density of $Z_i$ given $Z_{i-1} = z_{i-1}, \dots, Z_1 = z_1$ under $\mathbb{P}_\theta$. We have

$$p_n(\theta, z_1, \dots, z_n) = p_1(\theta, z_1) \prod_{i=2}^{n} p_i(\theta, z_i/z_{i-1}, \dots, z_1). \tag{15}$$

For $i \geqslant 2$, denote by

$$g_i(u_i) = g_i(\theta, u_i/z_{i-1}, \dots, z_1) \tag{16}$$

the conditional density under $\mathbb{P}_\theta$ of $U_i$ given $Z_{i-1} = z_{i-1}, \dots, Z_1 = z_1$. Then,

$$p_i(\theta, z_i/z_{i-1}, \dots, z_1) = \int_{\mathscr{U}} g_i(\theta, u_i/z_{i-1}, \dots, z_1)f(z_i/u_i)\,\mathrm{d}u_i. \tag{17}$$

It is well known from filtering theory that the predictive conditional densities $g_i$ can be obtained recursively. More precisely, let us set

$$\Phi_z^\theta(g)(u') = \frac{\int_{\mathscr{U}} g(u)f(z/u)p(\theta,u,u')\,\mathrm{d}u}{\int_{\mathscr{U}} g(u)f(z/u)\,\mathrm{d}u}. \quad (18)$$

Then, (12) is equal to

$$p_n^g(\theta,z_1,\dots,z_n) = p_1^g(\theta,z_1)\prod_{i=2}^{n} p_i^g(\theta,z_i/z_{i-1},\dots,z_1), \quad (19)$$

with

$$p_i^g(\theta,z_i/z_{i-1},\dots,z_1) = \int_{\mathscr{U}} g_i^g(\theta,u_i/z_{i-1},\dots,z_1)f(z_i/u_i)\,\mathrm{d}u_i, \quad (20)$$

where

$$g_i^g(\theta,./z_{i-1},\dots,z_1) = \Phi_{z_{i-1}}^\theta \circ \cdots \circ \Phi_{z_1}^\theta(g). \quad (21)$$

For more details, see [3,8,9].

## 3. Extension of the Leroux method to a general HMM

In 1992, Leroux has introduced, for finite $\mathscr{U}$, another useful contrast process. Our Assumptions (B) together with the Feller property of the chain enable us to extend this method to a general space $\mathscr{U}$.

Let us define using (10) for all $n\geqslant 1$

$$q_n(\theta,z_1,\dots,z_n) = \sup_{u\in\mathscr{U}} p_n(\theta,z_1,\dots,z_n/u). \quad (22)$$

We consider the associated process

$$q_n(\theta) = q_n(\theta,Z_1,\dots,Z_n). \quad (23)$$

For $n=1$, $q_1(\theta,z_1)=q_1(z_1)$ does not depend on $\theta$. Since $\mathscr{U}$ is general, we must prove that (22)–(23) are well defined (finite). We see below that the conditional densities $p_n(\theta,z_1,\dots,z_n/u)$ inherit the properties of $f(z/u)$.

**Proposition 3.1.** *For $n\geqslant 1$, $\theta\in\Theta$, for $\mu$ a.e. $(z_1,\dots,z_n)\in\mathbb{R}^n$, if* (B2) *and* (A2) *are verified, the function $u\to p_n(\theta,z_1,\dots,z_n/u)$ belongs to $C_b(\mathscr{U})$, and for all $u\in\mathscr{U}$, $p_n(\theta,z_1,\dots,z_n/u)>0$.*

**Proof.** For $n=1$, this is (B2). For $n=2$, using (10),

$$p_2(\theta,z_1,z_2/u) = f(z_1/u)\int_{\mathscr{U}} p(\theta,u,u')f(z_2/u')\,\mathrm{d}u' = f(z_1/u)P_\theta(f(z_2/.))(u).$$

Clearly, (B2) and (A2) imply that this function belongs to $C_b(\mathscr{U})$ and is positive. The conclusion is obtained for arbitrary $n$ by induction. □

Therefore, we can define the random variable with values in $\bar{\mathscr{U}}$

$$\hat{u}_n(\theta) = \hat{u}_n(\theta,Z_1,\dots,Z_n) \quad (24)$$

as any solution of $q_n(\theta)=q_n(\theta,Z_1,\dots,Z_n)=p_n(\theta,Z_1,\dots,Z_n/\hat{u}_n(\theta))$ and study $q_n(\theta)$.

**Theorem 3.1.** *Under* (A0)–(A3), (B1)–(B3), *the following holds*:

(i) *For all* $\theta$, $\mathbb{P}_{\theta_0}$*-a.s., as n tends to infinity,*

$$\frac{1}{n}\log q_n(\theta) \to H(\theta_0, \theta),$$

*where the limit* $H(\theta_0, \theta)$ *satisfies* $-\infty \leqslant H(\theta_0, \theta) < +\infty$.

(ii) *Moreover* $H(\theta_0, \theta) = \lim_n \mathbb{E}_{\theta_0} \frac{1}{n} \log q_n(\theta) = \inf_n \mathbb{E}_{\theta_0} \frac{1}{n} \log q_n(\theta)$.

**Proof.** For $n, m \geqslant 1$, we get using (10) that $p_{n+m}(\theta, z_1, \ldots, z_{n+m}/u)$ is equal to

$$f(z_1/u) \times \int_{\mathscr{U}^{n-1}} \mathrm{d}u_2 \ldots \mathrm{d}u_n \prod_{i=2}^{n} p(\theta, u_{i-1}, u_i) f(z_i/u_i) \times \left[ \int_{\mathscr{U}} p(\theta, u_n, u_{n+1}) p_m(\theta, z_{n+1}, \ldots, z_{n+m}/u_{n+1})\, \mathrm{d}u_{n+1} \right].$$

Therefore, bounding under the integral $p_m$ by $q_m$, for all $u$, $p_{n+m}(\theta, z_1, \ldots, z_{n+m}/u)$ is now lower than or equal to

$$f(z_1/u) \times \int_{\mathscr{U}^{n-1}} \mathrm{d}u_2 \ldots \mathrm{d}u_n \prod_{i=2}^{n} p(\theta, u_{i-1}, u_i) f(z_i/u_i) \times q_m(\theta, z_{n+1}, \ldots, z_{n+m}).$$

This is exactly equal to $p_n(\theta, z_1, \ldots, z_n/u) q_m(\theta, z_{n+1}, \ldots, z_{n+m})$. Taking the supremum over $u$ leads to, for all $z_1, \ldots, z_{n+m}$ (a.e. $\mu^{\otimes n+m}$),

$$q_{n+m}(\theta, z_1, \ldots, z_{n+m}) \leqslant q_n(\theta, z_1, \ldots, z_n) q_m(\theta, z_{m+1}, \ldots, z_{n+m}). \tag{25}$$

So, setting for $n < m$, $W_{n,m} = \log q_{m-n}(\theta, Z_{n+1}, \ldots, Z_m)$, we obtain that $W_{n,m}$ is a stationary and ergodic sequence with respect to the shift transformation $W_{n,m} \to W_{n+1,m+1}$ under $\mathbb{P}_{\theta_0}$, since, by (A3), $(Z_n)$ is a stationary and ergodic process under $\mathbb{P}_{\theta_0}$. Moreover, using (25), it is subadditive, i.e. for all $n < p < m$ ($\mathbb{P}_{\theta_0}$-a.s.)

$$W_{n,m} \leqslant W_{n,p} + W_{p,m}.$$

Therefore, we can apply Kingman's theorem for subadditive processes [13]: By (B3), we have $\mathbb{E}_{\theta_0}(W_{0,1}^+) = \mathbb{E}_{\theta_0}(\log^+(q_1(Z_1))) < \infty$. Hence, we get Theorem 3.1. □

**Remark 1.** Kingman's theorem ensures the existence of the deterministic limit $H(\theta_0, \theta)$ but this value may be equal to $-\infty$. Contrary to the classical ergodic theorem, it does not give a representation of the limit as the expectation of some random variable. This is why it is necessary to obtain such a representation by another proof.

## 4. Convergence of the loglikelihood

In this section, we study the convergence of the exact likelihood $p_n(\theta)$ and of $p_n^g(\theta)$ defined in (13)–(14). Let us set, under (A2)–(B2), (see (10))

$$f_n(\theta, u) = \log p_n(\theta, Z_1, \ldots, Z_n/u). \tag{26}$$

We now introduce some additional assumptions.

- (B4) For all $\theta$ such that $H(\theta_0,\theta) > -\infty$, the sequence defined in (24) satisfies $\mathbb{P}_{\theta_0}(\hat{u}_n(\theta) \in \mathscr{U}) \to 1$ as $n$ tends to infinity and it is $\mathbb{P}_{\theta_0}$-tight in $\mathscr{U}$.
- (B5) The function $u \to f_n(\theta, u)$ is $C^2$ on $\mathscr{U}$, $\mathbb{P}_{\theta_0}$-a.s.
- (B6) Let $B(\hat{u}_n(\theta),\varepsilon) = \{u \in \mathscr{U}, |u - \hat{u}_n(\theta)| \leqslant \varepsilon\}$. There exists an $\varepsilon_0 > 0$ such that

$$\frac{1}{n} \sup_{u \in B(\hat{u}_n(\theta),\varepsilon_0)} |f''_n(\theta,u)| \to 0$$

  in $\mathbb{P}_{\theta_0}$-probability. ($f''_n(\theta,u)$ is the second derivative with respect to $u$).

Assumptions (B4)–(B6) are new and replace the too stringent assumption that $\mathscr{U}$ is finite or compact. In the Kalman filter example, the random variable $\hat{u}_n(\theta)$ can be explicitly computed and all assumptions hold for this model. Now, we prove that, under the above additional assumptions, $\frac{1}{n}\log p_n^g(\theta)$, $\frac{1}{n}\log p_n(\theta)$ have the same limit as $\frac{1}{n}\log q_n(\theta)$ as $n$ tends to infinity.

**Theorem 4.1.** *Assume* (A0)–(A4) *and* (B1)–(B6). *For any density $g$ on $\mathscr{U}$ satisfying that $u \to g(u)$ is continuous and positive on $\mathscr{U}$, we have, in $\mathbb{P}_{\theta_0}$-probability, as n tends to infinity*

$$\lim \frac{1}{n} \log p_n^g(\theta) = H(\theta_0,\theta). \tag{27}$$

*In particular, the result holds for the exact likelihood.*

**Proof.** Clearly, for all $\theta$, and all $g$, $\mathbb{P}_{\theta_0}$-a.s., $p_n^g(\theta) \leqslant q_n(\theta)$ (see ((10)–(12), (22)). Therefore, by Theorem 3.1,

$$\limsup \frac{1}{n} \log p_n^g(\theta) \leqslant H(\theta_0,\theta). \tag{28}$$

The whole difficulty lies in getting the lower bound. If $H(\theta_0,\theta) = -\infty$, the result is immediate. Now, fix $\theta$ such that $H(\theta_0,\theta) > -\infty$. From now on, $\theta$ is omitted in the notation ($f_n(u) = f_n(\theta,u)$ and $\hat{u}_n = \hat{u}_n(\theta)$). Using (B4), for any $\eta > 0$, there is an integer $n_0$, a compact set $K \subset \mathscr{U}$ and $\varepsilon_1 = \varepsilon_1(K) > 0$ such that

$$\forall n \geqslant n_0, \quad \mathbb{P}_{\theta_0}(\hat{u}_n \in K) \geqslant 1 - \eta \quad \text{and} \quad \mathbb{P}_{\theta_0}(B(\hat{u}_n,\varepsilon_1) \subset K) \geqslant 1 - \eta. \tag{29}$$

We may choose $\varepsilon_1 \leqslant \varepsilon_0$ where $\varepsilon_0$ is given by (B6). Using (26) and (10)–(12), we get

$$p_n^g(\theta) = \int_{\mathscr{U}} g(u) \exp(f_n(u))\,\mathrm{d}u \geqslant \int_{B(\hat{u}_n,\varepsilon_1)} g(u)\exp(f_n(u))\,\mathrm{d}u. \tag{30}$$

Define $Z_n(\varepsilon)$ as

$$Z_n(\varepsilon) = \sup_{u \in B(\hat{u}_n,\varepsilon)} |f''_n(u)|. \tag{31}$$

For $u \in B(\hat{u}_n,\varepsilon_1)$, since $f'_n(\hat{u}_n) = 0$, we have

$$f_n(u) \geqslant f_n(\hat{u}_n) - \frac{\varepsilon_1^2}{2} Z_n(\varepsilon_1). \tag{32}$$

Thus, $p_n^g(\theta) \geqslant q_n(\theta)\exp[-\frac{\varepsilon_1^2}{2}Z_n(\varepsilon_1)]\int_{B(\hat{u}_n,\varepsilon_1)} g(u)\,\mathrm{d}u$. On the event $\{B(\hat{u}_n,\varepsilon_1) \subset K\}$, $\inf_{u\in B(\hat{u}_n,\varepsilon_1)} g(u) \geqslant \inf_{u\in K} g(u) = c(K) > 0$. Hence,

$$\frac{1}{n}\log p_n^g(\theta) \geqslant \frac{1}{n}\log q_n(\theta) - \frac{\varepsilon_1^2}{2n}Z_n(\varepsilon_1) + \frac{1}{n}\log 2\varepsilon_1 c(K).$$

Since $\varepsilon_1 \leqslant \varepsilon_0$, $Z_n(\varepsilon_1) \leqslant Z_n(\varepsilon_0)$, by Assumption (B6), $\frac{1}{n}Z_n(\varepsilon_1)$ tends to 0 in $\mathbb{P}_{\theta_0}$-probability, which leads to Theorem 4.1 using (29). Choosing $g$ equal to the stationary density $g(\theta,.)$, we get the result for the exact likelihood. □

The next question is now to identify the limit $H(\theta_0,\theta)$. We only do it at $\theta=\theta_0$. This requires strengthening some of the previous assumptions.

- (B4′) Assumption (B4) holds and there exist a positive integer $k$ and a constant $C$ such that, for all $\theta$, and all $n$, $\mathbb{E}_{\theta_0}|\hat{u}_n(\theta)|^k \leqslant C$.
- (B5′) Assumption (B5) holds and there exists $\varepsilon > 0$ such that

$$\frac{1}{n}\mathbb{E}_{\theta_0}\left[\sup_{u\in B(\hat{u}_n(\theta),\varepsilon)} |f_n''(\theta,u)|\right] \to 0.$$

**Proposition 4.1.** *Assume* (A0)–(A4), (B1)–(B3), (B4′)–(B5′). *Let $g$ be a positive, continuous density satisfying*

$$\exists C>0,\ \forall u\in(l,r)\quad |\log g(u)| \leqslant C(1+|u|^k).$$

*Then, for all $\theta$,*

$$\lim \frac{1}{n}\mathbb{E}_{\theta_0}\log p_n^g(\theta) = H(\theta_0,\theta). \tag{33}$$

**Proof.** Since $p_n^g(\theta) \leqslant q_n(\theta)$, by Theorem 3.1(ii), we have

$$\limsup \frac{1}{n}\mathbb{E}_{\theta_0}\log p_n^g(\theta) \leqslant H(\theta_0,\theta). \tag{34}$$

Using the r.v. $Z_n(\varepsilon)$ defined in (31), we have the following lower bound:

$$\frac{1}{n}\log p_n^g(\theta) \geqslant \frac{1}{n}\log q_n(\theta) - \frac{\varepsilon^2}{2n}Z_n(\varepsilon) + \frac{1}{n}\log\int_{B(\hat{u}_n,\varepsilon)} g(u)\,\mathrm{d}u. \tag{35}$$

Now, using (B4′)–(B5′), we get that $\limsup \frac{1}{n}\mathbb{E}_{\theta_0}\log p_n^g(\theta) \geqslant H(\theta_0,\theta)$. □

Let us note that Assumption (B4′) and the condition on $g$ are used to control the last term in the lower bound (35). They are fitted to the case $(l,r)=\mathbb{R}$. If $l$ or $r$ is finite, these conditions have to be adapted.

## 5. Entropy

This section is devoted to identifying the limit $H(\theta_0,\theta_0)$.

### 5.1. The prediction algorithm and its domain

Set

$$\mathscr{H}_\theta = \left\{ g : \mathscr{U} \to \mathbb{R}, \text{ continuous and } \frac{g}{g(\theta,.)} \in L^2_{\pi_\theta} \right\}, \tag{36}$$

where $\pi_\theta(\mathrm{d}u) = g(\theta, u)\,\mathrm{d}u$ is the stationary distribution of $P_\theta$ and $L^2_{\pi_\theta}$ is the space of square-integrable functions w.r.t. $\pi_\theta$. We consider, on $\mathscr{H}_\theta$, the topology associated with the following family of semi-norms: for all compact subsets $K$ of $\mathscr{U}$,

$$|g|_{K,\theta} = \sup_{u \in K} |g(u)| + \left\{ \int_{\mathscr{U}} \frac{g^2(u)}{g^2(\theta,u)} g(\theta,u)\,\mathrm{d}u \right\}^{1/2}.$$

Hence, $g_n \to g$ in $\mathscr{H}_\theta$ if and only if $g_n \to g$ uniformly on each compact subset of $\mathscr{U}$ and $\frac{g_n}{g(\theta,.)} \to \frac{g}{g(\theta,.)}$ in $L^2_{\pi_\theta}$. Endowed with this topology, $\mathscr{H}_\theta$ is a Polish space. Now, we define

$$\mathscr{F}_\theta = \left\{ g \in \mathscr{H}_\theta, g \geqslant 0, \int_{\mathscr{U}} g(u)\,\mathrm{d}u = 1 \right\}, \tag{37}$$

which is the set of probability densities belonging to $\mathscr{H}_\theta$. Clearly, $g(\theta, .)$ belongs to $\mathscr{F}_\theta$. Moreover, it is immediate to check that $\mathscr{F}_\theta$ is a closed subset of $\mathscr{H}_\theta$.

Now, let us recall the algorithm at $\theta$ that computes recursively the predictive conditional densities of $U_i$ given $Z_{i-1} = z_{i-1}, \ldots, Z_1 = z_1$ under $\mathbb{P}_\theta$, i.e. $g_i(\theta, u_i/z_{i-1}, \ldots, z_1)$ (see (16)–(18)). For $g : \mathscr{U} \to \mathbb{R}$ a probability density and $z \in \mathbb{R}$, let us set

$$\Phi^\theta_z(g) = \frac{A^\theta_z g}{h_z g} \tag{38}$$

with

$$h_z g = \int_{\mathscr{U}} f(z/u)\, g(u)\,\mathrm{d}u, \quad A^\theta_z g(u') = \int_{\mathscr{U}} f(z/u)\, g(u)\, p(\theta, u, u')\,\mathrm{d}u. \tag{39}$$

To obtain the successive conditional distributions, we must compute the iterates $\Phi^\theta_{z_n} \circ \Phi^\theta_{z_{n-1}} \circ \cdots \circ \Phi^\theta_{z_1}$ for $z_1, \ldots, z_n$ in $\mathbb{R}$. It is therefore central to find a proper space on which these iterates are well-defined.

**Proposition 5.1.** *Assume* (A0)–(A5), (B1)–(B2). *Then,*

(1) *For* $g \in \mathscr{F}_\theta$ *and* $\mu$*-a.e.* $z$, $\Phi^\theta_z(g) \in \mathscr{F}_\theta$. *If* $g > 0$ *and continuous, then* $\Phi^\theta_z(g) > 0$ *and continuous.*
(2) $g \to \Phi^\theta_z(g)$ *is continuous on* $\mathscr{F}_\theta$ *(in the topology of* $\mathscr{H}_\theta$*).*
(3) *For all* $g \in \mathscr{F}_\theta$ *and all* $n$, $(u, z_1, \ldots, z_n) \to \Phi^\theta_{z_n} \circ \Phi^\theta_{z_{n-1}} \circ \cdots \circ \Phi^\theta_{z_1}(g)(u)$ *is measurable on* $\mathscr{U} \times \mathbb{R}^n$.
(4) *Let* $(g_n)$ *be a sequence of functions belonging to* $\mathscr{F}_\theta$ *and let* $g \in \mathscr{F}_\theta$. *Assume that the sequence of probability measures* $\nu_n(\mathrm{d}u) = g_n(u)\,\mathrm{d}u$ *weakly converges to the probability measure* $\nu(\mathrm{d}u) = g(u)\,\mathrm{d}u$, *then, for all* $z$, *the sequence of probability measures* $\Phi^\theta_z(g_n)(u)\,\mathrm{d}u$ *weakly converges to the probability measure* $\Phi^\theta_z(g)(u)\,\mathrm{d}u$.

**Proof.** (1). Let $g \in \mathscr{F}_\theta$. Since $g \geqslant 0$, $g \neq 0$, using (B2), we get $h_z g > 0$. Therefore, $\Phi^\theta_z(g)$ is well-defined, non-negative and is a probability density. Now, we use reversibility to get

$$A_z^\theta g(u') = g(\theta, u') \int_{\mathscr{U}} f(z/u) \frac{g(u)}{g(\theta, u)} p(\theta, u', u)\, \mathrm{d}u. \tag{40}$$

This can be written as (see (B3) for $q_1(z)$)

$$\frac{A_z^\theta g}{g(\theta, .)} = P_\theta\left( f(z/.) \frac{g}{g(\theta, .)}\right) \leqslant q_1(z) P_\theta\left(\frac{g}{g(\theta, .)}\right). \tag{41}$$

Since $g/g(\theta, .) \in L^2_{\pi_\theta}$, we deduce that $A_z^\theta g/g(\theta, .) \in L^2_{\pi_\theta}$. Moreover, from (40), (A4) and (A5), we can obtain the continuity of the function $A_z^\theta g$. To complete the proof of (1), we use (A2)(ii).

To get (2), we prove the continuity of the operators $A_z^\theta$ and $h_z$ on $\mathscr{F}_\theta$. Both are linear on $\mathscr{H}_\theta$. For $g$ in $\mathscr{H}_\theta$,

$$|h_z g| \leqslant q_1(z) \int_{\mathscr{U}} \frac{|g(u)|}{g(\theta, u)} g(\theta, u)\, \mathrm{d}u \leqslant q_1(z) \left|\frac{g}{g(\theta, .)}\right|_{L^2_{\pi_\theta}}. \tag{42}$$

Now, suppose that, for functions $g_n, g$ in $\mathscr{F}_\theta$, the sequence $(g_n)$ converges to $g$ uniformly on each compact subset of $\mathscr{U}$. Since $g_n, g$ are probability densities, the pointwise convergence of $g_n$ to $g$ implies that, as $n$ tends to infinity

$$\int_{\mathscr{U}} |g_n(u) - g(u)|\, \mathrm{d}u \to 0. \tag{43}$$

(This is the Scheffé theorem). This in turn implies the weak convergence of $g_n(u)\,\mathrm{d}u$ to $g(u)\,\mathrm{d}u$. Since $u \to f(z/u)$ is continuous and bounded, we deduce that $h_z g_n \to h_z g$. Thus, $h_z$ is continuous on $\mathscr{F}_\theta$ (in the topology of $\mathscr{H}_\theta$). Now, using (41), we obtain

$$\left|\frac{A_z^\theta g}{g(\theta, .)}\right|_{L^2_{\pi_\theta}} \leqslant q_1(z) \left|\frac{g}{g(\theta, .)}\right|_{L^2_{\pi_\theta}}. \tag{44}$$

Consider again functions $g_n, g$ in $\mathscr{F}_\theta$ such that the sequence $(g_n)$ converges to $g$ uniformly on each compact subset of $\mathscr{U}$. Let $K$ be a compact subset of $\mathscr{U}$. We have

$$\sup_{u' \in K} |A_z^\theta (g_n - g)(u')| \leqslant \sup_{u' \in K, u \in \mathscr{U}} g(\theta, u') \frac{p(\theta, u', u)}{g(\theta, u)} \int_{\mathscr{U}} |g_n(u) - g(u)|\, \mathrm{d}u\, q_1(z). \tag{45}$$

Thus, using (43), (A5) and (44), we obtain that $A_z^\theta$ is continuous on $\mathscr{F}_\theta$. (This achieves (2)). To prove (3), let us check that, for $g \in \mathscr{F}_\theta$ and $z_1, \ldots, z_n \in \mathbb{R}$ (see (10)–(12))

$$\Phi_{z_n}^\theta \circ \Phi_{z_{n-1}}^\theta \circ \cdots \circ \Phi_{z_1}^\theta (g) = \frac{A_{z_n}^\theta \circ A_{z_{n-1}}^\theta \circ \cdots \circ A_{z_1}^\theta (g)}{p_n^g(\theta, z_1, \ldots, z_n)}. \tag{46}$$

For $n = 1$, it is the definition. For $n = 2$, using the linearity $h_z$ and $A_z^\theta$ and $h_{z_1} g > 0$, we get

$$\Phi_{z_2}^\theta \circ \Phi_{z_1}^\theta (g) = \frac{A_{z_2}^\theta \circ A_{z_1}^\theta g}{h_{z_2} \circ A_{z_1}^\theta g}. \tag{47}$$

The above denominator is equal to $\int_{\mathscr{U}} f(z_2/u')A^{\theta}_{z_1}(g)(u')\,\mathrm{d}u'$. Changing the order of integrations lead to

$$h_{z_2} \circ A^{\theta}_{z_1} g = \int_{\mathscr{U}} g(u)p_2(\theta, z_1, z_2/u)\,\mathrm{d}u.$$

The proof of (46) is achieved by induction. The denominator is measurable. Since $(u', z) \to A^{\theta}_{z}(g)(u')$ is measurable, the same holds for $(u, z_1, \ldots, z_n) \to A^{\theta}_{z_n} \circ A^{\theta}_{z_{n-1}} \circ \cdots \circ A^{\theta}_{z_1}(g)(u)$ by induction.

Let us prove (4). Suppose that, for $g_n, g$ in $\mathscr{F}_{\theta}$, $g_n(u)\,\mathrm{d}u$ weakly converges to $g(u)\,\mathrm{d}u$. Since $u \to f(z/u)$ is continuous and bounded, $h_z g_n$ tends to $h_z g$. By (A5)(iii), for all $u' \in \mathscr{U}$, the function $u \to f(z/u)p(\theta, u', u)/g(\theta, u)$ is also continuous and bounded. Therefore, using (40), for all $u'$, $A^{\theta}_{z} g_n(u')$ tends to $A^{\theta}_{z} g(u')$, so $\Phi^{\theta}_{z}(g_n)(u')$ tends to $\Phi^{\theta}_{z}(g)(u')$. Since these functions are probability densities, by Scheffé's theorem, we get the result. This completes the proof of Proposition 5.1. □

## 5.2. Conditional distributions given the infinite past

For $(z_n, n \in \mathbb{Z}) \in \mathbb{R}^{\mathbb{Z}}$, we denote by $\underline{z}_n = (z_n, z_{n-1}, \ldots)$ the vector of $\mathbb{R}^{\mathbb{N}}$ defined by the infinite past from $n$. Recall that, using (38)–(39),

$$g_{n+1}(\theta_0, ./Z_n, \ldots, Z_1) = \Phi^{\theta_0}_{Z_n} \circ \Phi^{\theta_0}_{Z_{n-1}} \circ \cdots \circ \Phi^{\theta_0}_{Z_1}(g(\theta_0, .)). \tag{48}$$

This is the conditional density, under $\mathbb{P}_{\theta_0}$, of $U_{n+1}$ given $Z_n, \ldots, Z_1$. Similarly, the conditional density of $U_1$ given $Z_0, Z_{-1} \ldots, Z_{-n+1}$ (under $\mathbb{P}_{\theta_0}$) is

$$g_{n+1}(\theta_0, ./Z_0, Z_{-1}, \ldots, Z_{-n+1}) = \Phi^{\theta_0}_{Z_0} \circ \Phi^{\theta_0}_{Z_{-1}} \circ \cdots \circ \Phi^{\theta_0}_{Z_{-n+1}}(g(\theta_0, .)). \tag{49}$$

This sequence converges in a sense precised in Proposition 5.3 to a function $\tilde{g}(\theta_0, ./\underline{Z}_0)$ that we first characterize.

**Proposition 5.2.** *Assume* (A0)–(A6). *There exists a regular version of the conditional distribution of $U_1$ given the infinite past $\underline{Z}_0$ under $\mathbb{P}_{\theta_0}$ having density $\tilde{g}(\theta_0, u/\underline{Z}_0)\,\mathrm{d}u$ satisfying*

(1) $\forall u \in \mathscr{U}$, $\mathbb{E}_{\theta_0}(p(\theta_0, U_0, u)/\underline{Z}_0) = \tilde{g}(\theta_0, u/\underline{Z}_0)$, $\mathbb{P}_{\theta_0}$-*a.s.*,
(2) $(u, \underline{Z}_0(\omega)) \to \tilde{g}(\theta_0, u/\underline{Z}_0(\omega))$ *is measurable*,
(3) $\mathbb{P}_{\theta_0}$-*a.s.*, $\tilde{g}(\theta_0, ./\underline{Z}_0)$ *belongs to* $\mathscr{F}_{\theta_0}$.

**Proof.** Let $\hat{v}(\theta_0, \mathrm{d}u_0; \underline{Z}_0(\omega))$ be a regular version of the conditional distribution under $\mathbb{P}_{\theta_0}$ of $U_0$ given $\underline{Z}_0$, defined for all $\omega \in \Omega$. Now, set

$$\tilde{g}(\theta_0, u/\underline{Z}_0(\omega)) = \int_{\mathscr{U}} p(\theta_0, u_0, u)\hat{v}(\theta_0, \mathrm{d}u_0; \underline{Z}_0(\omega)) \tag{50}$$

so that (1) holds. With our assumptions, (2) also holds and the above function is a probability density on $\mathscr{U}$. Using reversibility, we get

$$\tilde{g}(\theta_0, u/\underline{Z}_0(\omega)) = g(\theta_0, u) \int_{\mathscr{U}} \frac{p(\theta_0, u, u_0)}{g(\theta_0, u_0)} \hat{v}(\theta_0, \mathrm{d}u_0; \underline{Z}_0(\omega)). \tag{51}$$

By (A5)(iii), we deduce the continuity in $u$. It remains to prove that

$$u \to \int_{\mathscr{U}} \frac{p(\theta_0, u, u_0)}{g(\theta_0, u_0)} \hat{v}(\theta_0, du_0; \underline{Z}_0(\omega)) \in L^2_{\pi_{\theta_0}}. \tag{52}$$

By the Cauchy–Schwarz inequality, this is satisfied if

$$\int_{\mathscr{U}} g(\theta_0, u)\, du \int_{\mathscr{U}} \frac{p^2(\theta_0, u, u_0)}{g^2(\theta_0, u_0)} \hat{v}(\theta_0, du_0; \underline{Z}_0(\omega)) < \infty. \tag{53}$$

Changing the order of integrations, the above quantity is equal to

$$\mathbb{E}_{\theta_0}\left(\int_{\mathscr{U}} g(\theta_0, u) \frac{p^2(\theta_0, u, U_0)}{g^2(\theta_0, U_0)}\, du / \underline{Z}_0\right)(\omega). \tag{54}$$

This r.v. is finite $\mathbb{P}_{\theta_0}$-a.s. as soon as

$$\mathbb{E}_{\theta_0}\left(\int_{\mathscr{U}} g(\theta_0, u) \frac{p^2(\theta_0, u, U_0)}{g^2(\theta_0, U_0)}\, du\right) < \infty. \tag{55}$$

This is exactly our assumption (A6).

It remains to prove that the conditional distribution of $U_1$ given $\underline{Z}_0$ is exactly $\tilde{g}(\theta_0, u/\underline{Z}_0)\, du$. Hence, let us compute, for all $\varphi : \mathscr{U} \to [0,1]$ Borel, $\mathbb{E}_{\theta_0}(\varphi(U_1)/\underline{Z}_0)$. Using the Markov property of $(U_n, Z_n)$ and the special form of its transition probability (8) leads to

$$\mathbb{E}_{\theta_0}(\varphi(U_1)/\underline{U}_0, \underline{Z}_0) = \mathbb{E}_{\theta_0}(\varphi(U_1)/U_0, Z_0) = \mathbb{E}_{\theta_0}(\varphi(U_1)/U_0). \tag{56}$$

Hence, the result is obtained since

$$\mathbb{E}_{\theta_0}(\varphi(U_1)/\underline{Z}_0) = \int \hat{v}(\theta_0, du_0; \underline{Z}_0(\omega)) \int \varphi(u) p(\theta_0, u_0, u)\, du. \qquad \square \tag{57}$$

### *5.3. Convergence of the log-likelihood ratio at the true value of the parameter*

Now, we are able to give a meaning for the entropy of the stationary process $(Z_n, n \in \mathbb{Z})$ at $\theta_0$. Let

$$\tilde{p}(\theta_0, z/\underline{Z}_0) = \int_{\mathscr{U}} f(z/u) \tilde{g}(\theta_0, u/\underline{Z}_0)\, du = h_z(\tilde{g}(\theta_0, ./\underline{Z}_0)) > 0 \quad (\mathbb{P}_{\theta_0}\text{-a.s.}) \tag{58}$$

and define

$$\tilde{P}_{\theta_0}(dz/\underline{Z}_0) = \tilde{p}(\theta_0, z/\underline{Z}_0)\mu(dz). \tag{59}$$

Relation (59) defines a random probability measure which is a regular version of the conditional distribution, under $\mathbb{P}_{\theta_0}$, of $Z_1$ given $\underline{Z}_0$. Since $\tilde{p}(\theta_0, z/\underline{Z}_0) \leqslant q_1(z)$, we have, by (B3),

$$\mathbb{E}_{\theta_0} \log^+ \tilde{p}(\theta_0, Z_1/\underline{Z}_0) < +\infty. \tag{60}$$

Hence, we can set

$$-E(\theta_0) = \mathbb{E}_{\theta_0} \log \tilde{p}(\theta_0, Z_1/\underline{Z}_0), \tag{61}$$

where $-\infty \leqslant -E(\theta_0) < +\infty$. Taking the conditional expectation given $\underline{Z}_0$ yields

$$-E(\theta_0) = \mathbb{E}_{\theta_0}\left( \int_{\mathbb{R}} \log \tilde{p}(\theta_0, z/\underline{Z}_0) \tilde{P}_{\theta_0}(\mathrm{d}z/\underline{Z}_0) \right). \tag{62}$$

Thus, $E(\theta_0)$ is the expectation of the usual entropy of the distribution $\tilde{P}_{\theta_0}(\mathrm{d}z/\underline{Z}_0)$. Before studying the likelihood, we need some preliminary results.

**Proposition 5.3.** *The sequence of probability measures*

$$(g_{n+1}(\theta_0, u/Z_0, Z_{-1}, \ldots, Z_{-n+1})\,\mathrm{d}u)$$

(*see* (49)) *weakly converges*, $\mathbb{P}_{\theta_0}$-*a.s., to the probability measure* $\tilde{g}(\theta_0, u/\underline{Z}_0)\,\mathrm{d}u$.

**Proof.** Set $g_{n+1}(\theta_0, u/Z_0, Z_{-1}, \ldots, Z_{-n+1})\,\mathrm{d}u := \nu_{0,-n+1}(\theta_0, \mathrm{d}u)$ and $\nu_{0,-\infty}(\theta_0, \mathrm{d}u) := \tilde{g}(\theta_0, u/\underline{Z}_0)\,\mathrm{d}u$. For $x \in \mathbb{R}$ and $\omega \in \Omega$, set

$$F_n(x, \omega) = \int_{-\infty}^{x} \nu_{0,-n+1}(\theta_0, \mathrm{d}u; \omega) \quad \text{and} \quad F(x, \omega) = \int_{-\infty}^{x} \nu_{0,-\infty}(\theta_0, \mathrm{d}u; \omega), \tag{63}$$

which are continuous in $x$ and non-decreasing. For all $x \in \mathbb{R}$, $\mathbb{P}_{\theta_0}$-a.s., we have

$$F_n(x, .) = \mathbb{E}_{\theta_0}(1_{(-\infty,x]}(U_1)/Z_0, \ldots, Z_{-n+1}), \quad F(x, .) = \mathbb{E}_{\theta_0}(1_{(-\infty,x]}(U_1)/\underline{Z}_0). \tag{64}$$

By the martingale convergence theorem, we get that, as $n \to \infty$, $\forall x \in \mathbb{R}, \mathbb{P}_{\theta_0}$-a.s., $F_n(x, .) \to F(x, .)$. Therefore, there exists a set $N_{\theta_0}$ in $\mathscr{A}$ such that $\mathbb{P}_{\theta_0}(N_{\theta_0}) = 0$ and $\forall \omega \in N^c_{\theta_0}, \forall r \in Q, F_n(r, \omega) \to F(r, \omega)$. Now, fix $\omega \in N^c_{\theta_0}$ and $x \in \mathbb{R}$. For $\varepsilon > 0$, using the continuity of $F(., \omega)$, there exist $r', r'' \in Q$ such that

$$r' \leqslant x \leqslant r'' \quad \text{and} \quad F(x, \omega) - \varepsilon \leqslant F(r', \omega) \leqslant F(r'', \omega) \leqslant F(x, \omega) + \varepsilon.$$

The inequality $F_n(r', \omega) \leqslant F_n(x, \omega) \leqslant F_n(r'', \omega)$ implies $F(r', \omega) \leqslant \liminf_n F_n(x, \omega) \leqslant \limsup_n F_n(x, \omega) \leqslant F(r'', \omega)$. Hence, $F_n(x, \omega) \to F(x, \omega)$ and we have shown that, for all $\omega \in N^c_{\theta_0}$, the weak convergence of $\nu_{0,-n+1}(\theta_0, \mathrm{d}u; \omega)$ to $\nu_{0,-\infty}(\theta_0, \mathrm{d}u; \omega)$ holds. □

**Proposition 5.4.** *Let us set* $\tilde{g}_n(\theta_0, u) = \tilde{g}_n(\theta_0, u/\underline{Z}_{n-1})$. *Then, for all* $n \in \mathbb{Z}$, $\mathbb{P}_{\theta_0}$-*a.s.*, $\tilde{g}_{n+1}(\theta_0, .) = \Phi^{\theta_0}_{Z_n}(\tilde{g}_n(\theta_0, .))$.

**Proof.** Since $(U_n, Z_n)$ is strictly stationary, the conditional distribution, under $\mathbb{P}_{\theta_0}$, of $U_2$ given $\underline{Z}_1$ is $\tilde{g}(\theta_0, u/\underline{Z}_1)\,\mathrm{d}u$ and Proposition 5.3 leads to the weak convergence of $g_{n+2}(\theta_0, u/Z_1, Z_0, Z_{-1}, \ldots, Z_{-n+1})\,\mathrm{d}u$ to $\tilde{g}(\theta_0, u/\underline{Z}_1)\,\mathrm{d}u, \mathbb{P}_{\theta_0}$-a.s., where the densities are all in $\mathscr{F}_{\theta_0}$. We also have

$$g_{n+2}(\theta_0, ./Z_1, Z_0, Z_{-1}, \ldots, Z_{-n+1}) = \Phi^{\theta_0}_{Z_1}(g_{n+1}(\theta_0, ./Z_0, Z_{-1}, \ldots, Z_{-n+1})).$$

Using Propositions 5.1, (4) and 5.3, the sequence $g_{n+2}(\theta_0, u/Z_1, Z_0, Z_{-1}, \ldots, Z_{-n+1})\,\mathrm{d}u$ weakly converges, $\mathbb{P}_{\theta_0}$-a.s. to $\Phi^{\theta_0}_{Z_1}(\tilde{g}(\theta_0, ./\underline{Z}_0))$. Thus, we obtain

$$\Phi^{\theta_0}_{Z_1}(\tilde{g}(\theta_0, u/\underline{Z}_0))\,\mathrm{d}u = \tilde{g}(\theta_0, u/\underline{Z}_1)\,\mathrm{d}u.$$

Since the densities are continuous, we deduce $\Phi^{\theta_0}_{Z_1}(\tilde{g}(\theta_0, ./\underline{Z}_0)) = \tilde{g}(\theta_0, ./\underline{Z}_1)$. The result of Proposition 5.4 follows. □

The two previous propositions are also proved in [9] in the context of a specific model. An important consequence of these propositions is that, $\tilde{g}_n(\theta_0, u)\,\mathrm{d}u$ is the conditional distribution of $U_n$ given $\underline{Z}_{n-1}$ under $\mathbb{P}_{\theta_0}$. On $(\Omega, \mathscr{A}, \mathbb{P}_{\theta_0})$, the process $(U_n, Z_n, \tilde{g}_n(\theta_0, .))_{n \in \mathbb{Z}}$ with state space $\mathscr{U} \times \mathbb{R} \times \mathscr{F}_{\theta_0}$ is strictly stationary and ergodic. So, by Proposition 5.4, we have obtained a stationary regime for the Markov process $((U_n, Z_n, \Phi^{\theta_0}_{Z_n} \circ \Phi^{\theta_0}_{Z_{n-1}} \circ \cdots \circ \Phi^{\theta_0}_{Z_1}(g(\theta_0, .)), n \geqslant 1)$. Let us set

$$X_n = \log p_n(\theta_0, Z_1/Z_0, Z_{-1}, \ldots, Z_{-n+2}). \tag{65}$$

Then, $X_n = \log(\int f(Z_1/u) g_n(\theta_0, u/Z_0, Z_{-1}, \ldots, Z_{-n+2})\, \mathrm{d}u)$, and we can state:

**Theorem 5.1.** *Assume* (A0)–(A6), (B1)–(B3) *and that* (B4′)–(B5′) *hold at* $\theta = \theta_0$. *Assume moreover that the sequence of random variables* $(X_n^-)$ *is uniformly integrable. Then,* $H(\theta_0, \theta_0) > -\infty$ *and* $H(\theta_0, \theta_0) = -E(\theta_0)$.

**Proof.** Using (58)–(60), an application of the ergodic theorem yields

$$\frac{1}{n} \sum_{i=1}^{n} \log \tilde{p}(\theta_0, Z_i/\underline{Z}_{i-1}) \to -E(\theta_0).$$

Now, since $u \to f(Z_1/u)$ is continuous and bounded, using Proposition 5.3, $X_n$ defined in (65) tends to $X = \log \tilde{p}(\theta_0, Z_1/\underline{Z}_0)$, as $n$ tends to infinity, $\mathbb{P}_{\theta_0}$-a.s. Since, by (B3), the sequence $(X_n^+)$ is uniformly integrable, the additional assumption ensures that the same holds for $(|X_n|)$. So, first, we get that $\mathbb{E}_{\theta_0}|X| < \infty$ which implies $-E(\theta_0) > -\infty$ according to Definition (61). Second, $\mathbb{E}_{\theta_0}(X_n) \to \mathbb{E}_{\theta_0}(X)$. Now, by the strict stationarity, $\mathbb{E}_{\theta_0}(X_n) = \mathbb{E}_{\theta_0}(\log p_n(\theta_0, Z_n/Z_{n-1}, \ldots, Z_1))$. Taking Cesaro means, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\theta_0} \log p_i(\theta_0, Z_i/Z_{i-1}, \ldots, Z_1) = \frac{1}{n} \mathbb{E}_{\theta_0} \log p_n(\theta_0) \to -E(\theta_0).$$

Using Proposition 4.1 and (61) leads to the equality of the two limits $H(\theta_0, \theta_0) = -E(\theta_0)$. □

## 6. Specifying the model entirely on the Kalman filter

The Kalman filter is a hidden Markov model for which the behaviour of the likelihood is well known. Since the hidden state space is $\mathscr{U} = \mathbb{R}$, it is interesting to check the assumptions on this model, especially (B4)–(B6) and (B4′)–(B5′). Consider the one dimensional AR(1)-process $U_n = aU_{n-1} + \eta_n$, and the observed process $Z_n = U_n + \varepsilon_n$ defined in (2)–(3). We are interested in the estimation of $\theta = (a, \beta^2)$ and we shall suppose that $\gamma^2$ is known. The process $(U_n)$ is assumed in stationary regime: the marginal distribution of $(U_n)$ is the Gaussian law

$$\pi_\theta = \mathscr{N}(0, \tau^2) \quad \text{with } \tau^2 = \frac{\beta^2}{1 - a^2}. \tag{66}$$

Let $g(\theta, v)$ denote the density of $\pi_\theta$. The transition operator $P_\theta$ of $(U_n)$ has density equal to

$$p(\theta, u, v) = \frac{1}{\beta (2\pi)^{1/2}} \exp -\frac{(v - au)^2}{2\beta^2}. \tag{67}$$

Assumptions (A0)–(A5) hold. As for (A5)(iii), note that, since

$$\sup_{v \in \mathscr{U}} p(\theta, u, v)/g(\theta, v) \propto \exp((1 - a^2)u^2/2\beta^2),$$

this quantity is not uniformly bounded on the whole state space $\mathbb{R}$. Checking (A6) is also simple since $p(\theta, u, v)p(\theta, v, u)$ is up to a constant a two-dimensional Gaussian density. Consider now the assumptions on the conditional distribution of $Z_n$ given $U_n = u$. The density $f(z/u)$ is here

$$f(z/u) = \frac{1}{\gamma(2\pi)^{1/2}} \exp - \frac{(z-u)^2}{2\gamma^2}. \tag{68}$$

In this case, $q_1(z) = 1/\gamma(2\pi)^{1/2}$ is constant. Assumptions (B1)–(B3) are satisfied, and Theorem 3.1 holds. However, we will not use it to compute the limit. Instead, we use a $p_n^g(\theta)$ with a special $g$. To this end, we shall apply Theorem 4.1 and Proposition 4.1 after having checked Assumptions (B4)–(B6) and (B4′)–(B5′).

Let us compute $p_n^g(\theta, z_1, \ldots, z_n)$ (see (12)). For this, we need to specify the operator $\Phi_z^\theta$ (see (18)). In the Kalman filter model, this operator has the following special property: if $v_{(m,\sigma^2)}$ is the Gaussian density with mean $m$ and variance $\sigma^2$ (with the convention that the Dirac mass $\delta_m$ is a Gaussian law with nul variance and mean $m$), then, $\Phi_z^\theta(v_{(m,\sigma^2)})$ is also Gaussian. Therefore, it is enough to specify its mean and its variance. The following result is classically obtained by elementary computations

$$\Phi_z^\theta(v_{(m,\sigma^2)}) = v_{(\bar{m},\bar{\sigma}^2)} \tag{69}$$

with

$$\bar{m} = a(m\delta(\sigma^2) + z(1 - \delta(\sigma^2))), \quad \bar{\sigma}^2 = \beta^2 + a^2\sigma^2\delta(\sigma^2), \tag{70}$$

and

$$\delta(\sigma^2) = \frac{\gamma^2}{\gamma^2 + \sigma^2}. \tag{71}$$

Note that the degenerate case $v_{(u,0)} = \delta_u$ is included in these formulae with the convention $\sigma^2 = 0$. The mean $\bar{m}$ depends on $(m, \sigma^2)$, on $\theta$ and on the new observation $z$. A special feature of the Kalman filter is that the variance $\bar{\sigma}^2$ only depends on $\sigma^2$ and $\theta$ and neither on $m$ nor on $z$. The function $\sigma^2 \to \bar{\sigma}^2 = F^\theta(\sigma^2)$ is

$$F^\theta(v) = \beta^2 + a^2 \frac{v\gamma^2}{\gamma^2 + v}. \tag{72}$$

This function is convex increasing and has a unique stable fixed point $v(\theta)$ satisfying $\beta^2 \leqslant v(\theta) \leqslant \frac{\beta^2}{1-a^2}$.

Starting the iterations with $g = v_{(m,\sigma^2)}$, the density $g_i^g(\theta, ./z_{i-1}, \ldots, z_1)$ defined in (21) is Gaussian. We denote its mean and variance by

$$m_i(\theta, (m, \sigma^2), z_{i-1}, \ldots, z_1) \quad \text{and} \quad \sigma_i^2(\theta, \sigma^2). \tag{73}$$

We replace from now on the superscript $g$ by $(m, \sigma^2)$. Density (19) is now obtained as

$$p_n^{(m,\sigma^2)}(\theta, z_1, \ldots, z_n) \propto \prod_{i=1}^{n} (\sigma_i^2(\theta, \sigma^2) + \gamma^2)^{-1/2} \exp\left( - \frac{(z_i - m_i(\theta, (m, \sigma^2), z_{i-1}, \ldots, z_1))^2}{\sigma_i^2(\theta, \sigma^2) + \gamma^2} \right) \tag{74}$$

with the convention that $m_1 = m, \sigma_1^2 = \sigma^2$.

## 6.1. Checking the additional assumptions

Let us check (B4)–(B6), the assumptions that lead to Theorem 4.1. For this, we compute explicitly $\hat{u}_n(\theta, Z_1, \ldots, Z_n)$ defined in (24). Consider first the equation defining $\sigma_i^2(\theta, \sigma^2)$.

Using notation (72) and starting iterations with the initial density $v_{(m,\sigma^2)}$, we get $\sigma_1^2(\theta,\sigma^2) = \sigma^2$ and for $i \geqslant 2$,

$$\sigma_i^2(\theta,\sigma^2) = F_\theta \circ \cdots \circ F_\theta(\sigma^2) \tag{75}$$

is the $(i-1)$th iterate of $F_\theta$ starting from $\sigma^2$. By the properties of $F_\theta$, $\sigma_i^2(\theta,\sigma^2)$ converges as $i$ goes to infinity to the unique fixed point $v(\theta)$ of $F_\theta$. To simplify some notations below, whenever $\sigma^2 = 0$, we shall set

$$\sigma_i^2(\theta) = \sigma_i^2(\theta,0). \tag{76}$$

Consider now the recurrence equation defining $m_i(\theta,(u,0),Z_{i-1},\dots Z_1)$, i.e. the mean obtained when starting the iterations with the Dirac mass at $u$, after $(i-1)$ iterations corresponding to successive observations $Z_1,Z_2,\dots,Z_{i-1}$ (see (70)–(73)). We shall also simplify the notations and set, for $i \geqslant 2$,

$$m_i(\theta,u) = m_i(\theta,(u,0),Z_{i-1},\dots Z_1) \quad \text{and} \quad m_1(\theta,u) = u, \quad m_2(\theta,u) = au. \tag{77}$$

For $i \geqslant 1$, denote, using (71),

$$\delta_i = \delta(\sigma_i^2(\theta)) = \frac{\gamma^2}{\sigma_i^2(\theta)+\gamma^2}, \quad \alpha_0 = 1 \quad \text{and for } i \geqslant 2,\ \alpha_{i-1} = a^{i-1}\prod_{j=1}^{i-1}\delta_{i-j}. \tag{78}$$

We obtain that

$$m_i(\theta,u) = \alpha_{i-1}u + m_i(\theta,0), \quad m_1(\theta,0) = m_2(\theta,0) = 0. \tag{79}$$

For $i \geqslant 3$

$$m_i(\theta,0) = a\sum_{k=1}^{i-1}\alpha_k(1-\delta_{i-k})Z_{i-k}. \tag{80}$$

Therefore, the random function $f_n(\theta,u) = \log p_n(\theta,Z_1,\dots,Z_n/u)$ (see (26)) is equal to, $C$ denoting a given constant,

$$f_n(\theta,u) = -\frac{1}{2}\left(C + \sum_{i=1}^{n}\log(\sigma_i^2(\theta)+\gamma^2) + \frac{(Z_i - \alpha_{i-1}u - m_i(\theta,0)^2}{\sigma_i^2(\theta)+\gamma^2}\right). \tag{81}$$

Clearly, $u \to f_n(\theta,u)$ is a parabola, whose maximum is attained at point

$$\hat{u}_n(\theta) = \hat{u}_n(\theta,Z_1,\dots,Z_n) = \frac{A_n(\theta)}{D_n(\theta)}, \tag{82}$$

with

$$A_n(\theta) = \sum_{i=1}^{n}\frac{\alpha_{i-1}}{\sigma_i^2(\theta)+\gamma^2}(Z_i - m_i(\theta,0)), \quad D_n(\theta) = \sum_{i=1}^{n}\frac{\alpha_{i-1}^2}{\sigma_i^2(\theta)+\gamma^2}. \tag{83}$$

By (78), it follows that $|\alpha_{i-1}| \leqslant |a|^{i-1}$ with $|a|<1$. Since $\sigma_i^2(\theta)$ converges to $v(\theta)$, $D_n(\theta)$ converges to a positive constant $D(\theta)$, as $n$ tends to infinity. The numerator $A_n(\theta)$ is a random variable defined on $(\Omega,\mathscr{A},\mathbb{P}_{\theta_0})$, which is centred and Gaussian. Let us check that $A_n(\theta)$ converges in $L^2(\mathbb{P}_{\theta_0})$ to a limiting Gaussian random variable. Let $\|.\|_2$ denote the $L^2$-norm in $L^2(\mathbb{P}_{\theta_0})$. First note that, if $C(\theta_0) = \|Z_i\|_2$,

$$\|Z_i - m_i(\theta,0)\|_2 \leqslant C(\theta_0)\left(1+\sum_{k=1}^{i-1}|\alpha_k|\right) \leqslant C(\theta_0)\left(\frac{1}{1-|a|}\right).$$

Thus, for a constant $C$, $\|A_{n+m}(\theta) - A_n(\theta)\|_2 \leqslant C\sum_{i=n+1}^{n+m}\frac{|\alpha_{i-1}|}{\sigma_i^2(\theta)+\gamma^2}$. Since this upper bound tends to 0 as $n,m$ tend to infinity, the sequence $A_n(\theta)$ converges in $L^2(\mathbb{P}_{\theta_0})$ to a limiting random variable, say $A(\theta)$, and we obtain that, as $n$ tends to infinity

$$\hat{u}_n(\theta) \to \frac{A(\theta)}{D(\theta)} \quad \text{and} \quad \mathbb{E}_{\theta_0}\hat{u}_n(\theta)^2 \leqslant C. \tag{84}$$

Hence, we have (B4) and (B4′) with $k = 2$. As for (B5′), we see that $\frac{\partial^2}{\partial u^2} f_n(\theta, u)$, the second derivative of $f_n(\theta, u)$ w.r.t. $u$, has here a simple expression, independent of $u$,

$$\frac{\partial^2}{\partial u^2} f_n(\theta, u) = -D_n(\theta).$$

Therefore, (B5′) holds. Thus, Theorem 4.1 holds and Proposition 4.1 can be applied for any Gaussian density.

### 6.2. Computing the entropy $H(\theta_0, \theta_0)$ and the limit $H(\theta_0, \theta)$

According to Theorem 4.1, in $\mathbb{P}_{\theta_0}$-probability, for all $\theta$ and for all $g = v_{(m,\sigma^2)}$, $\frac{1}{n}\log p_n^{(m,\sigma^2)}(\theta)$ and $\frac{1}{n}\log q_n(\theta)$ have the same limit $H(\theta_0, \theta)$. The exact likelihood corresponds to $g = v_{(0,\tau^2)}$ (see (66)). The interest of Theorem 4.1 is that we can choose $g$ to obtain this common limit. Our choice leads to simpler computations.

For $g$, let us consider the Gaussian density with $m = 0$ and $\sigma^2 = v(\theta)$ the fixed point of (72) and set

$$p_n^{(0,v(\theta))}(\theta) = p_n^s(\theta). \tag{85}$$

Then, for all $i$, $\sigma_i(\theta, v(\theta)) = v(\theta)$ and

$$\delta(\sigma_i(\theta, v(\theta))) = \delta(v(\theta)) = \frac{\gamma^2}{\gamma^2 + v(\theta)} := \delta(\theta). \tag{86}$$

The iterations on the means simplify into $m_1 = 0$, $m_2 = a(1 - \delta(\theta))Z_1$ and,

$$m_i(\theta, (0, v(\theta)), Z_{i-1}, \ldots, Z_1) = a(1 - \delta(\theta)) \sum_{k=1}^{i-1} (a\delta(\theta))^{k-1} Z_{i-k}. \tag{87}$$

Let us set

$$H_z^\theta(m) = a(m\delta(\theta) + z(1 - \delta(\theta))). \tag{88}$$

Then, the algorithm defined in (70) is simply given by $m \to \bar{m} = H_z^\theta(m)$, and, for $i \geqslant 2$,

$$m_i(\theta, (0, v(\theta)), Z_{i-1}, \ldots, Z_1) = H_{Z_{i-1}}^\theta \circ \cdots \circ H_{Z_1}^\theta(0). \tag{89}$$

Therefore, the function $p_n^s(\theta)$ now satisfies (up to a constant)

$$\frac{1}{n}\log p_n^s(\theta) = -\frac{1}{2}\left(\log(\gamma^2 + v(\theta)) + \frac{1}{n}\sum_{i=1}^n \frac{(Z_i - m_i(\theta, (0, v(\theta)), Z_{i-1}, \ldots, Z_1))^2}{\gamma^2 + v(\theta)}\right). \tag{90}$$

We proceed now following the method of Section 5 and introduce successive iterations starting from the past. Let us consider

$$m_i(\theta, (0, v(\theta)), Z_0, \ldots, Z_{-i+2}) = H_{Z_0}^\theta \circ \cdots \circ H_{Z_{-i+2}}^\theta(0) \tag{91}$$

$$= a(1 - \delta(\theta)) \sum_{k=0}^{i-2} (a\delta(\theta))^k Z_{-k}. \tag{92}$$

Since the process $(Z_n, n \in \mathbb{Z})$ is, under $\mathbb{P}_{\theta_0}$, strictly stationary and Gaussian, and since $a\delta(\theta) < 1$, we have as $i$ tends to infinity, in $L^2(\mathbb{P}_{\theta_0})$, (and a.s.)

$$H^{\theta}_{Z_0} \circ \cdots \circ H^{\theta}_{Z_{-i+2}}(0) \to m(\theta, \underline{Z}_0), \tag{93}$$

where

$$m(\theta, \underline{Z}_0) = H^{\theta}_{Z_0} \circ \cdots \circ H^{\theta}_{Z_{-i+2}}(0) \circ \cdots = a(1 - \delta(\theta)) \sum_{k=0}^{+\infty} (a\delta(\theta))^k Z_{-k}. \tag{94}$$

In this model, Assumptions (B4′)–(B5′) hold for all $\theta$. Moreover, the random variables

$$X_n(\theta) = \log p_n(\theta, Z_1/Z_0, \ldots, Z_{-n+2})$$

satisfy $|X_n(\theta)| \leqslant C(Z_1^2 + m_n^2(\theta, (0, v(\theta)), Z_0, \ldots, Z_{-n+2}))$. So, they are uniformly integrable. Theorem 5.1 (applied for all $\theta$) yields that the limit of expression (90) is, up to a constant,

$$H(\theta_0, \theta) = -\frac{1}{2}\left(\log(\gamma^2 + v(\theta)) + \frac{1}{\gamma^2 + v(\theta)} \mathbb{E}_{\theta_0}((Z_1 - m(\theta, \underline{Z}_0))^2)\right). \tag{95}$$

We can now relate the above result to the one obtained in a previous paper. Let us introduce the random Gaussian distribution, well defined under $\mathbb{P}_{\theta_0}$, with mean $m(\theta, \underline{Z}_0)$ and variance $v(\theta) + \gamma^2$:

$$\tilde{P}_\theta = \mathcal{N}(m(\theta, \underline{Z}_0), v(\theta) + \gamma^2). \tag{96}$$

Then, if $K(P, Q)$ denotes the Kullback information of $P$ with respect to $Q$,

$$H(\theta_0, \theta_0) - H(\theta_0, \theta) = \mathbb{E}_{\theta_0}(K(\tilde{P}_{\theta_0}, \tilde{P}_\theta)). \tag{97}$$

Hence, we recover the identifiability assumption introduced in [8, pp. 306–308], where this quantity is proved to be non-negative and equal to 0 if and only if $\theta = \theta_0$.

Of course, in this model, the asymptotic properties of the exact MLE of $\theta$ are well known from the theory of Gaussian ARMA-processes. Our approach gives a new light based on the point of view of HMMs.

## 7. Other examples

It is possible to check several of our assumptions on other models. Each group of assumptions implies a result which has its own interest, the more complete being obtained under the whole set of assumptions.

Let us first look at Assumptions (A) which concern the hidden chain only. A special case with particular interest is obtained when $(U_n)$ derives from a regular sampling of a diffusion process $(V_t)$ on $(l, r)$ with $-\infty \leqslant l < r \leqslant +\infty$

$$\mathrm{d}V_t = b(\theta, V_t)\,\mathrm{d}t + a(\theta, V_t)\,\mathrm{d}W_t, \quad U_n = V_{n\Delta}, \tag{98}$$

where $\Delta > 0$ is fixed and $W$ is a standard Brownian motion. Under classical regularity assumptions on $b(\theta, .)$ and $a(\theta, .)$, Assumptions (A0)–(A5) may be easily checked and are standard. Assumption (A6) may possibly be checked using explicit expressions of the transition density.

Let us look at Assumptions (B).

**Example 1.** Consider an additive model $Z_n = U_n + \varepsilon_n$ where the noises $(\varepsilon_n)$ are $\mathcal{N}(0, 1)$. Assumptions (B1)–(B3) are automatically satisfied as we have seen in the previous section.

**Example 2.** When the state space of $(V_t)$ is $(0, \infty)$, setting $Z_n = U_n^{1/2}\varepsilon_n$ yields a class of HMMs which are often called stochastic volatility models. If, moreover $(\varepsilon_n)$ is a sequence of i.i.d. random variables with distribution $\mathcal{N}(0, 1)$, then

$$f(z/u) = \frac{1}{(2\pi u)^{1/2}} \exp\left(-\frac{z^2}{2u}\right). \tag{99}$$

We can compute $q_1(z) = (1/(2\pi e)^{1/2}|z|^{-1}$. Assumptions (B1)–(B2) are immediate. Assumption (B3) is a weak moment assumption on the stationary distribution of $(V_t)$. Assumptions (B4)–(B6) remain unchecked but Theorem 3.1 holds.

**Example 3.** Another model is proposed in [6,9,10]. The hidden chain is a standard Gaussian AR(1) process. The observation is given by $Z_n = U_n\varepsilon_n$, where the noise has a non-Gaussian distribution. For all $n$, $\varepsilon_n$ has the distribution of $\varepsilon\Gamma^{-1/2}$ where $\varepsilon$ and $\Gamma$ are independent random variables, $\varepsilon$ is a symmetric Bernoulli variable taking values $+1, -1$ with probability $\frac{1}{2}$ and $\Gamma$ has an exponential distribution with known parameter $\lambda > 0$. The exact likelihood is explicit and the checking of assumptions is on going work.

## 8. Concluding remarks and open problems

To complete the proof of consistency for the MLE, there remains to relate the limit $H(\theta_0, \theta)$ to a Kullback information as we have done in the Kalman filter example. This is really difficult. The difficulty comes from the fact that identifying this limit requires proving stability of the stochastic algorithm, for $(Z_n)$ under $\mathbb{P}_{\theta_0}$, $g_{n+1} = \phi^{\theta}_{Z_n}(g_n)$. The stability property needed here may be in a very weak sense as in [15] and should be related to Assumption (B4).

It is possible, up to very slight changes, to extend our results to the case where the hidden chain $(U_n)$ has a state space equal to an open convex subset of $\mathbb{R}^k$. Therefore, the case of discrete observations of continuous time stochastic volatility models (as in [7]) may be considered.

In the examples above, we focus on hidden chains which are discretisations of diffusion processes. However, other classes of ergodic Markov chains may be considered. For instance, Barndorff-Nielsen and Shephard [1] consider stochastic volatility models where the volatility is given by a non-Gaussian Ornstein–Uhlenbeck process (O–U Lévy process). The discretisation of such processes yields ergodic Markov chains on $(0, +\infty)$.

## Acknowledgements

This work was partially supported by the Research Training Network "Statistical Methods for Dynamical Stochastic Models" under the Improving Human Potential Program, financed by the Fifth Framework Program of the European Commission.

The authors wish to thank an anonymous referee for useful suggestions and comments.

## References

[1] O.E. Barndorff-Nielsen, N. Shephard, Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics, J. Roy. Statist. Soc. Ser. B 63 (2) (2001) 167–241.

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

[3] P. Del Moral, A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms, Ann. Inst. H. Poincaré Probab. Statist. 37 (2) (2001) 155–194.
[4] R. Douc, C. Matias, Asymptotics of the maximum likelihood estimator for general hidden Markov models, Bernoulli 7 (3) (2001) 381–420.
[5] R. Douc, E. Moulines, T. Ryden, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, Ann. Statist. 32 (5) (2004) 2254–2304.
[6] V. Genon-Catalot, A non-linear explicit filter, Statist. Probab. Lett. 61 (2003) 145–154.
[7] V. Genon-Catalot, T. Jeantheau, C. Larédo, Stochastic volatility models as hidden Markov models and statistical applications, Bernoulli 6 (6) (2000) 1051–1079.
[8] V. Genon-Catalot, T. Jeantheau, C. Larédo, Conditional likelihood estimators for hidden Markov models and stochastic volatility models, Scand. J. Statist. 30 (2) (2003) 297–316.
[9] V. Genon-Catalot, M. Kessler, Random scale perturbation of an AR(1)-process and its properties as a non-linear explicit filter, Bernoulli 10 (4) (2004) 701–720.
[10] V. Genon-Catalot, M. Kessler, Statistical inference for a random scale perturbation of an AR(1)-process, Prepublication 2004-10, MAP5, Université René Descartes, Paris 5, 2004.
[11] E. Ghysels, A. Harvey, E. Renault, Stochastic volatility, in: G.S. Maddala, C.R. Rao (Eds.), Handbook of Statistics 14: Statistical Methods in Finance, North-Holland, Amsterdam, 1996, pp. 119–192.
[12] J.L. Jensen, N.V. Petersen, Asymptotic normality of the maximum likelihood estimator in state space models, Ann. Statist. 27 (1999) 514–535.
[13] J.F.C. Kingman, Subadditive processes, in: P.L. Hennequin (Ed.), Ecole d'été de probabilités de Saint-Flour V-1976, Lecture Notes in Mathematics, vol. 539, Springer, Berlin, 1976, pp. 167–223.
[14] H.S. Künsch, State space and hidden Markov models, in: O.E. Barndorff-Nielsen, D.R. Cox, C. Klüppelberg (Eds.), Complex Stochastic Systems, Chapman and Hall/CRC, Boca Raton, 2001, pp. 109–173.
[15] B.G. Leroux, Maximum likelihood estimation for hidden Markov models, Stochastic Process Appl. 40 (1992) 127–143.
[16] N. Shephard (Ed.), Stochastic Volatility: Selected Readings. Series: Advanced Texts in Econometrics, Oxford University Press, Oxford, 2005.
[17] M. Sørensen, Prediction-based estimating functions, Econometrics J. 3 (2000) 123–147.