Statistical inference for a random scale perturbation of an AR(1)-process .

VALENTINE GENON-CATALOT[1] and MATHIEU KESSLER[2]

May 4, 2004

[1] *Laboratoire MAP 5 (CNRS-UMR 8145), UFR de Mathématiques et Informatique, Université Paris 5 René Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France*
[2] *Departamento de Matemàtica Aplicada y Estadstica, Universidad Politcnica de Cartagena, Paseo Alfonso XIII, 30203 Cartagena, Spain.*

Abstract

We study the properties of a hidden Markov model whose hidden chain is a standard AR(1) process observed with a random multiplicative perturbation. The exact likelihood is explicit. We also give several approximations of the exact likelihood.

MSC: primary 93E11, 60G35; secondary 62C10.

Keywords: filtering, discrete time observations, hidden Markov models, parametric inference.

Running title: Random scale perturbation of an AR(1).

1 Introduction

Statistical inference for hidden Markov models (HMMs) has been the subject of many recent contributions. In these models, the process of interest is a Markov chain (U_n) which is not observed. Given the whole sequence of state variables (U_n), the observed variables (Z_n) are conditionally independent and the conditional distribution of Z_i only depends on the corresponding state variable U_i. A concrete description is often obtained as follows. Suppose that (ε_n) is a sequence of independent and identically distributed random variables (the noise), independent of the unobserved Markov chain (U_n). Now let the observed process be given, for all n, by

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

Then, the sequence (Z_n) satisfies the above properties. In examples where these models are involved, the function F is most often known together with the common distribution of the ε_n's. (For general properties, see e.g. Elliott et al., 1995).

Statistical inference based on the observation of $(Z_1, \ldots, Z_n)$ is a difficult subject that has been investigated recently under restrictions on the state space of the hidden Markov chain. For an ergodic hidden chain with finite state space, Leroux's seminal paper (1992) proves the strong consistency of the exact maximum likelihood estimator. Then, asymptotic normality is proved in Bickel et al, 1998. The extension to the case where the hidden chain has a compact state space is treated in Jensen and Petersen, 1999, and more completely in Douc and Matias, 2001. When the state space of the hidden chain is not compact, for instance the real line, the asymptotic behaviour of the exact maximum likelihood is not known. The difficulties encountered for this study mainly come from the fact that the exact likelihood is not generally given by an explicit formula. Moreover, although this likelihood is known in theory, it is numerically untractable due to the high dimensional integrals that are involved in its formula.

Hence, there is a real interest in finding models where the hidden chain has a non compact state space and this likelihood has an explicit expression so that a direct study becomes feasible theoretically and numerically. The most popular example in this respect is the Kalman filter where the hidden Markov chain is an autoregressive process of order 1 and noises are additive and Gaussian. The asymptotic behaviour of unknown parameters in the Kalman filter is well known since the observed process is Gaussian and the likelihood can be explicitely studied (see e.g. Azencott and Dacunha-Castelle, 1984 , Genon-Catalot et al, 2003). Another model with explicit likelihood is presented in Genon-Catalot (2003): the hidden process is also a linear autoregressive model but the noises are multiplicative and have a specific distribution so that the observed process is non Gaussian. In Genon-Catalot and Kessler (2003), the filtering and prediction properties of this model have been studied. Indeed, all the distributions of the prediction and optimal filters (i.e. respectively the conditional distributions of U_n given $Z_{n-1}, \ldots, Z_1$ or given $Z_n, \ldots, Z_1$) are explicit so that the numerical implementation of the filters is simple and easy. Stability properties are also proved. In this paper, we focus on the statistical properties of the model and especially on the estimation of the unknown parameters present in the distribution of the hidden chain (the AR(1)-process). We apply for this model the general approach developped in Genon-Catalot et al (2003) based on the conditional likelihood with respect to the infinite past to derive the asymptotic behaviour of the exact likelihood at the true value of the parameter. We propose and study various approximations of the exact likelihood.

In Section 2, we present the model and the statistical problem. In Section 3, we recall the general method to obtain the exact likelihood of a HMM and specify it for our model relying on some formulae given in the appendix. Section 4 contains our main theoretical results. First, we prove stability results which are useful to obtain the asymptotic behaviour of the exact likelihood and of the approximations given in Section 5. We prove the convergence of the log-likelihood at the true value of the parameter. We cannot prove its convergence for all values of the parameter but the model enlights the difficulties to obtain this convergence. In Section 5, we propose two approximations of the likelihood and study their asymptotic behaviour. In Section 6, we consider Ryden's approach (1994) to estimate the unknown parameters.

Section 7 contains some concluding remarks and a discussion on open problems. The Appendix contains some formulae from Genon-Catalot and Kessler (2003).

2 Model and assumptions

Suppose that we observe random variables $Z_1, \ldots, Z_n, \ldots$ given by:

$$Z_n = U_n \varepsilon_n, \tag{2}$$

$$U_n = a\, U_{n-1} + \eta_n, \tag{3}$$

where (ε_n), (η_n) are two independent sequences of independent identically distributed real valued random variables and a is a real number. Assume that η_n has distribution $\mathcal{N}(0, \beta^2)$ and that the initial variable U_1 is independent of $(\varepsilon_n, n \geq 1,\ \eta_n, n \geq 2)$. As mentionned above, the sequence (U_n) (the signal) is unobserved and only, the sequence (Z_n) is observable.

Here, as in the well known Kalman filter, the signal (3) is a standard AR(1) model. In particular, if the initial variable U_1 is Gaussian, then $(U_n, n \geq 1)$ is a Gaussian process. However, in (2), the noise (ε_n) is not additive but multiplicative. To ensure explicit computations, we can no more consider a Gaussian distribution for the noise. We assume that, for all n, ε_n has the distribution of $\varepsilon\ \Gamma^{-1/2}$ where ε and Γ are independent random variables, ε is a symetric Bernoulli variable taking values $+1, -1$ with probability $1/2$ and Γ has an exponential distribution with parameter λ, a positive number. Then, for all $u \in \mathbb{R}$, $u \neq 0$, the distribution of $Z = u\varepsilon_1$ is given by:

$$F_u(dz) = f_u(z)\ dz, \quad \text{with} \quad f_u(z) = \frac{\lambda u^2}{|z|^3} \exp(-\frac{\lambda u^2}{z^2}) \tag{4}$$

For $u = 0$, $F_0(dz) = \delta_0(dz)$ is the Dirac mass at 0. The distribution $F_u(dz)$ is the conditional distribution of Z_n given the whole sequence $(U_n, n \geq 1)$ when $U_n = u$. If U_n is centered Gaussian with standard deviation σ_n, then the marginal distribution of Z_n is a $\sqrt{\lambda}\sigma_n\ t(2)$ where $t(2)$ denotes the Student distribution with 2 degrees of freedom. This distribution has moments of order $\delta < 2$ so that the model is well fitted to heavy-tailed data. It is proved in Genon-Catalot (2003) that, under the above assumption on the noise distribution (i.e. (4)), the joint distribution of $(Z_1, \ldots, Z_n)$ is explicit and simple. We recall below the way it is obtained. If the random variables ε_n have distribution $\varepsilon\Gamma^{-1/2}$ with Γ a Gamma variable with parameters k (integer) and λ, then the joint distribution of $(Z_1, \ldots, Z_n)$ is also explicit.

In what follows, we assume that the parameter λ specifying the distribution of the noise (see (4)) is known and that the unknown parameter is $\theta = (a, \beta^2)$. We assume that $a^2 < 1$ and that the process (U_n) is in stationary regime, i.e. that, for all n, U_n has distribution

$$\nu^\theta = \mathcal{N}(0, \sigma^2(\theta)), \tag{5}$$

with

$$\sigma^2(\theta) = \frac{\beta^2}{1 - a^2}. \tag{6}$$

In such a case, the joint process (U_n, Z_n) is strictly stationary and ergodic. And we shall assume that our observation $(Z_1, \ldots, Z_n)$ is extracted from the strictly stationary and ergodic process $((U_n, Z_n), n \in \mathbb{Z})$. Let us denote by $\mathbb{P}_\theta$ the distribution of $(U_n, Z_n), n \in \mathbb{Z}$ on the canonical space $\Omega = \mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}$ endowed with its usual Borel σ-field $\mathcal{A}$, and denote also by $((U_n, Z_n), n \in \mathbb{Z})$ the canonical process. We assume that our observation $(Z_1, \ldots, Z_n)$ is ruled by the true value $\theta_0 = (a_0, \beta_0^2)$ of the parameter.

3 Exact Likelihood

3.1 Derivation for general hidden Markov models

Let us first recall some general formulae to obtain the exact likelihood function associated with the observation of $(Z_1, \ldots, Z_n)$ given by (2-(3). For this, we refer e.g. to Genon-Catalot and al (2003), Genon-Catalot and Kessler (2003). We keep general notations to indicate that this derivation holds for general HMMs. Under $\mathbb{P}_\theta$, let us denote by $p_n(\theta, z_1, \ldots, z_n)$ the joint density of $(Z_1, \ldots, Z_n)$. Then, the density of the first observation Z_1 is equal to

$$p_1(\theta, z_1) = \int_{\mathbb{R}} g(\theta, u) f_u(z_1)\, du \tag{7}$$

where $g(\theta, u)$ denotes the stationary density of the hidden chain, i.e. the normal density with variance (6) and $f_u(z)$ is the conditional density of Z_1 given $U_1 = u$ given in (4). For all n, by our assumptions, Z_n has density (7).

Now, the joint process (U_n, Z_n) is Markov with transition $p(\theta, u, u') f_{u'}(z') du' dz'$ where $p(\theta, u, u')$ is the transition density of (U_n) specified by (3) i.e. the Gaussian density with mean au and variance β^2.

For $n \geq 2$, integrating the joint density of $((U_i, Z_i), i = 1, \ldots, n)$ which is obtained by the Markov property of the joint process, we get

$$p_n(\theta, z_1, \ldots, z_n) = \int_{\mathbb{R}^n} g(\theta, u_1) f_{u_1}(z_1) \prod_{i=2}^{n} p(\theta, u_{i-1}, u_i) f_{u_i}(z_i)\, du_1 \ldots du_n. \tag{8}$$

However, the above expression is generally hardly tractable. A more appropriate expression is obtained through the following formula

$$p_n(\theta, z_1, \ldots, z_n) = p_1(\theta, z_1) \prod_{i=2}^{n} p_i(\theta, z_i / z_{i-1}, \ldots, z_1) \tag{9}$$

where $p_i(\theta, z_i / z_{i-1}, \ldots, z_1)$ denotes the conditional density of Z_i given $(Z_{i-1}, \ldots, Z_1) = (z_{i-1}, \ldots, z_1)$. Because (Z_n) is not a Markov process, each of these conditional densities depends both on i and on the whole set of previous variables $(z_{i-1}, \dots, z_1)$. They are obtained as follows. Recall that $\nu^\theta(du)$ denotes the distribution of U_1 under $\mathbb{P}_\theta$ (see (5)) and set, for $n > 1$

$$\nu^\theta{}_{n|n-1:1}(du) = L(U_n \mid Z_{n-1}, \dots, Z_1/\mathbb{P}_\theta). \tag{10}$$

Then, the conditional density of Z_n given $(Z_{n-1}, \dots, Z_1)$ under $\mathbb{P}_\theta$ is given by

$$p_n(\theta, z_n/Z_{n-1}, \dots, Z_1) = \int_{\mathbb{R}} f_u(z_n)\nu^\theta_{n|n-1:1}(du). \tag{11}$$

Therefore, the exact likelihood function

$$L_n(\theta) = p_1(\theta, Z_1)\prod_{i=2}^{n} p_i(\theta, Z_i/Z_{i-1}, \dots, Z_1), \tag{12}$$

requires the knowledge of (10) and then the computation of (11).

3.2 Application to the model.

Now, we specialize the previous formulae to the model studied here. For this, we briefly recall some definitions and previous results of Genon-Catalot (2003) and Genon-Catalot and Kessler (2003). More details are given in the Appendix.

The computation of the successive conditional distributions $\nu^\theta_{n|n-1:1}$ is obtained by a recursive algorithm which is well known in filtering theory. However, the exact computation of these distributions is generally intractable except in very few models. The model investigated here belongs to these few. Indeed, it is proved in the above references that, if the distribution of U_1 belongs to a specific class, including the centered Gaussian laws, then, for all n, $\nu^\theta_{n|n-1:1}$ also belongs to this class. The class is composed of distributions of the form $\nu(du) = \nu_{\sigma,\alpha}(du) = g(u)du$, with

$$g(u) = \sum_{i\geq 0} \alpha_i \frac{1}{\sigma} g_i(\frac{u}{\sigma}) \tag{13}$$

where

$$g_i(u) = (2\pi)^{-1/2}\frac{u^{2i}}{C_{2i}}\exp(-\frac{u^2}{2}), \tag{14}$$

and $C_{2i} = E(X^{2i})$, for X a standard Gaussian variable. The above distribution is called Serial Gaussian distribution. It is specified by a scale parameter $\sigma > 0$ and a mixture parameter $\alpha = (\alpha_i, i \geq 0)$ (with $\forall i \geq 0, \alpha_i \geq 0$ and

$\sum_{i\geq 0}\alpha_i = 1$). We shall use the abbreviation SG(σ,α) for such a distribution. The centered Gaussian distribution $\mathcal{N}(0,\sigma^2)$ is SG$(\sigma,\alpha^{(s)})$ with

$$\alpha^{(s)} = (1,0,\ldots). \tag{15}$$

For all n, (10) is also a SG distribution and we denote its parameters by

$$(\sigma^2_{n-1:1}(\theta)), \alpha^{n-1:1}(\theta)) \tag{16}$$

with

$$\alpha^{n-1:1}(\theta)) = (\alpha_i^{n-1:1}(\theta), i \geq 0)). \tag{17}$$

The above parameters are random variables which are functions of $(Z_{n-1},\ldots,Z_1)$. For all n, Z_n, θ, the mappings $\Phi^\theta_{Z_n}$:

$$\nu^\theta_{n|n-1:1} \to \nu^\theta_{n+1|n:1} = \Phi^\theta_{Z_n}(\nu^\theta_{n|n-1:1}) \tag{18}$$

can be explicitely expressed in terms of the parameters of SG-distributions. The exact expressions are given in the Appendix. We obtain (10) by the iteration

$$\nu^\theta_{n|n-1:1} = \Phi^\theta_{Z_{n-1}} \circ \ldots \Phi^\theta_{Z_1}(\nu^\theta). \tag{19}$$

Because the distribution of U_1 has a mixture parameter given by (15)(it is a centered Gaussian), it holds that the distribution (10)(see also (19)and the Appendix) is a mixture of at most n components, i.e.

$$\forall i \geq n, \alpha_i^{n-1:1}(\theta) = 0. \tag{20}$$

Actually, the simulation study performed in Genon-Catalot and Kessler (2003) shows that the number of significantly non nul mixture coefficients is 1 or 2 (*i.e.* $\alpha_0^{n-1:1}(\theta) + \alpha_1^{n-1:1}(\theta)$ is close to 1). For all n, the conditional density of Z_n given $(Z_{n-1},\ldots,Z_1)$ is explicit. Indeed, when $f_u(z)$ is equal to (4), and since $\nu^\theta_{n|n-1:1}$ belongs to the SG class, the integral in (11) can be explicitely computed.

Proposition 3.1. *Suppose that ν is SG(σ,α), with density given by (13)-(14), then, the density*

$$h(z) = \int_{I\!R} f_u(z)\nu(du)$$

with $f_u(z)$ defined in (4) is given by

$$h(z) = \sum_{i\geq 0} \alpha_i h_i(\frac{z}{\lambda^{1/2}\sigma}), \quad h_i(z) = \frac{(2i+1)z^{2i}}{(z^2+2)^{i+3/2}}. \tag{21}$$

Proposition 3.1 is obtained by a simple computation using (4)-(13)-(14) and the formula $C_{2i+2} = (2i+1)C_{2i}$ for Gaussian even moments. The densities appearing in Proposition 3.1 are mixtures of distributions having a density proportional to z^{2i} with respect to a $\lambda^{1/2}\sigma t(2i+2)$ ($t(2i+2)$ denotes the Student distribution with $2i+2$ degrees of freedom). We thus obtain:

$$p_1(\theta, Z_1) = h_1(\theta, Z_1) = \frac{\lambda\, \sigma^2(\theta)}{\left(Z_1^2 + 2\lambda\sigma^2(\theta)\right)^{3/2}}. \tag{22}$$

This is a $\lambda^{1/2}\sigma(\theta)t(2)$ distribution. It is worth noting that the parameter λ is not identifiable in this model. This is why it is assumed to be known.

And, for $i \geq 2$, we have (see (20))

$$p_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) = \tag{23}$$

$$h_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) \times \sum_{j=0}^{i-1} \alpha_j^{i-1:1}(\theta) \frac{(2j+1)Z_i^{2j}}{\left(Z_i^2 + 2\lambda\sigma_{i-1:1}^2(\theta)\right)^j}, \tag{24}$$

where

$$h_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) = \frac{\lambda\, \sigma_{i-1:1}^2(\theta)}{\left(Z_i^2 + 2\lambda\sigma_{i-1:1}^2(\theta)\right)^{3/2}}. \tag{25}$$

Thus, the exact maximum likelihood (m.l.e.) $\hat{\theta}_n$ obtained by maximisation of (12) can be explicitely computed from the above expressions.

4 Main results.

To obtain results on the asymptotic behaviour under $\mathbb{P}_{\theta_0}$ of the normalized loglikelihood $\frac{1}{n} \log L_n(\theta)$, we need to study some stability properties of the measure-valued process (19).This process is called the prediction filter. We explain below the kind of stability properties that are needed and the properties that we have obtained.

4.1 Statement of the problem.

The likelihood function of HMMs has a specific form that we enlight now. For this, let us introduce some notations and a preliminary lemma. Let $\mathcal{P}(\mathbb{R})$ denote the set of probability measures on $\mathbb{R}$ and consider the mapping on $\mathbb{R} \times \mathcal{P}(\mathbb{R})$ defined by

$$(z, \nu) \rightarrow l(z, \nu) = \log \int_{\mathbb{R}} f_u(z)\, \nu(du) \geq -\infty. \tag{26}$$

Lemma 4.1. *The mapping (26) is measurable when $\mathcal{P}(I\!\!R)$ is endowed with the Borel σ-field associated with the topology of weak convergence of probability measures.*

The proof is given in the Appendix.

Then, by (11),

$$\frac{1}{n}\log L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} l(Z_i, \nu^{\theta}_{i|i-1:1}). \tag{27}$$

The measure-valued process $\nu_1 = \nu^{\theta}$, and for $n > 1$, $\nu_n = \nu^{\theta}_{n|n-1:1}$ obeys the recursive equation (see(18) and in the Appendix (97)-(105))

$$\nu_{n+1} = \Phi^{\theta}_{Z_n}(\nu_n). \tag{28}$$

Therefore, the joint process $X_n = (U_n, Z_n, \nu_n, n \geq 1)$ defined on $(\Omega, \mathcal{A}, I\!\!P_{\theta_0})$ is Markov, with state space $E = I\!\!R \times I\!\!R \times \mathcal{P}(I\!\!R)$. Its distribution depends both on θ_0 and θ. Hence, the problem of finding the limit of the normalized loglikelihood is directly connected with the problem of stability for this Markov process. Although $(U_n, Z_n, n \geq 1)$ is stationary by assumption, $(X_n, n \geq 1)$ is not (because the initial variable contains ν_1 which is deterministic) and we have to find its stationary distribution (if any) to obtain the limit of the normalized log-likelihood process.

Our approach is to find, for all θ a measure valued process, with values in the class of SG distribution, which is a strictly stationary solution of (28) under $I\!\!P_{\theta_0}$. Equation (28) being explicitely expressed in terms of the parameters of the SG distributions, it is therefore equivalent to find a strictly stationary process $(V_n = \sigma_n^2, \alpha^{(n)})$ with values in $I\!\!R^+ \times \{(\alpha_i, i \geq 0) \in [0,1]^{I\!\!N}, \sum_{i\geq 0}\alpha_i = 1\}$ solution of (28). We treat first the the scale parameter for which we have a complete result.

4.2 Stability for the scale parameter

In this section, we study equation (28) for the scale parameter which is given by an autonomous recursive equation.

Let us consider the function defined on $[0, +\infty)$ (see (98)-(99)- (100)):

$$F_z^{\theta}(v) = \beta^2 + a^2\, \frac{vz^2}{z^2 + 2\lambda\, v}. \tag{29}$$

which is the expression of (28) for the square of the scale parameter. For each θ, we look for a strictly stationary process $(V_n, n \in Z\!\!\!Z)$, defined on

$(\Omega, \mathcal{A}, \mathbb{P}_{\theta_0})$, solution of

$$\forall n \in \mathbb{Z}, V_n = F^{\theta}_{Z_n}(V_{n-1}) \quad (V_n = \sigma_n^2). \tag{30}$$

To study (30), we use the approach presented in Bougerol (1993) to obtain such a solution by a general theorem on iterations of Lipschitz random functions on a complete separable metric space. The result proved below (Proposition 4.1) is an extension of Proposition 3.1 of Genon-Catalot and Kessler (2003).

Define the closed interval:

$$I(\theta) = [\beta^2, \sigma^2(\theta) = \frac{\beta^2}{1-a^2}]. \tag{31}$$

Lemma 4.2. *For $z \neq 0$, the function F^{θ}_z is increasing from $I(\theta)$ onto $I(\theta)$ and is Lipschitz with $0 \leq \frac{dF^{\theta}_z}{dv} \leq a^2$. For $z = 0$, $F^{\theta}_0(v) = \beta^2$ for all v.*

Proof. We have

$$\frac{dF^{\theta}_z}{dv}(v) = \frac{a^2 z^4}{(z^2 + 2\lambda\, v)^2}.$$

For all $v \in I(\theta)$, the following holds:

$$\beta^2 \leq F^{\theta}_z(v) \leq F^{\theta}_z(\frac{\beta^2}{1-a^2}) \leq \frac{\beta^2}{1-a^2}.$$

So, the proof is complete □

Remark.

1) Since $I(\theta)$ is bounded from below by β^2, for all ν in the SG-class, the image $\Phi^{\theta}_z(\nu)$ has positive scale parameter and, hence, has a density.

2) From lemma 4.2, we see that, for all $v \in I(\theta)$, and in particular for $v = \frac{\beta^2}{1-a^2} = \sigma^2(\theta)$,

$$\frac{\beta^2}{F^{\theta}_z(v)} \in [1-a^2, 1] \quad \text{and} \quad 1 - \frac{\beta^2}{F^{\theta}_z(v)} \in [0, a^2]. \tag{32}$$

The above intervals are very small when a is close to 0, which is a situation close to independence. This remark will be of use below to build approximations of the exact likelihood.

For $z = (z_n, n \in \mathbb{Z}) \in \mathbb{R}^{\mathbb{Z}}$ and $n \in \mathbb{Z}$, let us denote by $\underline{z}_n = (z_n, z_{n-1}, \ldots)$ the element of $\mathbb{R}^{\mathbb{N}}$ which is the infinite past of z starting from n.

Proposition 4.1. *For all θ (with $a^2 < 1$), there exists, on $(\Omega, \mathcal{A}, I\!P_{\theta_0})$, a measurable function $V(\theta, \underline{Z}_n)$, such that the process*

$$(\bar{V}_n = V(\theta, \underline{Z}_n), n \in \mathbb{Z}) \tag{33}$$

is the unique strictly stationary solution process of (30). This process takes values in $I(\theta)$ (see (31)), is ergodic and

1) *For all $v_0 \in I(\theta)$, $n \in \mathbb{Z}$, as p tends to $+\infty$,*

$$V_{n:n-p}(\theta, v_0) = F^{\theta}_{Z_n} \circ F^{\theta}_{Z_{n-1}} \circ \ldots \circ F^{\theta}_{Z_{n-p}}(v_0) \to V(\theta, \underline{Z}_n),$$

almost surely under $I\!P_{\theta_0}$.

2) *Moreover, as n tends to $+\infty$,*

$$V_{n:1}(\theta, v_0) = F^{\theta}_{Z_n} \circ F^{\theta}_{Z_{n-1}} \circ \ldots \circ F^{\theta}_{Z_1}(v_0)$$

satisfies

$$V_{n:1}(\theta, v_0) - V(\theta, \underline{Z}_n) \to 0,$$

almost surely under $I\!P_{\theta_0}$.

3) *The function $\theta \to V(\theta, \underline{Z}_0)$ is $I\!P_{\theta_0}$-a.s. continuous on Θ.*

4) *Let $\mu_{(\theta_0,\theta)}(du, dz, dv)$ denote the distribution, under $I\!P_{\theta_0}$, of $(U_1, Z_1, V(\theta, \underline{Z}_0))$. Then, $(U_n, Z_n, V_{n-1:1}(\theta, v_0))$ converges in distribution under $I\!P_{\theta_0}$ to $\mu_{(\theta_0,\theta)}$ as n tends to $+\infty$.*

Proof. To obtain 1), we apply Theorem 3.1, p.955 of Bougerol, 1993 and its Corollaries 3.2-3.3. By Lemma 4.2, $(F^{\theta}_{Z_n})$ is an ergodic sequence of Lipschitz maps on the separable complete metric space $I(\theta)$ and the Lipschitz coefficient of $(F^{\theta}_{Z_n})$ is $\rho(F^{\theta}_{Z_n}) = a^2$. To get our result, we only need to check that, for some $v \in I(\theta)$, $I\!E_{\theta_0} \log^+ |F^{\theta}_{Z_1}(v) - v|$ is finite. This follows from the fact that the interval $I(\theta)$ is bounded from below by $\beta^2 > 0$. Moreover, the following holds

$$|V_{n:1}(\theta, v_0) - V(\theta, \underline{Z}_n)| \le a^{2n}|v_0 - V(\theta, \underline{Z}_0) \le a^{2n}\frac{2\beta^2}{1-a^2}. \tag{34}$$

since

$$V(\theta, \underline{Z}_n) = F^{\theta}_{Z_n} \circ F^{\theta}_{Z_{n-1}} \circ \ldots \circ F^{\theta}_{Z_1}(V(\theta, \underline{Z}_0)).$$

This gives both points 2) and 3). Since (U_n, Z_n) is strictly stationary, point 4) is a simple consequence of point 2). The proof is now complete. □

We are now able to conclude.

Corollary 4.1. *1) As n tends to $+\infty$,*

$$\sigma^2_{0:(-n+2)}(\theta) = F^\theta_{Z_0} \circ F^\theta_{Z_{-1}} \circ \ldots \circ F^\theta_{Z_{-n+2}}(\sigma^2(\theta)) \tag{35}$$

tends to $V(\theta, \underline{Z}_0)$, almost surely under $I\!P_{\theta_0}$.

2) As n tends to $+\infty$,

$$\sigma^2_{n:1}(\theta) = F^\theta_{Z_n} \circ F^\theta_{Z_{n-1}} \circ \ldots \circ F^\theta_{Z_1}(\sigma^2(\theta)) \tag{36}$$

defined in (16), which is the square of the scale parameter of (10) (see also (19)) satisfies, almost surely under $I\!P_{\theta_0}$:

$$\sigma^2_{n:1}(\theta) - V(\theta, \underline{Z}_n) \to 0. \tag{37}$$

Proof. It is an immediate consequence of Proposition 4.1 1) and 2) with $v_0 = \sigma^2(\theta) \in I(\theta)$ □

4.3 Stability at the true value of the parameter.

We need to prove the existence of stationary solution for (28), *i.e.* a solution involving both the scale and the mixture parameter. Actually, we only have the result at the true value θ_0. This was proved in Genon-Catalot and Kessler (2003). We recall the results there obtained.

First note that

$$\nu^{\theta_0}_{n|n-1:1} = \Phi^{\theta_0}_{Z_{n-1}} \circ \ldots \circ \Phi^{\theta_0}_{Z_1}(\nu^{\theta_0}). \tag{38}$$

is the conditional distribution, under $I\!P_{\theta_0}$, of U_n given $Z_{n-1}, \ldots, Z_1$. Similarly, the conditional distribution of U_1 given $Z_0, Z_{-1} \ldots, Z_{-n+2}$ (under $I\!P_{\theta_0}$) is

$$\nu^{\theta_0}_{1|0:(-n+2)} = \Phi^{\theta_0}_{Z_0} \circ \Phi^{\theta_0}_{Z_{-1}} \circ \ldots \circ \Phi^{\theta_0}_{Z_{-n+2}}(\nu^{\theta_0}). \tag{39}$$

Proposition 4.2. *The sequence of random probability measures (39)) weakly converges, as n tends to $+\infty$, $I\!P_{\theta_0}$-a.s., to a probability measure $\nu^{\theta_0}_{1|-\infty}$. This distribution is SG with scale parameter $V(\theta_0, \underline{Z}_0)^{1/2}$ (obtained in Proposition 4.1) and has a mixture parameter $\alpha(\theta_0, \underline{Z}_0)$ which is a measurable function of $\underline{Z}_0$. Moreover, setting*

$$\forall n \in Z\!\!\!Z, \nu_n(\theta_0) = SG(V(\theta_0, \underline{Z}_{n-1})^{1/2}, \alpha(\theta_0, \underline{Z}_{n-1})) \tag{40}$$

then, for all $n \in \mathbb{Z}$, almost surely under $\mathbb{P}_{\theta_0}$,

$$\nu_{n+1}(\theta_0) = \Phi^{\theta_0}_{Z_n}(\nu_n(\theta_0)). \tag{41}$$

The process $(\nu_n(\theta_0), n \in \mathbb{Z})$ with values in the class of SG distributions is the unique strictly stationary and ergodic solution of (28) at θ_0.

Remark.

1) This result relies on several properties and tools: first, the mappings Φ^θ_z are continuous with respect to the topology of weak convergence of probability measures; second, the SG-class is, in some sense, closed under weak convergence; finally, because $\nu^{\theta_0}_{1|0:(-n+2)}$ is the conditional distribution of U_1 given $(Z_0, \ldots, Z_{-n+2}$, it is possible to use a martingale limit theorem. This tool is not at hand to study $\nu^{\theta}_{1|0:(-n+2)}$ under $\mathbb{P}_{\theta_0}$.

2) For each n, $\nu_n(\theta_0)$ (see (40)) is the conditional distribution of U_n given $\underline{Z}_{n-1}$ under $\mathbb{P}_{\theta_0}$. On $(\Omega, \mathcal{A}, \mathbb{P}_{\theta_0})$, the process $(U_n, Z_n, \nu_n(\theta_0)_{n\in\mathbb{Z}}$ is strictly stationary and ergodic. So, we have obtained a stationary regime for the Markov process $(U_n, Z_n, \Phi^{\theta_0}_{Z_{n-1}} \circ \ldots \circ \Phi^{\theta_0}_{Z_1}(\nu^{\theta_0}), n \geq 1)$.

The following consequence of Proposition 4.2 is useful for the sequel.

Proposition 4.3. *For all integer k,*

$$\mathbb{E}_{\theta_0}(\sum_{j\geq 0} j^k \alpha_j(\theta_0, \underline{Z}_0) < +\infty. \tag{42}$$

Proof. Let $k \geq 1$. Recall that C_{2k} denotes the $2k$-th order moment of a standard Gaussian variable and has the explicit expression:

$$C_{2k} = \frac{(2k)!}{2^k k!}. \tag{43}$$

We have

$$\mathbb{E}_{\theta_0} U_1^{2k} = C_{2k}\sigma^{2k}(\theta_0), \tag{44}$$

where $\sigma^2(\theta_0)$ is the variance of the stationary distribution of (U_n) (see (6)).

Now, by Proposition 4.2, the conditional distribution of U_1 given $\underline{Z}_0$ is $\mathrm{SG}(V^{1/2}(\theta_0, \underline{Z}_0), \alpha(\theta_0, \underline{Z}_0)$. Using formulae (13)-(14), we get

$$\mathbb{E}_{\theta_0}(U_1^{2k}/\underline{Z}_0) = V(\theta_0, \underline{Z}_0)^k \sum_{j\geq 0} \frac{C_{2j+2k}}{C_{2j}} \alpha_j(\theta_0, \underline{Z}_0). \tag{45}$$

Since $V(\theta_0, \underline{Z}_0) \geq \beta_0^2$, taking expectations in (45) and using (44), we obtain

$$\mathbb{E}_{\theta_0}(\sum_{j\geq 0} \frac{C_{2j+2k}}{C_{2j}} \alpha_j(\theta_0, \underline{Z}_0) \leq \beta_0^{-2} C_{2k} \sigma^{2k}(\theta_0).$$

Now, by the Stirling formula, $\frac{C_{2j+2k}}{C_{2j}} \sim (2j)^k$. So, we get the result. □

4.4 Convergence of the likelihood at the true value of the parameter

We now use the previous result to derive the asymptotic behaviour of the exact likelihood at θ_0. Let

$$\tilde{p}(\theta_0, z/\underline{Z}_0) = \int_{\mathbb{R}} f_u(z) \nu^{\theta_0}_{1|0:-\infty}(du) \tag{46}$$

and

$$\tilde{P}_{\theta_0}(dz) = \tilde{p}(\theta_0, z/\underline{Z}_0)\ dz. \tag{47}$$

The above relation defines a random probability measure which is clearly a regular version of the conditional distribution, under $\mathbb{P}_{\theta_0}$, of Z_1 given $\underline{Z}_0$. More precisely, let us set (see (25) and Proposition 4.1)

$$h(\theta, z/\underline{Z}_0) = \frac{\lambda\, V(\theta, \underline{Z}_0)}{(z^2 + 2\lambda\, V(\theta, \underline{Z}_0))^{3/2}}. \tag{48}$$

This is the Student density $(\lambda\ V(\theta, \underline{Z}_0))^{1/2}\ t(2)$ (the scale parameter is random and a measurable function of $\underline{Z}_0$)). It is defined, for all θ, under $\mathbb{P}_{\theta_0}$. Now, we have (see (23) and Proposition 4.2)

$$\tilde{p}(\theta_0, z/\underline{Z}_0) = h(\theta_0, z/\underline{Z}_0) \sum_{j\geq 0} \alpha_j(\theta_0, \underline{Z}_0) \frac{(2j+1)z^{2j}}{(z^2 + 2\lambda\, V(\theta_0, \underline{Z}_0))^j}. \tag{49}$$

Let us define

$$u(\theta_0, \underline{Z}_1) = \log \tilde{p}(\theta_0, Z_1/\underline{Z}_0). \tag{50}$$

Lemma 4.3. *The following holds.*

i) $\mathbb{E}_{\theta_0} \mid u(\theta_0, \underline{Z}_1) \mid < +\infty$.

ii) Setting $H(\theta_0) = \mathbb{E}_{\theta_0} u(\theta_0, \underline{Z}_1)$, *we have*

$$H(\theta_0) = \mathbb{E}_{\theta_0} \left(\int_{\mathbb{R}} \log \tilde{p}(\theta_0, z/\underline{Z}_0) \tilde{P}_{\theta_0}(dz) \right). \tag{51}$$

Proof. i) Let us set

$$x = \frac{z^2}{(z^2 + 2\lambda\, V(\theta_0, \underline{Z}_0))} \in [0, 1).$$

Then, using that $\sum_{j\geq 0}(2j+1)x^j = \frac{1+x}{(1-x)^2}$, we obtain

$$0 < \tilde{p}(\theta_0, Z_1/\underline{Z}_0) \leq \frac{1}{(Z_1^2 + 2\lambda\, V(\theta_0, \underline{Z}_0))^{1/2}}.$$

Then, we use the bounds for $V(\theta_0, \underline{Z}_0)$ and get

$$0 < \tilde{p}(\theta_0, Z_1/\underline{Z}_0) \leq \frac{1}{(2\lambda\, \beta^2)^{1/2}}.$$

On the other hand, using $2j + 1 \geq 1$ and the concavity of the logarithm, we obtain

$$\log \tilde{p}(\theta_0, Z_1/\underline{Z}_0) \geq \tag{52}$$

$$\log h(\theta_0, Z_1/\underline{Z}_0) + \sum_{j\geq 0} j\alpha_j(\theta_0, \underline{Z}_0) \times \log\,(\frac{Z_1^2}{Z_1^2 + 2\lambda\, V(\theta_0, \underline{Z}_0)}) \tag{53}$$

The distribution of Z_1 under $\mathbb{P}_{\theta_0}$ is the $\lambda^{1/2}\sigma(\theta_0)t(2)$ distribution. Using that $V(\theta_0, \underline{Z}_0)$ belongs to $I(\theta_0)$, we immediately check that

$$\mathbb{E}_{\theta_0} \mid \log h(\theta_0, Z_1/\underline{Z}_0) \mid < +\infty$$

and

$$\mathbb{E}_{\theta_0} \mid \log\,(\frac{Z_1^2}{(Z_1^2 + 2\lambda\, V(\theta_0, \underline{Z}_0)})\mid^2 < +\infty.$$

Since, using Proposition4.3,

$$\mathbb{E}_{\theta_0}(\sum_{j\geq 0} j\alpha_j(\theta_0, \underline{Z}_0))^2 \leq \mathbb{E}_{\theta_0} \sum_{j\geq 0} j^2\alpha_j(\theta_0, \underline{Z}_0) < +\infty,$$

we get that $\log \tilde{p}(\theta_0, Z_1/\underline{Z}_0)$ is bounded from below by an integrable random variable which implies the result.
ii) Using that $\tilde{P}_{\theta_0}$ is the conditional distribution of Z_1 given $\underline{Z}_0$, we obtain

$$\mathbb{E}_{\theta_0}(u(\theta_0, /\underline{Z}_1)//\underline{Z}_0)) = \int_{\mathbb{R}} \log \tilde{p}(\theta_0, z/\underline{Z}_0)\tilde{P}_{\theta_0}(dz). \tag{54}$$

This implies (51). □

Thus, $E(\theta_0) = -H(\theta_0)$ is the expectation of the usual entropy of the distribution $\tilde{P}_{\theta_0}(dz)$. Now, we are able to prove the following.

Theorem 4.1. *We have* $\frac{1}{n}\log L_n(\theta_0) \to H(\theta_0)$ *defined in (51),* $\mathbb{P}_{\theta_0}$*-a.s..*

Proof. By the ergodic theorem, which can be applied by Lemma 4.3, i), we get

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

Moreover, since, for $Z_1 \neq 0$, $u \to f_u(Z_1)$ is continuous and bounded,

$$\int f_u(Z_1)\ \nu^{\theta_0}_{1|0:(-n+2)}(du) = p_n(\theta_0, Z_1/Z_0, Z_{-1}, \ldots, Z_{-n+2})$$

converges to

$$\tilde{p}(\theta_0, Z_1/\underline{Z}_0)$$

as n tends to infinity, $\mathbb{P}_{\theta_0}$-a.s.. Hence,

$$\log p_n(\theta_0, Z_1/Z_0, Z_{-1}, \ldots, Z_{-n+2}) - \log \tilde{p}(\theta_0, Z_1/\underline{Z}_0) \to 0.$$

By the strict stationarity of $(Z_i, i \in \mathbb{Z})$, we obtain that, as $i \to +\infty$,

$$\log p_i(\theta_0, Z_i/Z_{i-1}, \ldots, Z_1) - \log \tilde{p}(\theta_0, Z_i/\underline{Z}_{i-1}) \to 0.$$

Taking Cesaro means, as $n \to +\infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \big(\log p_i(\theta_0, Z_i/Z_{i-1}, \ldots, Z_1) - \log \tilde{p}(\theta_0, Z_i/\underline{Z}_{i-1})\big) \to 0.$$

In the expression above, the first term in the first sum is taken equal to $\log p_1(\theta_0, Z_1)$. Finally,

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

All these convergences holds $\mathbb{P}_{\theta_0}$-a.s. □

5 Approximations of the likelihood.

We only have the asymptotic behaviour of the normalized log-likelihood function at the true value of the parameter. Therefore, we introduce now approximations of the exact likelihood. Previous simulation results on the mixture parameter suggest that, for small values of a, there are only one or two components of the mixture which appear to be significantly non nul. This suggests to replace the exact likelihood by an explicit function based on a fixed number of mixture components. As a substitute to the exact likelihood $L_n(\theta)$, let us consider the following functions. The first one is based on a one-component mixture and is given by (see (22)-(25))

$$L_n^{(1)}(\theta) = h_1(\theta, Z_1) \times \prod_{i=2}^{n} h_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1). \tag{55}$$

The second one is based on a two-component mixture:

$$L_n^{(2)}(\theta) = g_1(\theta, Z_1) \times \prod_{i=2}^{n} g_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1), \tag{56}$$

where

$$g_1(\theta, Z_1) = h_1(\theta, Z_1) f_1(\theta, Z_1), \tag{57}$$

$$g_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) = h_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1)\ f_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) \tag{58}$$

$$f_1(\theta, Z_1) = \frac{\beta^2}{\sigma^2(\theta)} + (1 - \frac{\beta^2}{\sigma^2(\theta)}) \frac{3\ Z_1^2}{Z_1^2 + 2\lambda\sigma^2(\theta)}, \tag{59}$$

$$f_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) = \frac{\beta^2}{\sigma_{i-1:1}^2(\theta)} + (1 - \frac{\beta^2}{\sigma_{i-1:1}^2(\theta)}) \frac{3\ Z_i^2}{Z_i^2 + 2\lambda\sigma_{i-1:1}^2(\theta)}. \tag{60}$$

Indeed, the simulation results also show that the mixture parameter is close to the one used above. To make notations consistent, we set

$$\sigma_{0:1}^2(\theta) = \sigma^2(\theta).$$

We denote by $\tilde{\theta}_n^{(i)}$ the estimator obtained by maximisation of $L_n^{(i)}(\theta)$, for $i = 1, 2$.

Let us define for $n \geq 1$,

$$\nu_{n|n-1:1}^{(1)}(\theta) = \mathcal{N}(0, \sigma_{n-1:1}^2(\theta)). \tag{61}$$

This measure valued process has values in the set of centered Gaussian distributions with variances given by the recursive algorithm (99).

Consider also

$$\nu^{(2)}_{n|n-1:1}(\theta) = \mathrm{SG}(\sigma_{n-1:1}(\theta)), (\frac{\beta^2}{\sigma^2_{n-1:1}(\theta)}, 1 - \frac{\beta^2}{\sigma^2_{n-1:1}(\theta)})), \qquad (62)$$

This process has values in the subset of SG-distributions with a two-component mixture parameter. Its scale and mixture parameter are defined through (99). Changing $(Z_{n-1}, \ldots, Z_1)$ into $(Z_0, \ldots, Z_{-n+2})$, we may set

$$\nu^{(1)}_{1|0:(-n+2)}(\theta) = \mathcal{N}(0, \sigma^2_{0:(-n+2)}(\theta)) \qquad (63)$$

and

$$\nu^{(2)}_{1|0:(-n+2)}(\theta) = \mathrm{SG}(\sigma_{0:(-n+2)}(\theta)), (\frac{\beta^2}{\sigma^2_{0:(-n+2)}(\theta)}, 1 - \frac{\beta^2}{\sigma^2_{0:(-n+2)}(\theta)})). \qquad (64)$$

Due to Proposition 4.1, the following random probability measures are well defined for all θ:

$$\nu^{(1)}_{1|0:-\infty}(\theta) = \mathcal{N}(0, V(\theta, \underline{Z}_0)) \qquad (65)$$

$$\nu^{(2)}_{1|0:(-\infty)}(\theta) = \mathrm{SG}(V^{1/2}(\theta, \underline{Z}_0)), (\frac{\beta^2}{V(\theta, \underline{Z}_0)}, 1 - \frac{\beta^2}{V(\theta, \underline{Z}_0)})). \qquad (66)$$

We have the following result.

Proposition 5.1. *For all θ and for $i = 1, 2$, $\nu^{(i)}_{1|0:(-n+2)}(\theta)$ converges weakly, $I\!P_{\theta_0}$-a.s., to $\nu^{(i)}_{1|0:(-\infty)}(\theta)$ as n tends to $+\infty$.*

The above proposition follows from a simple application of Corollary 4.1. The essential difference with Proposition 4.2 is that the above convergences hold, not only at θ_0, but for all θ.

Let us also notice that (see (26)), for $i = 1, 2$,

$$\frac{1}{n} \log L^{(i)}_n(\theta) = \frac{1}{n} \sum_{j=1}^{n} l(Z_j, \nu^{(i)}_{j|j-1:1}(\theta)). \qquad (67)$$

The above equalities are analogous to (27). In analogy with (46)-(47), let us define the following random probability measures:

$$\tilde{p}^{(i)}(\theta, z/\underline{Z}_0) = \int_{I\!R} f_u(z) \nu^{(i)}_{1|0:(-\infty)}(\theta)(du) \qquad (68)$$

and

$$\tilde{P}^{(i)}_\theta(dz) = \tilde{p}^{(i)}(\theta, z/\underline{Z}_0)\, dz. \qquad (69)$$

Thus, we have (see (48) and Proposition 3.1)

$$\tilde{p}^{(1)}(\theta, z/\underline{Z}_0) = \int_{\mathbb{R}} f_u(z) \nu^{(1)}_{1|0:(-\infty)}(\theta)(du) = h(\theta, z/\underline{Z}_0), \tag{70}$$

and we now compute

$$\tilde{p}^{(2)}(\theta, z/\underline{Z}_0) = \int_{\mathbb{R}} f_u(z) \nu^{(2)}_{1|0:(-\infty)}(\theta)(du) = g(\theta, z/\underline{Z}_0), \tag{71}$$

where

$$g(\theta, z/\underline{Z}_0) = h(\theta, z/\underline{Z}_0) \times f(\theta, z/\underline{Z}_0) \tag{72}$$

and

$$f(\theta, z/\underline{Z}_0) = \frac{\beta^2}{V(\theta, \underline{Z}_0)} + (1 - \frac{\beta^2}{V(\theta, \underline{Z}_0)}) \frac{3\, z^2}{z^2 + 2\lambda V(\theta, \underline{Z}_0)}. \tag{73}$$

Finally, let us set

$$u^{(1)}(\theta, \underline{Z}_1) = \log h(\theta, Z_1/\underline{Z}_0) \tag{74}$$

$$u^{(2)}(\theta, \underline{Z}_1) = \log g(\theta, Z_1/\underline{Z}_0). \tag{75}$$

We have the following result.

Theorem 5.1. *For all θ and for $i = 1, 2$,*

1) $\mathbb{E}_{\theta_0} \mid u^{(i)}(\theta, \underline{Z}_1) \mid < \infty.$

2)

$$\frac{1}{n} L_n^{(i)}(\theta) \to \mathbb{E}_{\theta_0} u^{(i)}(\theta, \underline{Z}_1)$$

almost surely under $\mathbb{P}_{\theta_0}$-a.s., as n tends to $+\infty$.

3)

$$\mathbb{E}_{\theta_0} u^{(i)}(\theta, \underline{Z}_1) = H(\theta_0) - \mathbb{E}_{\theta_0}(K(\tilde{P}_{\theta_0}, \tilde{P}_\theta^{(i)})),$$

where $K(P, Q)$ denotes the Kullback information of P with respect to Q.

Proof. We only study the case $i = 1$, the other case being analogous. Using that $V(\theta, \underline{Z}_0)$ belongs to $I(\theta)$, we obtain

$$c - \log(Z_1^2 + c') \le u^{(1)}(\theta, \underline{Z}_1) \le C,$$

where c, c', C are constants depending on the bounds of the interval $I(\theta)$. Under $\mathbb{P}_{\theta_0}$, Z_1 has distribution $\lambda^{1/2}\sigma(\theta_0)t(2)$. Hence, $\mathbb{E}_{\theta_0} \mid \log{(Z_1^2 + c')} \mid < \infty$. So, we get 1).
By the ergodic theorem, the following convergence holds $\mathbb{P}_{\theta_0}$-a.s.:

$$\frac{1}{n}\sum_{i=1}^{n} u^{(1)}(\theta, \underline{Z}_i) \to \mathbb{E}_{\theta_0} u^{(1)}(\theta, \underline{Z}_1). \tag{76}$$

Using the explicit expressions (25)-(48), some computations show that

$$\mid \log h_i(\theta, Z_i/Z_{i-1}, \ldots, Z_1) - u^{(1)}(\theta, \underline{Z}_i) \mid \leq C \mid \sigma^2_{n-1:1}(\theta) - V(\theta, \underline{Z}_i) \mid,$$

where C is a constant depending on θ. By Corollary 4.1, the right-hand side above tends to 0 as i tends to infinity. Taking Cesaro means and using (76), we get 2).
As for (54)-(51), we have

$$\mathbb{E}_{\theta_0}(u^{(1)}(\theta, \underline{Z}_1)/\underline{Z}_0) = \int_{\mathbb{R}} \log \tilde{p}^{(1)}(\theta, z/\underline{Z}_0)\tilde{P}_{\theta_0}(dz). \tag{77}$$

Therefore,

$$\mathbb{E}_{\theta_0}(u^{(1)}(\theta, \underline{Z}_1) - u(\theta_0, \underline{Z}_1)/\underline{Z}_0) = -K(\tilde{P}_{\theta_0}, \tilde{P}_{\theta}^{(1)}). \tag{78}$$

This gives the result. □

By Theorem 5.1, we know that, for all θ,

$$\mathbb{E}_{\theta_0}(u^{(i)}(\theta, \underline{Z}_1) - u(\theta_0, \underline{Z}_1)) = -\mathbb{E}_{\theta_0}(K(\tilde{P}_{\theta_0}, \tilde{P}_{\theta}^{(i)}) \leq 0. \tag{79}$$

To derive consistency of the estimators $\tilde{\theta}_n^{(i)}$ based on $L_n^{(i)}(\theta)$, the above result is not enough. We should also prove that, for $i = 1, 2$, the function

$$\theta \to \mathbb{E}_{\theta_0}(K(\tilde{P}_{\theta_0}, \tilde{P}_{\theta}^{(i)})$$

admits a *unique* minimum at some value $\theta_0^{(i)}$. Ideally, this value should be equal to θ_0. In this case, using a general theorem of consistency such as the one proved in Genon-Catalot et al (2003), we would be able to conclude that the estimators $\tilde{\theta}_n^{(i)}$ converge to $\theta_0^{(i)}$. Nevertheless, as we will show via numerical simulations, estimation by exact maximum likelihood and by using $L_n^{(2)}(\theta)$ yield very close results.

6 Ryden's method

Because exact likelihoods of hidden Markov models are generally rather difficult to study and to compute, Ryden (1994) suggests a class of estimators which are consistent and asymptotically normal in the context where the hidden chain has a finite number of states. The method is called maximum likelihood split data (MLSD) and is based on a simple and natural idea. Suppose that the number of data is nm, *i.e.* a multiple of some integer m. For small m, the joint density $p_m(\theta, z_1, \dots, z_m)$ of $(Z_1, \dots, Z_m)$ is not so difficult to derive even in general models. Therefore, Ryden's method is to replace the exact likelihood by the likelihood of a n-sample of m-dimensional variables having density $p_m(\theta, z_1, \dots, z_m)$. Ryden's likelihood is thus given (for nm data) by

$$\mathcal{L}_n^{(m)}(\theta) = \prod_{k=1}^{n} p_m(\theta, Z_{m(k-1)+1}, \dots, Z_{mk}). \tag{80}$$

A simple application of the ergodic theorem implies that, $\mathbb{P}_{\theta_0}$-a.s.,

$$\frac{1}{n} \log \mathcal{L}_n^{(m)}(\theta) \to \mathbb{E}_{\theta_0} \log p_m(\theta, Z_1, \dots, Z_m). \tag{81}$$

Then, under the appropriate identifiability assumption, *i.e.* that the above limit has a unique maximum at $\theta = \theta_0$, a standard proof of consistency for maximum contrast estimators implies the consistency of Ryden's MLSD estimator. The method has advantadges and drawbacks. The advantadge is that it leads to consistent estimators. The drawback is that, for small m, Ryden's likelihood may be very far from the exact likelihood. Indeed, the numerical simulation results in his paper mention that for $m \geq 20$, the maximum likelihood split data estimator is as good as the exact maximum likelihood estimator. But then, the method is as intractable as the exact likelihood method.

We will illustrate Ryden's method with $m = 1$ and $m = 2$. Let us set

$$\mathcal{L}_n^{(1)}(\theta) = \prod_{k=1}^{n} p_1(\theta, Z_i), \tag{82}$$

where $p_1(\theta, z)$ is the common density of the Z_i's and is given in (22). It is immediate to check that, using this MLSD, only one parameter will be identified, namely $\sigma^2(\theta)$, the stationary variance of (U_n).

Secondly, consider that we have $2n$ data, and set

$$\mathcal{L}_n^{(2)}(\theta) = \prod_{k=1}^{n} p_2(\theta, Z_{2k-1}, Z_{2k}). \tag{83}$$

We will now give the explicit expression of $\mathcal{L}_n^{(2)}(\theta)$. By (23), we have:

$$p_2(\theta, Z_1, Z_2) = p_1(\theta, Z_1) \times p_2(\theta, Z_2/Z_1). \tag{84}$$

To compute $p_2(\theta, Z_2/Z_1)$, we need $\sigma^2_{1:1}(\theta)$ and $\alpha^{1:1}(\theta)$ which is only composed of two non nul values. These are easily obtained using the formulae of the appendix. After some computations, we get

$$\sigma^2_{1:1}(\theta) = \sigma^2(\theta)\frac{Z_1^2 + 2\lambda\beta^2}{Z_1^2 + 2\lambda\sigma^2(\theta)}, \tag{85}$$

$$\alpha_0^{1:1}(\theta) = \frac{\beta^2}{\sigma^2_{1:1}(\theta)} = (1-a^2)\frac{Z_1^2 + 2\lambda\sigma^2(\theta)}{Z_1^2 + 2\lambda\beta^2}, \tag{86}$$

$$\alpha_1^{1:1}(\theta) = 1 - \frac{\beta^2}{\sigma^2_{1:1}(\theta)} = a^2\frac{Z_1^2}{Z_1^2 + 2\lambda\beta^2}. \tag{87}$$

Now, using (83), it is easily seen that we can identify two parameters, namely β^2 and a^2.

7 Concluding remarks and open problems

In this paper, we have studied a hidden Markov model for which the exact likelihood is simple and explicit. the exact maximum likelihood can be easily computed and performs well on numerical simulated data. We have obtained the limit of the normalized log-likelihood function at the true value θ_0 of the parameter. For other values of the parameter, a difficulty arises which is specific to hidden Markov models. The likelihood at θ_0 may be studied by means of a theoretical tool, namely the conditional likelihood with respect to the infinite past $\underline{Z}_0$ of the observations. This process is defined by

$$\prod_{i=1}^{n} \tilde{p}(\theta_0, Z_i/\underline{Z}_{i-1}).$$

The random variable $\tilde{p}(\theta_0, Z_i/\underline{Z}_{i-1})$ is defined $\mathbb{P}_{\theta_0}$-a.s. It depends on the whole set of coordinates $\underline{Z}_i$. Since the distributions of $\underline{Z}_i$ under $\mathbb{P}_{\theta_0}$ and under $\mathbb{P}_\theta$ are not absolutely continuous, we cannot deduce the existence of $\tilde{p}(\theta, Z_i/\underline{Z}_{i-1})$ under $\mathbb{P}_{\theta_0}$ from its existence under $\mathbb{P}_\theta$ as it is the case when only a finite number of coordinates are involved.

The processes introduced in Section 5 to approximate the exact likelihood are well defined for all θ and seem rather close to the exact likelihood. They provide simple and explicit estimators which perform as well as the exact maximum likelihood estimator for small values of a. Still, open problems remain since we have not proved that the functions

$$\theta \rightarrow \mathbb{E}_{\theta_0}(K(\tilde{P}_{\theta_0}, \tilde{P}_\theta^{(i)}))$$

have a unique minimum.

Acknowledgments. 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.

References

[1] Azencott R. and Dacunha-Castelle D., 1984, Séries d'observations irrégulières. Modélisation et prévision. *Masson, Paris.*

[2] Bickel P.J., Ritov Y. and Ryden T., (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Annals of Stat.* **26**, 1614-1635.

[3] Bougerol P. (1993). Kalman filtering with random coefficients and contractions. *Siam. J. Control and Optimization* **31** (4), 942-959.

[4] Douc P. and Matias L. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7** (3), 381-420.

[5] Elliott, R.J., Lakhdar A. and Moore J.B. (1995). *Hidden Markov models: Estimation and control*, Application of Mathematics, Berlin: Springer.

[6] Genon-Catalot V., (2002), A non linear explicit filter, *Statistics and probability letters.*

[7] Genon-Catalot V., Jeantheau T. and Larédo C. (2003). Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scand. J. Statist.*, **30** (2), 297-316.

[8] Genon-Catalot V., Kessler M. (2003), Random scale perturbation of an AR(1) process and its properties as a non linear explicit filter, to appear in *Bernoulli*.

[9] Jensen J.L. and Petersen N.V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Stat.* **27** (2), 514-535.

[10] Leroux B.G. (1992). Maximum likelihood estimation for hidden Markov models. *Stoch. Proc. and their appl.* **40**, 127-143.

[11] Ryden T., (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.* **22**, 1884-1895.

8 Appendix

8.1 Formulae for the filtering algorithm

We briefly recall results of Genon-Catalot (2003) and Genon-Catalot and Kessler (2003) useful for the present paper. In filtering theory, it is well known that, given an initial distribution for U_1, it is possible to compute successively the following conditional distributions:

$$L(U_1) \xrightarrow{\text{updating}} L(U_1|Z_1) \xrightarrow{\text{prediction}} L(U_2|Z_1) \xrightarrow{\text{updating}} L(U_2|Z_2, Z_1) \xrightarrow{\text{prediction}} L(U_3|Z_2, Z_1) \ldots.$$

These distributions are however generally not explicit: the problem of the exact computation becomes quickly intractable unless these distributions all belong to the same parametric family. In this case, it is enough to express the algorithm in terms of the parameters. It is the case with the present model. When U_1 belongs to the class $\mathcal{F}$ of SG distributions, then, the above conditional distributions all belong to the SG class. We recall here the complete definition of this class.

Definition 8.1. *The class $\mathcal{F}$ of SG distributions consists of all the distributions $\nu = \nu_{\sigma,\alpha}$, where $\sigma \geq 0$ and $\alpha = (\alpha_i, i \geq 0)$ is a series of weight coefficients, $\forall i \geq 0, \alpha_i \geq 0$ and $\sum_{i\geq 0} \alpha_i = 1$, defined by*

a) *If $\sigma = 0$, for any α, we set $\nu_{0,\alpha}(du) = \delta_0(du)$.*

b) *When* $\sigma > 0$, $\nu(du) = \nu_{\sigma,\alpha}(du) = g(u)du$, *with*

$$g(u) = \sum_{i\geq 0} \alpha_i \frac{1}{\sigma} g_i(\frac{u}{\sigma}) \tag{88}$$

where

$$g_i(u) = (2\pi)^{-1/2} \frac{u^{2i}}{C_{2i}} \exp(-\frac{u^2}{2}), \tag{89}$$

and $C_{2i} = E(X^{2i})$, *for* X *a standard Gaussian variable.*

The Serial Gaussian distribution with parameters σ and $\alpha = (\alpha_i, i \geq 0)$ is abbreviated by $\mathrm{SG}(\sigma, \alpha)$.

Thus, for positive σ, a $\mathrm{SG}(\sigma, \alpha)$ distribution is a mixture distribution which is specified by a scale parameter σ and a mixture parameter α. Each SG distribution is symmetric and $\mathcal{F}$ contains the centered Gaussian laws. By standard series expansion, it can also be seen to contain the symmetric mixture of Gaussian distributions, $\frac{1}{2}\mathcal{N}(-m, \sigma^2) + \frac{1}{2}\mathcal{N}(m, \sigma^2)$.

Let us set, for $n \geq 1$:

$$\nu^\theta_{n|n:1}(du) = L(U_n \mid Z_n, \ldots, Z_1 / \mathbb{P}_\theta). \tag{90}$$

$$\nu^\theta{}_{n+1|n:1}(du) = L(U_{n+1} \mid Z_n, \ldots, Z_1 / \mathbb{P}_\theta) \tag{91}$$

The measure valued process $(\nu^\theta{}_{n+1|n:1})$ is called the prediction filter and was defined in (10). The process $(\nu^\theta_{n|n:1})$ is the optimal filter. The updating and prediction steps of the filtering algorithm can be described introducing the following operators. Let $\mathcal{P}(\mathbb{R})$ denote the set of probability measures on $\mathbb{R}$. For $\nu \in \mathcal{P}(\mathbb{R})$, the probability measure $\varphi_z(\nu)$ is defined, for any Bounded Borel function h on $\mathbb{R}$, by

$$\varphi_z(\nu)h = \frac{\nu(f_.(z)h)}{\nu(f_.(z))}, \tag{92}$$

with the convention that $0/0 = 0$. It is the posterior distribution on u when the prior is ν and the observation is $Z = u\varepsilon_1$. Notice that the operator φ_z does not depend on the unknown parameter θ but only on the distribution of the noise (ε_n). On the other hand, for $\mu \in \mathcal{P}(\mathbb{R})$, the probability measure $\psi^\theta(\mu) = \mu\, P^\theta$ is defined by

$$\psi^\theta(\mu)(h) = \mu\, P^\theta h = \int_{\mathbb{R}\times\mathbb{R}} p(\theta, u, u')h(u')\mu(du)du'. \tag{93}$$

(We have denoted by P^θ the one-step transition operator of the AR(1) process and $p(\theta, u, u')$ is the associated transition density). The filtering algorithm then writes for $n \geq 1$

$$\begin{aligned} \textit{updating} &: \quad \nu^\theta_{n|n:1} = \varphi_{Z_n}(\nu^\theta_{n|n-1:1}) \\ \textit{prediction} &: \quad \nu^\theta_{n+1|n:1} = \psi^\theta(\nu^\theta_{n|n:1}). \end{aligned}$$

To initiate the algoritihm, we set

$$\nu^\theta_{1|0:1}) = \nu^\theta \tag{94}$$

where ν^θ is defined in (5).

Let us denote by

$$\Phi^\theta_z = \psi^\theta \circ \varphi_z \tag{95}$$

the resulting composition (see also (18). The mappings (92)-(93)-(95) are from the SG class into the SG class and therefore can be expressed in terms of the parameters. We shall use the same notation Φ^θ_z to denote the application

$$(\sigma^2, \alpha) \to \Phi^\theta_z(\sigma^2, \alpha) = (\bar{\sigma}^2(z), \bar{\alpha}(z)) \tag{96}$$

which specifies the parameters.

Thus, under $\mathbb{P}_\theta$, the conditional distribution of U_n given $(Z_{n-1}, \ldots, Z_1)$ is equal to

$$\nu^\theta_{n|n-1:1} = \Phi^\theta_{Z_{n-1}} \circ \ldots \circ \Phi^\theta_{Z_1}(\nu^\theta). \tag{97}$$

The distribution ν^θ of U_1 is centered Gaussian (see (5)-(6)). Therefore, it is the SG distribution with scale parameter (6) and mixture parameter $\alpha^{(s)}$ specified by $\alpha^{(s)}_0 = 1$.

We have,

$$(\sigma^2_{n-1:1}(\theta), \alpha^{n-1:1}(\theta)) = \Phi^\theta_{Z_{n-1}} \circ \ldots \circ \Phi^\theta_{Z_1}(\sigma^2(\theta), \alpha^{(s)}), \tag{98}$$

Because the scale parameter is ruled by an autonomous algorithm, we use a special notation for it and set

$$\bar{\sigma}^2(z) = F^\theta_z(\sigma^2), \tag{99}$$

where

$$F^\theta_z(\sigma^2) = \beta^2 + a^2 \frac{\sigma^2 \; z^2}{z^2 + 2\lambda \, \sigma^2}. \tag{100}$$

Thus

$$\sigma^2_{n-1:1}(\theta) = F^\theta_{Z_{n-1}} \circ \ldots \circ F^\theta_{Z_1}(\sigma^2(\theta)) \tag{101}$$

the square of the scale parameter obtained for $\nu^{\theta}_{n|n-1:1}$. Formula (100) shows that $\sigma^2_{n-1:1}(\theta) \geq \beta^2 > 0$.

The evolution of the mixture parameters are given by somehow more intricate relations. The mixture coefficient $\bar{\alpha}(z)$ depends both on σ^2 and α as follows.

For $k \geq 0$

$$\bar{\alpha}_k(z) = \left(1 - \frac{\beta^2}{\bar{\sigma}^2(z)}\right)^k \sum_{i \geq k} \binom{i}{k} \left(\frac{\beta}{\bar{\sigma}(z)}\right)^{2(i-k)} \hat{\alpha}_i(z), \tag{102}$$

where if $\sigma\ z \neq 0$,

$$\hat{\alpha}_0(z) = 0, \quad \hat{\alpha}_i(z) = \alpha_{i-1} \frac{h_{i-1}(\frac{z}{\lambda^{1/2}\sigma})}{h(\frac{z}{\lambda^{1/2}\sigma})},\ i \geq 1, \tag{103}$$

where

$$h(z) = \sum_{i \geq 0} \alpha_i h_i(z), \quad h_i(z) = \frac{(2i+1)z^{2i}}{(z^2+2)^{i+3/2}}. \tag{104}$$

If $\sigma z = 0$, $\bar{\alpha}_0(z) = 1$.

Moreover for the updating step $\varphi_z(\sigma^2, \alpha) = (\hat{\sigma}^2(z), \hat{\alpha}(z))$ with $\hat{\alpha}(z)$ is given above and

$$\hat{\sigma}^2(z) = \sigma^2 \frac{z^2}{z^2 + 2\lambda\sigma^2}. \tag{105}$$

From (103)-(102), we see that if $\alpha_i = 0, i \geq p$, then $\hat{\alpha}_i(z) = \bar{\alpha}_i(z) = 0, i \geq p+1$ (see (20).

8.2 Proof of Lemma 4.1

Proof. We only need to prove that $(z, \nu) \to \nu(f_{.}(z))$ is measurable. For all $z \neq 0$, the function $u \to f_u(z)$ is continuous and bounded on $\mathbb{R}$. Therefore, if ν_n converges wealy to ν, $\nu_n(f_{.}(z))$ converges to $\nu(f_{.}(z))$. Moreover, it is easily seen that, if $z_n, z > 0$ (resp. $z_n, z < 0$) and $z_n \to z$, then

$$\sup_{u \in \mathbb{R}} |f_u(z_n) - f_u(z)| \to 0.$$

And, $\nu(f_{.}(0)) = 0$. Hence, $\nu(f_{.}(z))$ is equal to a continuous function of (z, ν) on $\{z > 0\}$, $\{z < 0\}$, $\{z = 0\}$. □