The likelihood ratio test for the number of components in a mixture with Markov regime.

E. Gassiat* and C. Keribin†

Abstract: We study the LRT statistic for testing a mixture of two populations with Markov regime against a single population in the mixture. We prove that the LRT statistic converges to infinity in probability as the number of observations tends to infinity. This is a consequence of a convergence result of the LRT statistic for a subproblem where the parameters are restricted to a subset of the whole parameter set.

A.M.S. Classification: 62 F 05, 62 H 30, 62 A 10.

Key Words: Likelihood Ratio Test; Hidden Markov Model; Order of a Mixture.

*Laboratoire Modélisation Stochastique et Statistique, Université d'Orsay, Bat. 425, 91 405 Orsay, France,

†Laboratoire Analyse et Probabilité, Université d'Evry-Val d'Essonne, Boulevard des Coquibus, 91 025 Evry.

1 Introduction

Hidden Markov Models have been used for a long time as a useful modelling tool. Applications include speech processing (Rabiner, 1989), neurophysiology (Fredkin, Rice, 1992), biology (Churchill, 1989), see also Mac Donald and Zucchini (1997). A hidden Markov model (HMM) is a discrete time stochastic process $(Y_k)_{k\in\mathbb{N}}$ such that the distribution of the process may be well explained through a non observable Markov chain $(X_k)_{k\in\mathbb{N}}$ taking values in a finite state space. Given (X_k), (Y_k) is a sequence of independent random variables, and the distribution of Y_n given (X_k) is that of Y_n given X_n. Thus, the one dimensional marginal of the process, namely the distribution of Y_n, appears to be a mixture of populations where the number of populations m is the cardinality of the state space of the hidden Markov chain. The observations may then be said to follow a "mixture model with Markov regime".
Asymptotic properties of maximum likelihood estimators (m.l.e.) for the estimation of the parameters in a HMM have received early interest when the (Y_k) take values in a finite set (Baum, Petrie, 1966), and recent interest in the general case (Leroux, 1992; Ryden, 1995; Bickel and Ritov, 1996; Bickel, Ritov and Ryden, 1998; Mevel, 1997; Vandekerkhove, 1997). Various algorithms have been proposed to achieve the m.l.e., see for instance Mevel (1997) or Vandekerkhove (1997).

The problem of estimating m, the number of populations, or of testing between different values of m, is an important one either for numerical features (a larger value than the true one leads to flat likelihoods) or for its interpretation as a meaningful parameter. Earlier works on the subject propose to use likelihood methods (Leroux, 1992; Ryden, 1995; Finesso, 1990), but fail to prove the consistency of the estimator (or the asymptotic level of the test) because of an intrinsic difficulty of the model: the usual regularity conditions are not fulfilled under the null hypothesis, so that the χ-square theory may not apply. Corresponding problem for mixtures with i.i.d. regime has a long history (see Lindsay, 1995) but the asymptotic distribution of the LRT (likelihood ratio test) has been recently solved by Dacunha-Castelle and Gassiat (1996, 1997b) and the consistency of the penalized likelihood estimator has been proved by Keribin (1998).

The aim of this paper is the study of the LRT for testing the number of populations in a mixture with Markov regime. Intuition and methods of Dacunha-Castelle and Gassiat (1996, 1997b) may be applied, and it appears that the computation of the asymptotic distribution in the general setting is much more complicated than in the i.i.d. regime situation. However, by applying the method to a subproblem of the general problem, we obtain meaningful results. Let us now be more precise.

Let

$$\mathcal{F} = \{f_\gamma, \gamma \in \Gamma\}$$

be a set of probability densities with respect to a given measure ν on a metric space, where Γ is a compact subset of some euclidian space. Let $(X_k)_{k\in\mathbb{N}}$ be a stationary Markov chain taking values in a set of m distinct values of Γ, $\{\gamma_1, \dots, \gamma_m\}$, with transition matrix Π. The observations are $(Y_1, \dots, Y_n)$. Conditionally to $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$ are independent variables, each Y_i having conditional distribution $f_{X_i} \cdot \nu$. Define

$$T_n(m) = \sup\{\ell_n(\Pi, \gamma_1, \dots, \gamma_m; m) \ : \ \Pi \ m \times m \text{ stochastic matrix}, (\gamma_1, \dots, \gamma_m) \in \Gamma^m\}, \tag{1}$$

in which $\ell_n(\Pi, \gamma_1, \dots, \gamma_m; m)$ is the log-likelihood of the observations under the parameter $(\Pi, \gamma_1, \dots, \gamma_m)$ when the number of hidden states is m. One of our main results is that the LRT statistic $T_n(2) - T_n(1)$ for testing i.i.d. observations of one population against 2 populations with Markov regime converges in probability to infinity under the null hypothesis (see Theorem 2.5). This result is a consequence of another convergence result in the sub-problem of testing a fixed contamination with Markov regime. Let f_{γ_0} and f_{γ_1} be fixed densities in $\mathcal{F}$, $\gamma_0 \neq \gamma_1$. Let $(X_k)_{k\in\mathbb{N}}$ take values in $\{\gamma_0, \ \gamma_1\}$, with initial distribution

$$P(X_0 = \gamma_0) = 1 - u, \ \ P(X_0 = \gamma_1) = u$$

and transition matrix

$$\Pi_{u,q} = \begin{pmatrix} 1 - \frac{qu}{1-u} & \frac{qu}{1-u} \\ q & 1 - q \end{pmatrix} \tag{2}$$

Here, $0 \le q \le 1$, $0 \le u \le 1$, and q and u are also constrained by

$$0 \le \frac{qu}{1-u} \le 1.$$

We want to test that the Y_i's are i.i.d. with common density f_{γ_0}, against that $(Y_n)_{n\in\mathbb{N}}$ follows a HMM for some parameters u, q, and with fixed γ_0 and γ_1. Obviously, the HMM reduces to i.i.d. variables with common density f_{γ_0} if and only if $u = 0$, so that the test we consider is that of H_0 : "$u = 0$" against H_1 : "$u \neq 0$". Let H be the Hilbert space $L^2(f_{\gamma_0} \cdot \nu)$, and

$$N^2 = \|\frac{f_{\gamma_1} - f_{\gamma_0}}{f_{\gamma_0}}\|_H^2.$$

For any $\eta \in]0, 1[$, let

$$\mathcal{Q}_\eta = \{q \in [0, 1] \ : \ (1 - q)^2(N^2 + 1) \le 1 - \eta\} = [q_\eta, 1] \tag{3}$$

with

$$q_\eta = 1 - \sqrt{\frac{1-\eta}{N^2+1}} > 0.$$

When the q's are restricted to belong to $\mathcal{Q}_\eta$, let $\mathcal{P}_\eta$ be the parameter set of possible (u, q). The log-likelihood $\ell_n(\Pi_{u,q}, \gamma_0, \gamma_1; 2)$ will now be denoted by $\ell_n(u, q)$. Notice that under the null hypothesis, the parameter q is undefined, the log-likelihood does not depend on q, and is $\ell_n(0, q)$ for any q. The LRT statistic is now, when the q's are restricted to belong to $\mathcal{Q}_\eta$:

$$V_n(\eta) = \sup\{\ell_n(u, q) - \ell_n(0, q) \ : \ (u, q) \in \mathcal{P}_\eta\}. \tag{4}$$

The first main convergence result is that $V_n(\eta)$ converges to half the supremum of the square of the positive part of some gaussian process (see Theorem 2.1). This result is similar to that obtained for testing the number of populations in a mixture, or for testing the order of a stationary ARMA process, see Dacunha-Castelle and Gassiat (1996, 1997b). As η goes to 0, the covariance of infinitely many members of the process goes to 0, and the convergence to infinity of $T_n(2) - T_n(1)$ is thus an easy consequence.
Convergence to infinity of LRT's has been observed earlier in two papers about mixtures with i.i.d. regime: when the parameter set is unbounded, see Hartigan (1985), and when the family $\mathcal{F}$ is not regular, two situations in which the set of scores may be proved to be non compact, see Ciuperca (1999). Here, the parameter set is compact, and the family $\mathcal{F}$ is regular. Our result may be compared to a similar result, not proved in any paper but which is also an easy consequence of Theorem 4.5 in Dacunha-Castelle and Gassiat (1996), that the LRT for testing the order of an ARMA model tends to infinity when the poles and roots are not restricted to a compact subset of the open unit complex disk.
Our result leaves open the problem of the choice of the penalization rate when estimating the number of populations using penalized likelihood estimators, for mixtures with Markov regime. However, for translation or scale mixtures, the method proposed by Dacunha-Castelle and Gassiat (1997a) may be applied to mixtures with Markov regime to obtain a consistent estimate of the number of populations.

The paper is organized as follows. In section 2, we state the assumptions and our main results. Section 3 is devoted to numerical experiments, and section 4 to the proofs of intermediate results.

2 Main results

We study the asymptotic properties of the LRT under the null hypothesis, in which the Y_i's are assumed to be i.i.d. with distribution $f_{\gamma_0}\nu$. Let us introduce the assumptions:

- (A1) For $\gamma \in \{\gamma_0, \gamma_1\}$, under H_0, $E(\mid \log f_\gamma(Y_1) \mid) < +\infty$.
- (A2) Under H_0, the support of the random variable $\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y_1)$ is $\mathbb{R}^+$.
- (A3) Under H_0, $E(\frac{f_{\gamma_0}}{f_{\gamma_1}}(Y_1)) < +\infty$.
- (A4) Under H_0, $E\left(\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y_1)\right)^3 < +\infty$.

Let now $(\xi_q)_{q\in\mathcal{Q}_\eta}$ be the continuous gaussian process with covariance

$$R(q_1, q_2) = \frac{C(q_1, q_2)}{C(q_1, q_1)^{1/2} C(q_2, q_2)^{1/2}} \tag{5}$$

with

$$C(q_1, q_2) = \frac{1 - (1 - q_1)(1 - q_2)}{1 - (1 - q_1)(1 - q_2)(N^2 + 1)}, \tag{6}$$

Then

Theorem 2.1 *Assume that (A1), (A2), (A3) and (A4) hold. Under H_0, $V_n(\eta)$ converges in distribution to*

$$\frac{1}{2} \sup_{q\in\mathcal{Q}_\eta} \xi_q^2 \cdot 1_{\xi_q \geq 0}.$$

Proof of Theorem 2.1.

Theorem 2.1 will be proved by investigating the asymptotic behavior of the maximum of the likelihood for fixed q as a process with $\mathcal{Q}_\eta$ the time set, and then taking the maximum for q over $\mathcal{Q}_\eta$. For an HMM with parameters $\Pi_{u,q}$, γ_0 and γ_1, let $(p_k(u,q))_{k\geq 1}$ be the prediction filter, that is, for any positive k:

$$p_k(u, q) = P(X_k = \gamma_1 / Y_1, \ldots, Y_{k-1}). \tag{7}$$

It is well known that the $p_k(u,q)$'s satisfy the recursive equation:

$$p_{k+1}(u, q) = \frac{\frac{qu}{(1-u)}(1 - p_k) f_{\gamma_0}(Y_k) + (1 - q) p_k f_{\gamma_1}(Y_k)}{(1 - p_k) f_{\gamma_0}(Y_k) + p_k f_{\gamma_1}(Y_k)} \tag{8}$$

with initialization $p_1(u,q) = u$. Parameter (u,q) is omitted for p_k on the right hand of the recurrence equation for readibility. This will be the case also in further recurrence equations concerning the derivatives of the prediction filter. Now the log-likelihood may be written:

$$\ell_n(u,q) - \ell_n(0,q) = \sum_{k=1}^{n} \log[1 + p_k(u,q)\frac{f_{\gamma_1} - f_{\gamma_0}}{f_{\gamma_0}}(Y_k)] \tag{9}$$

Define

$$S_n(u,q) = \frac{1}{n}[\ell_n(u,q) - \ell_n(0,q)].$$

Asymptotic properties of $S_n(u,q)$ and its derivatives will be investigated using the properties of the Markov chains $(Y_k, p_k(u,q))_{k\in\mathbb{N}}$, $(Y_k,$ derivatives of $p_k(u,q))_{k\in\mathbb{N}}$. This point of view is the one adopted by Mevel (1997), Mevel and LeGland (1999) to obtain the asymptotic properties of the maximum likelihood estimator (m.l.e.) of the parameters when the number of populations m is known.
Let now $(\widehat{u},\widehat{q})$ be the m.l.e. for (u,q), and $\widehat{u}_q$ the m.l.e. for u when q is fixed. We have:

Proposition 2.2 *Assume (A1), (A2) (A3) and (A4). Then* $\sup_{q\in\mathcal{Q}_\eta}\widehat{u}_q$ *and* $\widehat{u}$ *converge to* 0 *in probability under* H_0.

Remark: This result could be derived from Theorem 3 of Leroux (1992). However, we give here an alternative specific proof which shows the effect of overestimating m. The limiting contrast function of $S_n(u,q)$ will be proved to be

$$S(u,q) = -\int K(f_{\gamma_0}, (1-p)f_{\gamma_0} + pf_{\gamma_1})d\mu_{u,q}(p), \tag{10}$$

in which $K(\cdot,\cdot)$ is the Kullbak information (see for instance Dacunha-castelle and Duflo, 1993, for a definition) and $\mu_{u,q}$ is a probability distribution on $[0,1]$ such that, for $u=0$, $\mu_{0,q}$ is the dirac at 0. Notice that $S(u,q)$ is always finite by (A1), and, since the Kullback information is always non negative, for any (u,q), $S(u,q)\leq 0$, and $S(0,q)=0$. Proposition 2.2 will follow from the fact that $S(u,q)=0$ only for $u=0$, and that the convergence of $S_n(u,q)$ to $S(u,q)$ is uniform enough, see section 4. We have:

$$V_n(\eta) = \sup_{q\in\mathcal{Q}_\eta} nS_n(\hat{u}_q, q).$$

But using Proposition 2.2, for any q in $\mathcal{Q}_\eta$, Taylor expansion leads to

$$S_n(\widehat{u}_q, q) = -\frac{1}{2}\frac{(\frac{\partial}{\partial u}S_n(0,q))^2}{\frac{\partial^2}{\partial u^2}S_n(0,q)}\cdot(1 + R_n(q)) \tag{11}$$

in which $R_n(q)$ converges in probability to 0 uniformly in q as soon as

$$\frac{\frac{\partial^2}{\partial u^2}S_n(\tilde{u},q)}{\frac{\partial^2}{\partial u^2}S_n(0,q)}$$

converges to 1 uniformly in q for any $\tilde{u}$ converging in probability to 0 uniformly in q. Now,

$$\frac{\partial^2}{\partial u^2}S_n(u,q) = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(1+p_k(u,q)\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k))^2}(\frac{\partial p_k}{\partial u}(u,q))^2(\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k))^2 + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+p_k(u,q)\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k)}\frac{\partial^2 p_k}{\partial u^2}(u,q)\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k). \tag{12}$$

In particular, for $u=0$, one obtains

$$\frac{\partial^2}{\partial u^2}S_n(0,q) = -\frac{1}{n}\sum_{i=1}^{n}(\frac{\partial p_k}{\partial u}(0,q))^2(\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k))^2 + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 p_k}{\partial u^2}(0,q)\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k). \tag{13}$$

In this sum, the second term is centered, so that as soon as laws of large numbers apply, the limiting value is $-N^2$ times the expectation of $(\frac{\partial p_k}{\partial u}(0,q))^2$ under the stationary distribution. This convergence result holds more generally and we have:

Proposition 2.3 *Assume that (A2), (A3) and (A4) hold. Then, for any random $\tilde{u}$ converging in probability to 0 uniformly in q,*

$$\frac{\partial^2}{\partial u^2}S_n(\tilde{u},q)$$

converges in probability to $-N^2C(q,q)$ uniformly over $\mathcal{Q}_\eta$.

Proposition 2.3 leads to

$$V_n(\eta) = \frac{1}{2}\left[\sup_{q\in\mathcal{Q}_\eta}\frac{(\sqrt{n}\frac{\partial}{\partial u}S_n(0,q))^2}{N^2C(q,q)}\right](1+o_P(1)) \tag{14}$$

and Theorem 2.1 will follow from

Proposition 2.4 *Assume (A1), (A2) and (A3). Then, the process*

$$\sqrt{n}\left(\frac{\partial}{\partial u}S_n(0,q)\right)_{q\in\mathcal{Q}_\eta}$$

converges uniformly to the continuous gaussian process with covariance $N^2C(\cdot,\cdot)$.

Comments.

The score process at $u=0$ and for fixed q is:

$$\frac{\partial}{\partial u}S_n(0,q) = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial p_k}{\partial u}(0,q)\frac{f_{\gamma_1}-f_{\gamma_0}}{f_{\gamma_0}}(Y_k)$$

in which $(\frac{\partial p_k}{\partial u}(0,q))_{k\in\mathbb{N}}$ is a Markov chain satisfying the recursion:

$$\frac{\partial p_{k+1}}{\partial u}(0,q) = q + (1-q)\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y_k)\frac{\partial p_k}{\partial u}(0,q) \tag{15}$$

with initialization $\frac{\partial p_1}{\partial u}(0,q) = 1$. But in case an invariant distribution exists and has a second moment σ^2, (15) leads to:

$$\sigma^2 = q^2 + 2q(1-q) + (1-q)^2(N^2+1)\sigma^2.$$

Since for any q, $q^2 + 2q(1-q) \geq 0$, this equation leads to a non negative σ^2 if and only if $1-(1-q)^2(N^2+1) \geq 0$. This last condition is thus a necessary one to obtain a convergence in distribution of $\sqrt{n}\frac{\partial}{\partial u}S_n(0,q)$ and a law of large numbers for $\frac{\partial^2}{\partial u^2}S_n(0,q)$. This is why the parameter set is cut in two different regions, the union of all $\mathcal{P}_\eta$, and its complementary set. Our conjecture is that, on this complementary set, $nS_n(\hat{u}_q,q)$ converges to half the square of a standard gaussian white noise.

To state the unboundedness of the LRT statistic for testing between one or two populations in a mixture with Markov regime, we introduce the following assumption:

- **(I)** $T_n(1) - \ell_n(0,q)$ is uniformly tight under H_0.

Notice that a sufficient condition for (I) to hold is that the family $\mathcal{F}$ is regular enough. For instance, differentiability in quadratic mean with respect to γ and existence and inversibility of the Fisher information suffice since in this case $T_n(1)-\ell_n(0,q)$ converges to a $\chi^2(1)$ (see for instance Van der Vaart, 1999). We now obtain the unconstrained result:

Theorem 2.5 *Assume (I), and that (A1), (A2), (A3) and (A4) hold for any* γ_0 *and* γ_1 *in* Γ. *Then, under* H_0, $T_n(2) - T_n(1)$ *converges to infinity in probability.*

Remark.
The same result holds if the parameter space is restricted to all possible (u,q) such that

$$(1-q)^2(N^2+1) \leq 1,$$

as may be seen from the proof. This means that the unboundedness of the LRT statistic is not due to the fact that the model contains non ergodic hidden Markov chains (by taking $q=0$), but to the fact that the alternative may be too far from the null hypothesis, also with respect to the distance between the populations involved in the mixture.

Proof of Theorem 2.5.

Let $f_{\gamma_0}\nu$ be the true (unknown) distribution of the i.i.d. variables (Y_k). Let γ_1 be a point in Γ distinct of γ_0.

Since the LRT statistic for i.i.d. observations with the model $\mathcal{F}$ is tight, $T_n(2) - T_n(1)$ converges to infinity in probability as soon as $V_n(0)$ does.
Let $\rho < 1$, and define for all integer k:

$$r_k = \rho^{k^2}.$$

Define also $q_k \in [0,1]$ by

$$(1-q_k)^2(N^2+1) = 1 - r_k,$$

Obviously,

$$V_n(0) \geq \sup\{\ell_n(\widehat{u_{q_k}}, q_k) - \ell_n(0, q_k) \ : \ k \in \mathbb{N}\}. \tag{16}$$

Now, observe that

$$\lim_{k \neq l \to +\infty} R(q_k, q_l) = 0,$$

so that

$$\sup_{k \leq L} \xi_{q_k}^2 \cdot 1_{\xi_{q_k} \geq 0}$$

converges in probability to $+\infty$ as L tends to $+\infty$. For any positive M, for any positive ϵ, there exists L_0 such that

$$P(\sup_{k \leq L_0} \xi_{q_k}^2 \cdot 1_{\xi_{q_k} \geq 0} > M) \geq 1 - \epsilon.$$

But using portmanteau's Lemma, the fact that $\sup\{\ell_n(\widehat{u_{q_k}}, q_k) - \ell_n(0, q_k) \ : \ k \leq L_0\}$ converges in distribution to $\sup_{k \leq L_0} \xi_{q_k}^2 \cdot 1_{\xi_{q_k} \geq 0}$, and inequality (16), we have that

$$\liminf_{n \to +\infty} P(V_n(0) > M) \geq 1 - \epsilon.$$

Since this holds for any ϵ, $V_n(0)$ tends to $+\infty$ in probability.

3 Numerical experiments

In this section, we numerically investigate the behavior of the LRT statistic. The true model is an i.i.d. process with gaussian distribution $\mathcal{N}(0,1)$ of mean 0 and variance 1. We test i.i.d. observations of one population with distribution $\mathcal{N}(0,1)$ against two populations with Markov regime, the distributions of the observations conditionally to the Markov chain being $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$. In other words $f_{\gamma_0}(x) = (2\pi)^{-1/2}\exp -\frac{x^2}{2}$ and $f_{\gamma_1}(x) = (2\pi)^{-1/2}\exp -\frac{(x-1)^2}{2}$.

We first draw the likelihood map, then we study the convergence in distribution of $V_n(\eta)$ (Theorem (2.1)) and we finally experiment the convergence to infinity of the LRT statistic (Theorem (2.5)).

3.1 Likelihood map

We compute the log-likelihood ratio (LLR)

$$\ell_n(u,q) - \ell_n(0,q) = \sum_{k=1}^{n} \log[1 + p_k(u,q)\frac{f_{\gamma_1} - f_{\gamma_0}}{f_{\gamma_0}}(Y_k)],$$

using the recursive equation of the prediction filter (8) for $(q,u) \in [0,1] \times [0,1]$, such that $qu/(1-u) \leq 1$, by step of 0.01. The typical aspect of all our LLR maps is the following: we observe the LLR to fast decrease with u, and refine the map for u near 0. Two cases happen:

- Case A: the maximum is 0, reached for $\hat{u} = 0$, and the LLR is an always decreasing function of u (Fig. 1),
- Case B: the maximum is stricly positive and is reached for $\hat{u}$ close to 0 (Fig. 2).

When the number of observations increases, $\hat{u}$ gets closer to 0. For 10000 observations, $\hat{u}$ is of the order of 0.01, for 50000 observations, $\hat{u}$ is of the order of 0.005, and for 100000 observations, $\hat{u}$ is of the order of 0.001.

As the LLR can take large negative values, we cut off the values larger than -3 and replaced them by -3 to allow the visualization of the maximum. In the same way, we represent the points of the domain where $qu/(1-u) \geq 1$ with the value -4.

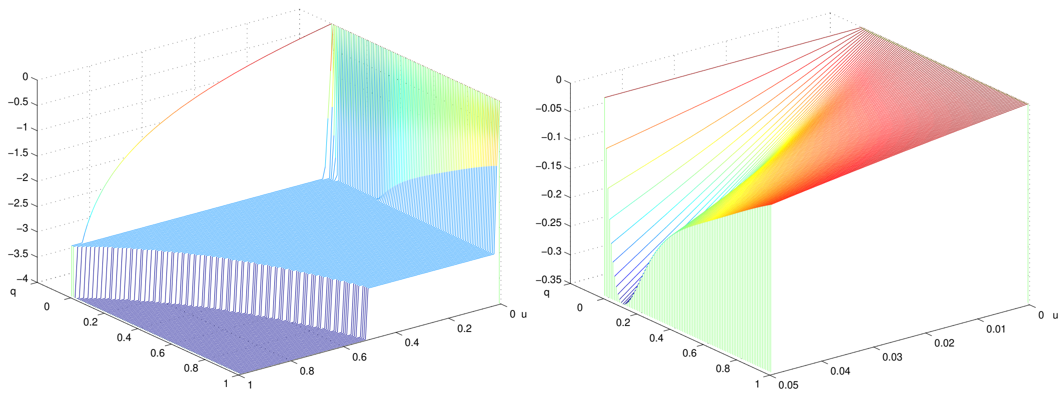

Figure 1: *Case A:On left: on the whole domain* $(q,u) \in [0,1] \times [0,1]$ *by steps of* 0.01. *On right: refinement* $(q,u) \in [0,1] \times [0,0.05]$ *by steps of* 5.10^{-4} *for* u *and* 0.01 *for* q

3.2 Distribution of $V_n(\eta)$

The aim of this section is to study the rate in the distributional convergence of the LRT statistic on $\mathcal{P}_\eta$ to the asymptotic distribution given in Theorem (2.1):

$$\frac{1}{2} \sup_{q \in \mathcal{Q}_\eta} \xi_q^2 \cdot 1_{\xi_q \geq 0},$$

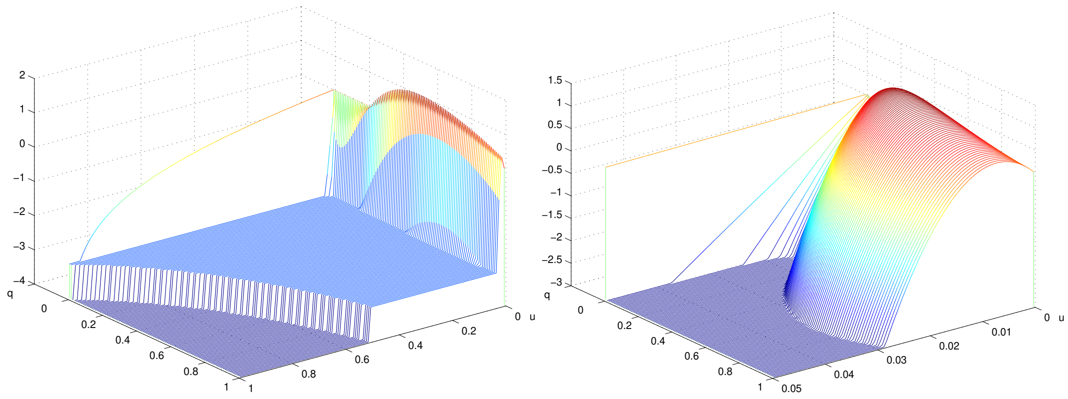

Figure 2: *Case B. On left: on the whole domain* $(q,u) \in [0,1] \times [0,1]$ *by steps of* 0.01*. On right: refinement* $(q,u) \in [0,1] \times [0,0.05]$ *by steps of* 5.10^{-4} *for* u *and* 0.01 *for* q

to be able to say for how many observations the asymptotic approximation holds. We encounter difficulties to have a significant result on the whole $\mathcal{P}_\eta$ domain, so we do the simulations on a simplified model, where the supremum is taken over two values of q.

$$\frac{1}{2} \sup_{q \in \{1, 0.4\}} \xi_q^2 \cdot 1_{\xi_q \geq 0}.$$

In this case, we can easily simulate the theorical law, using the following property: if X_1 and X_2 are independent normal deviates $\mathcal{N}(0,1)$ and if

$$Y_1 = \sigma_1 X_1, \quad Y_2 = \sigma_2(\rho X_1 + \sqrt{1-\rho^2} X_2),$$

then Y_1 and Y_2 are dependent random variables, normally distributed with standard deviations σ_1 and σ_2, and with correlation coefficient ρ.

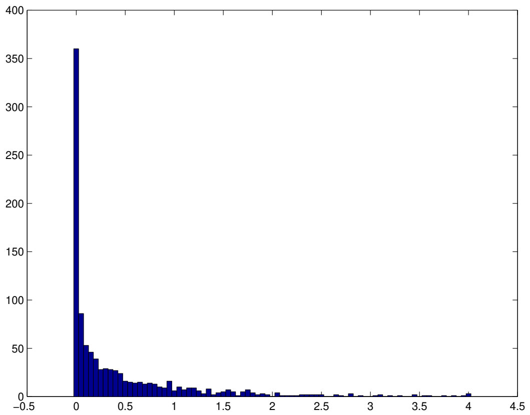

Figure 3: *Histogram of the simulated law* $\frac{1}{2} \sup_{q \in \{1,0.4\}} \xi_q^2 \cdot 1_{\xi_q \geq 0}$*. with* 1000 *points*

Fig. 3 shows the histogram of the simulated law with 1000 points, with $\rho = 0.18$ corresponding to the correlation when $q = 1$ and $q = 4$. The outline does not significantly change when we increase the number of points. Notice that the weight for $[0, 0.05]$ is around 370. For two independent normal deviates, this weight is around 340, whereas for $\rho = 1$, this weight is around 620.

To approximate the distribution of the LRT statistic $V_n(0.4, 1)$ in this case, we draw m samples of length n, and compute for each sample the maximum likelihood ratio value for the two values of q. For this computation, we use a Newton algorithm, using the fact, checked on a great number of samples, that the log-likelihood has an unique maximum near $u = 0$. Other simulations with an EM algorithm (Dempster *and al.* (1977)) and various initial points did not give better results. We first take $m = 500$ and $n = 2,000$, $n = 5,000$, $n = 10,000$ (Fig. 4 left). The weight of $[0, 0.05]$ is around 250. Compared with the simulated law $370/2 = 185$, it is much more higher. If we now increase the number of samples to $m = 5,000$ of size $n = 10,000$ (Fig. 4 right), this weight $2450/10 = 245$ is not significantly modified.

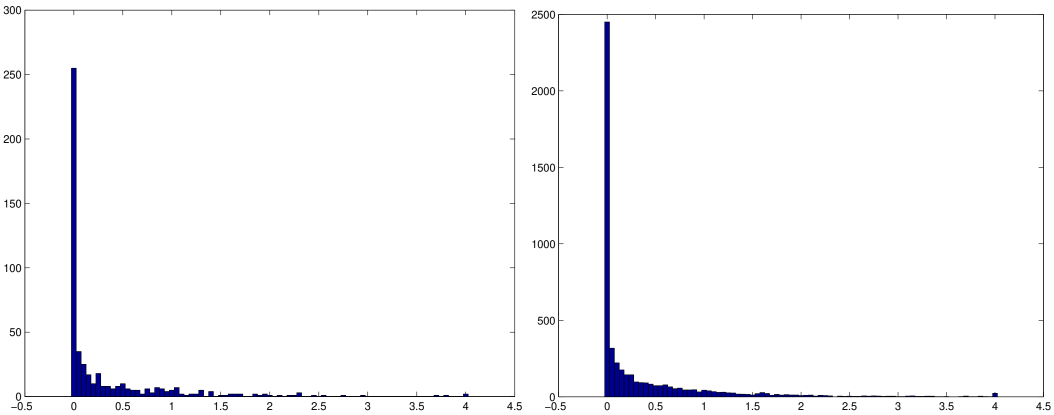

Figure 4: *Histogram of the simulated LRT statistic* $V_n(0.4, 1)$, $m = 500$ *(on left) and* $m = 5000$ *(on right) samples of size* $n = 10,000$

If we increase the size of the samples to $n = 100,000$ for 500 samples (Fig. 5 left), this weight begins to decrease (around 230), and this is more significant for $m = 50$ samples of size $1,000,000$ (Fig. 5 right), where the weight is of the order of $20 \times 10 = 200$.

This allows us to conclude that the convergence is extremely slow, and we can relate this fact to the slow convergence of the LRT statistic for a mixture of two normal distributions as pointed out by Atwood *and al.* (1996) or Chuang and Mendell (1997). Bootstrap techniques might also be another way to investigate the distribution, see McLachlan (1987) or Feng and McCulloch (1996) for the case of i.i.d mixtures.

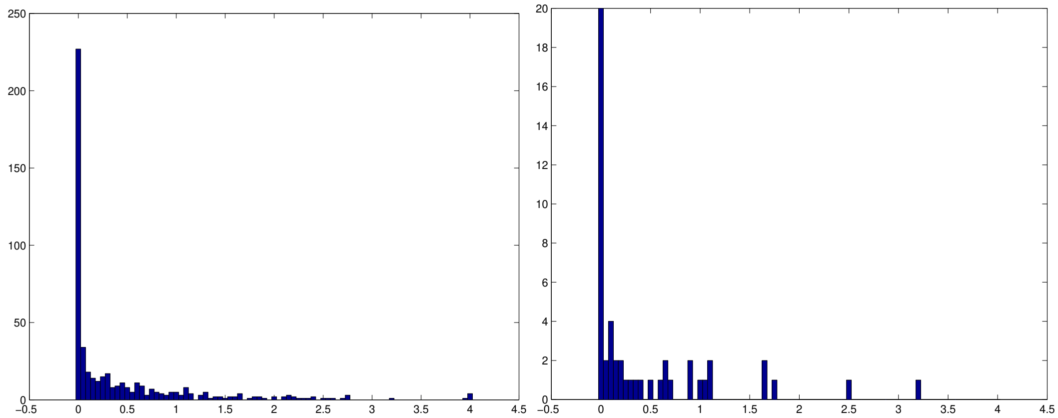

Figure 5: *Histogram of the simulated LRT statistic* $V_n(0.4, 1)$, $m = 500$ *samples of size* $n = 100,000$ *(on left) and* $m = 50$ *samples of size* $n = 1,000,000$ *(on right)*

3.3 Convergence of the LRT statistic to infinity

At which rate does the LRT on $\mathcal{P}_0$ converge to infinity? We numerically investigate the evolution of the LRT with the length n of the sample. As pointed out in the proof of Theorem (2.5), it is sufficient to take a series q_k such that

$$(1 - q_k)^2(N^2 + 1) = 1 - \rho^{k^2}.$$

We choose $\rho = 0.85$, $k = 0, \ldots, 5$, and compute the LRT from $n = 50$ to $n = 100,000$ by steps of 50 (Fig. 6).

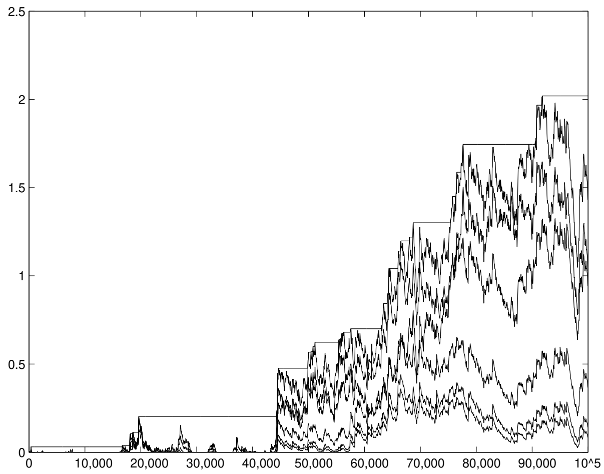

Figure 6: *Histogram of the simulated LRT statistic,* $m = 500$ *samples of size* $n = 100,000$

The result shows that the convergence is very slow, but can be seen. For the i.i.d. mixture of two normals, Hartigan (1985) gave a $\frac{1}{2} \log\log(n)$ rate. Such rate cannot be experimentally derived, it has to be theoretically investigated for mixtures with Markov regime in a further work.

4 Proofs.

4.1 Proof of Proposition 2.2.

Proposition 2.2 is a consequence of the following Lemmas.

Lemma 4.1 *Assume (A2). Let (u,q) be a fixed parameter. Then under H_0, $(Y_k, p_k(u,q))_{k\in\mathbb{N}}$ is a uniformly ergodic chain with invariant probability $\pi = f_{\gamma_0}\nu \otimes \mu_{u,q}$. For $u=0$, $\mu_{0,q}$ is the dirac at 0, and for $u \neq 0$,*

$$\int p d\mu_{u,q} \geq \frac{qu}{2(1-u)}. \tag{17}$$

Proof of Lemma 4.1. Under H_0, $(Y_k, p_k(u,q))_{k\in\mathbb{N}}$ is a Markov chain since the (Y_k) are i.i.d. and by equation (8). Moreover, for any k, Y_k and $p_k(u,q)$ are independent variables so that the Lemma follows as soon as $(p_k(u,q))_{k\in\mathbb{N}}$ is a uniformly ergodic chain with invariant probability $\mu_{u,q}$. For $u=0$, $p_k(0,q)=0$ for all k. For $u\neq 0$, $(p_k(u,q))_{k\in\mathbb{N}}$ is a Feller chain, ψ-irreductible with ψ the Lebesgue measure on $(\frac{qu}{1-u}, 1-q)$ thanks to (A2), thus a T-chain by Theorem 6.2.9, and then a uniformly ergodic chain by Theorem 16.2.5 of Meyn and Tweedie (1993). Now using equation (8) one obtains, using the fact that $p_k(u,q)$ is $\sigma(Y_1,\ldots,Y_{k-1})$-measurable for all k, independent of Y_k, and using Jensen's inequality applied to the function $1/x$:

$$E(p_{k+1}(u,q)/Y_1,\ldots,Y_{k-1}) \geq \frac{qu}{(1-u)}(1-p_k(u,q))$$

so that

$$E(p_{k+1}(u,q)) \geq \frac{qu}{(1-u)}(1-E(p_k(u,q)))$$

and equation (17) follows by taking the limit as k goes to infinity.

Lemma 4.2 *Assume (A1) and (A2). Then under H_0, $S_n(u,q)$ converges a.s. to $S(u,q)$. Moreover, for any positive δ,*

$$\sup_{u\geq\delta,(u,q)\in\mathcal{P}_\eta} S(u,q) = -2\epsilon(\delta,\eta) < 0$$

Proof of Lemma 4.2. The a.s. convergence of $S_n(u,q)$ to $S(u,q)$ follows from the first point of Lemma 4.1. Now,

$$S(u,q) = \int f_{\gamma_0}(x)d\nu(x)[\int \log(1-p+p\frac{f_{\gamma_1}}{f_{\gamma_0}}(x))d\mu_{u,q}(p)]$$

and by Jensen's inequality

$$\int \log(1-p+p\frac{f_{\gamma_1}}{f_{\gamma_0}}(x))d\mu_{u,q}(p) \leq \log(1-\int pd\mu_{u,q}(p)+\int pd\mu_{u,q}(p)\frac{f_{\gamma_1}}{f_{\gamma_0}}(x))$$

so that

$$S(u,q) \leq -K(f_{\gamma_0},(1-\int pd\mu_{u,q}(p))f_{\gamma_0}+\int pd\mu_{u,q}(p)f_{\gamma_1}).$$

Using now the second point of Lemma 4.1 one obtains

$$\sup_{u\geq\delta,(u,q)\in\mathcal{P}_\eta} S(u,q) \leq -\inf\{K(f_{\gamma_0},(1-p)f_{\gamma_0}+pf_{\gamma_1}) \ : \ u\geq\delta,(u,q)\in\mathcal{P}_\eta, p\geq\frac{qu}{2(1-u)}\}.$$

Now using (A1) $K(f_{\gamma_0},(1-p)f_{\gamma_0}+pf_{\gamma_1})$ is a continuous function of p and the result follows.

Lemma 4.3 *There exist a positive number M_1 such that, for any integer k,*

$$E\left(\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k}{\partial u}(u,q)|\right) < M_1, \quad E\left(\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k}{\partial q}(u,q)|\right) < M_1$$

Proof of Lemma 4.3.

Let us give the iterative model for the first derivative with respect to u and with respect to q of $p_k(u,q)$ for any value of (u,q). From now on, we shall denote by W_k the random variable $\frac{f_{\gamma_1}(Y_k)}{f_{\gamma_0}(Y_k)}$. W_k is non negative and its expectation under H_0 equals 1. The initial value is

$$\frac{\partial p_1}{\partial u}(u,q)=1, \frac{\partial p_1}{\partial q}(u,q)=0,$$

and for any positive integer k:

$$\begin{aligned}\frac{\partial p_{k+1}}{\partial u}(u,q) &= \frac{q}{(1-u)^2}\frac{(1-p_k)}{(1-p_k)+p_kW_k} \\ &+\left(1-\frac{q}{(1-u)}\right)\frac{W_k}{[(1-p_k)+p_kW_k]^2}\frac{\partial p_k}{\partial u} \end{aligned} \tag{18}$$

$$\begin{aligned}\frac{\partial p_{k+1}}{\partial q}(u,q) &= \frac{\frac{u}{(1-u)}(1-p_k)-p_kW_k}{(1-p_k)+p_kW_k} \\ &+\left(1-\frac{q}{(1-u)}\right)\frac{W_k}{[(1-p_k)+p_kW_k]^2}\frac{\partial p_k}{\partial q}. \end{aligned} \tag{19}$$

Direct computation leads to

$$|\frac{\partial p_{k+1}/\partial u}{(1-p_{k+1})^2}| \leq \frac{q}{(1-u)^2[(1-\frac{qu}{1-u})\bigwedge q]^2}+\frac{1-\frac{q}{(1-u)}}{(1-\frac{qu}{(1-u)})^2}W_k|\frac{\partial p_k/\partial u}{(1-p_k)^2}|, \tag{20}$$

in which $a\bigwedge b$ means $\inf\{a,b\}$. We have

$$\sup\{\frac{q}{(1-u)^2[(1-\frac{qu}{1-u})\bigwedge q]^2} \ : \ (u,q)\in\mathcal{P}_\eta\} = M_2 < +\infty$$

and also

$$\sup\{\frac{1-\frac{q}{(1-u)}}{(1-\frac{qu}{(1-u)})^2}\ :\ (u,q)\in\mathcal{P}_\eta\} = 1-h_1<1$$

so that

$$\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_{k+1}/\partial u}{(1-p_{k+1})^2}| \ \leq\ M_2+(1-h_1)W_k\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k/\partial u}{(1-p_k)^2}|. \qquad (21)$$

This leads to

$$E\left[\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_{k+1}/\partial u}{(1-p_{k+1})^2}|\right]\leq M_2+(1-h_1)E\left[\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k/\partial u}{(1-p_k)^2}|\right].$$

By induction one obtains that for any integer k

$$E\left[\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k/\partial u}{(1-p_k)^2}|\right]\leq\frac{M_2}{h_1}.$$

Since obviously

$$\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k}{\partial u}|\leq\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k/\partial u}{(1-p_k)^2}|$$

one obtains the first inequality of Lemma 4.3 with $M_1=\frac{M_2}{h_1}$. The second inequality may be obtained similarly.

Let us come back to the proof of Proposition 2.2. For any positive δ, we have

$$P\left(\sup_{q\in\mathcal{Q}_\eta}\widehat{u}_q\geq\delta\right)\geq P\left(\widehat{u}\geq\delta\right)$$

so that it is enough to prove the Proposition for $\sup_{q\in\mathcal{Q}_\eta}\widehat{u}_q$. Now, let (u_i,q_i) be the centers of $N(h)$ squares of edge of lentgh h covering $\mathcal{P}_\eta^\delta=\{(u,q)\in\mathcal{P}_\eta : u\geq\delta\}$. We have

$$\begin{aligned}P\left(\sup_{q\in\mathcal{Q}_\eta}\widehat{u}_q\geq\delta\right) &\leq P\left(\sup_{(u,q)\in\mathcal{P}_\eta^\delta}S_n(u,q)\geq 0\right)\\ &\leq \sum_{i=1}^{N(h)}P\left(S_n(u_i,q_i)\geq-\epsilon(\delta,\eta)\right)\\ &+ P\left(\sup_{|u-u'|\leq h,|q-q'|\leq h,(u,q)\in\mathcal{P}_\eta^\delta}|S_n(u,q)-S_n(u',q')|\geq\epsilon(\delta,\eta)\right).\end{aligned}$$

Now, as soon as $|u-u'|\leq h, |q-q'|\leq h, (u,q)\in\mathcal{P}_\eta^\delta$,

$$|S_n(u,q)-S_n(u',q')|\leq\frac{h}{n}\sum_{k=1}^{n}(\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k}{\partial u}(u,q)|+\sup_{(u,q)\in\mathcal{P}_\eta}|\frac{\partial p_k}{\partial q}(u,q)|)m(Y_k),$$

where m is the function

$$m(x) = \max\left(\frac{f_{\gamma_0}}{f_{\gamma_1}}(x); \frac{f_{\gamma_1}}{f_{\gamma_0}}(x)\right) - 1$$

Use Lemma 4.1, Lemma 4.2, Markov inequality, the fact that Y_k is independent of the p_k and its derivatives, and Lemma 4.3, to conclude that

$$\limsup_{n\to+\infty} P\left(\sup_{q\in\mathcal{Q}_\eta} \widehat{u}_q \geq \delta\right) = 0$$

and Proposition 2.2 follows.

4.2 Proof of Proposition 2.3.

We shall first prove:

Lemma 4.4 *Assume (A2), (A3) and (A4). The random variable*

$$\frac{\partial^2}{\partial u^2} S_n(0,q)$$

converges in probability to $N^2C(q,q)$ uniformly over $\mathcal{Q}_\eta$.

Proof of Lemma 4.4.
We have

$$\frac{\partial^2}{\partial u^2} S_n(0,q) = -\frac{1}{n}\sum_{i=1}^{n}(\frac{\partial p_k}{\partial u}(0,q))^2(W_k-1)^2 + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 p_k}{\partial u^2}(0,q)(W_k-1).$$

We shall use the following trick **(T1)** for proving weak laws of large numbers for Markov chains. Let $(\Phi_k)_{k\in\mathbb{N}}$ be a Markov chain in $\mathbb{R}^d$ for some dimension d. Assume that g is a norm-like function, that is a non negative function such that $\lim_{|x|\to+\infty} g(x) = +\infty$.

- 1. If there exists an initial value x such that

$$\liminf_{k\to+\infty} E[g(\Phi_k)/\Phi_0 = x] < +\infty$$

  then by Proposition 12-1-3 of Meyn and Tweedie (1993), the Markov chains admits an invariant probability μ.

- 2. If moreover for any initial points v_1 and v_2, $\Phi_k^{v_1} - \Phi_k^{v_2}$ tends in distribution to 0, then by using 1.IV.22 of Duflo (1996), the invariant probability is unique, and Φ_k converges in distribution to μ.

- 3. Then, by using Proposition 6-4-2 of Meyn and Tweedie (1993), $(\Phi_k)_{k\in\mathbb{N}}$ is an e-chain.

- 4. If moreover there exists a reachable state, then for any real function f such that $f \leq V$,
$$\frac{1}{n}\sum_{k=1}^{n} f(\Phi_k)$$
converges in probability to $\int f d\mu$ by a similar proof as for Theorem 18-5-1 of Meyn and Tweedie (1993).

We shall apply **(T1)** to $\Phi_k = (W_k, \frac{\partial p_k}{\partial u}(0,q), \frac{\partial^2 p_k}{\partial u^2}(0,q))$. Let us recall the recursive equations

$$\begin{aligned}
\frac{\partial p_{k+1}}{\partial u}(0,q) &= q + (1-q)W_k \frac{\partial p_k}{\partial u}(0,q) && (22)\\
\frac{\partial^2 p_{k+1}}{\partial u^2}(0,q) &= 2q - 2qW_k\frac{\partial p_k}{\partial u}(0,q) - 2(1-q)(W_k-1)W_k\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2 \\
&\quad +(1-q)W_k\frac{\partial^2 p_k}{\partial u^2}(0,q) && (23)
\end{aligned}$$

with initialization $\frac{\partial p_1}{\partial u}(0,q) = 1$ and $\frac{\partial^2 p_1}{\partial u^2}(0,q) = 0$. Notice that $\frac{\partial p_k}{\partial u}(u,q)$ is non negative for all integer k. (22) leads to

$$E\left(\frac{\partial p_{k+1}}{\partial u}(0,q)\right) = 1$$

for any integer k, and

$$E\left(\frac{\partial p_{k+1}}{\partial u}(0,q)\right)^2 = q^2 + 2q(1-q) + (1-q)^2(N^2+1)E\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2$$

so that by induction, for any integer k,

$$E\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2 \leq \frac{2}{\eta}.$$

Thus
$$E\left|\frac{\partial^2 p_{k+1}}{\partial u^2}(0,q)\right| \leq \frac{8}{\eta N^2} + (1-q)E\left|\frac{\partial^2 p_k}{\partial u^2}(0,q)\right|$$
and, by induction, for any integer k,

$$E\left|\frac{\partial^2 p_{k+1}}{\partial u^2}(0,q)\right| \leq \frac{8}{\eta N^2 q}.$$

Point 1. of **(T1)** holds with $g(x,y,z) = x^2 + y^2 + |z|$. To prove Point 2., notice first that the W_k are i.i.d. so that it is enough to prove it for the second and third coordinates of Φ_k. Let $v_1 = (y_1, z_1)$ and $v_2 = (y_2, z_2)$ be two different initial points for

$(\frac{\partial p_k}{\partial u}(0,q), \frac{\partial^2 p_k}{\partial u^2}(0,q))$. Let $\Delta_k(y_1,y_2)$ be the difference of $\frac{\partial p_k}{\partial u}(0,q)$ with initial values y_1 and y_2. Let now $S_k(y_1,y_2)$ be the sum of $\frac{\partial p_k}{\partial u}(0,q)$ with initial values y_1 and y_2, and $\Delta_k^2(v_1,v_2)$ be the difference of $\frac{\partial^2 p_k}{\partial u^2}(0,q)$ when the initial values are v_1 and v_2. Equations (22) and (23) give

$$\Delta_{k+1}(y_1,y_2) = (1-q)W_k\Delta_k(y_1,y_2)$$

$$S_{k+1}(y_1,y_2) = 2q + (1-q)W_kS_k(y_1,y_2)$$

$$\Delta_{k+1}(y_1,y_2)S_{k+1}(y_1,y_2) = 2q(1-q)W_k\Delta_k(y_1,y_2) + (1-q)^2W_k^2\Delta_k(y_1,y_2)S_k(y_1,y_2)$$

$$\Delta_{k+1}^2(v_1,v_2) = -2qW_k\Delta_k(y_1,y_2) - 2(1-q)(W_k-1)W_k\Delta_k(y_1,y_2)S_k(y_1,y_2)$$
$$+(1-q)W_k\Delta_k^2(v_1,v_2)$$

By iterating these equations, one obtains that, for any integer k,

$$E|\Delta_k(y_1,y_2)| = (1-q)^{k-1}|y_1-y_2|,$$

$$E|\Delta_k(y_1,y_2)S_k(y_1,y_2)| \le \frac{2q}{\eta_*}(1-q)^{k-1}|y_1-y_2|,$$

as soon as η_* is such that $\eta_* \le \eta$ and $|x_1^2 - y_1^2| \le 2q/\eta_*$, and then

$$E|\Delta_k^2(v_1,v_2)| \le 2(1-q)^{k-1}|y_1-y_2| + \frac{4}{\eta_*}(1-q)^kN^2|y_1-y_2|$$

as soon as also $|z_1 - z_2| \le 2|y_1-y_2| + \frac{4}{\eta_*}(1-q)N^2|y_1-y_2|$, which is always possible by choosing η_* small enough. Consequently, $\Delta_k(y_1,y_2)$ and $\Delta_k^2(v_1,v_2)$ tend in distribution to 0, and point 2 of **(T1)** holds. Since by using assumption (A3) the point $(0,q,2q)$ is obviously reachable, point 4. also holds, and for any q in $\mathcal{Q}_\eta$, the random variable

$$\frac{\partial^2}{\partial u^2}S_n(0,q)$$

converges in probability to $N^2C(q,q)$.

Let us now prove that this convergence is uniform. We have, using (22), for any q and q':

$$\begin{aligned}\frac{\partial p_{k+1}}{\partial u}(0,q) - \frac{\partial p_{k+1}}{\partial u}(0,q') &= (q-q')[1 - W_k\frac{\partial p_k}{\partial u}(0,q)] \\ &\quad +(1-q')W_k\left(\frac{\partial p_k}{\partial u}(0,q) - \frac{\partial p_k}{\partial u}(0,q')\right).\end{aligned}$$

Using Lemma 4.3 one obtains by induction

$$E\sup_{|q-q'|\le\delta, q,q'\in\mathcal{Q}_\eta}|\frac{\partial p_k}{\partial u}(0,q) - \frac{\partial p_k}{\partial u}(0,q')| \le \delta\cdot\frac{1+M_1}{q_\eta} \qquad (24)$$

Also,

$$\begin{aligned}\left(\frac{\partial p_{k+1}}{\partial u}(0,q)\right)^2 &= q^2+2q(1-q)W_k\frac{\partial p_k}{\partial u}(0,q)\\ &\quad +(1-q)^2W_k^2\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2,\end{aligned}$$

so that,

$$E\sup_{q\in\mathcal{Q}_\eta}\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2\le\frac{M_1+2}{\eta}. \tag{25}$$

Further, for any q and q',

$$\begin{aligned}\left(\frac{\partial p_{k+1}}{\partial u}(0,q)\right)^2-\left(\frac{\partial p_{k+1}}{\partial u}(0,q')\right)^2 &= q^2-q'^2+(2q(1-q)-2q'(1-q'))W_k\frac{\partial p_k}{\partial u}(0,q)\\ &\quad 2q'(1-q')W_k\left(\frac{\partial p_k}{\partial u}(0,q)-\frac{\partial p_k}{\partial u}(0,q')\right)\\ &\quad +((1-q)^2-(1-q')^2)W_k^2\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2\\ &\quad +(1-q')^2W_k^2\left[\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2-\left(\frac{\partial p_k}{\partial u}(0,q')\right)^2\right]\end{aligned}$$

and by induction,

$$E\sup_{|q-q'|\le\delta,q,q'\in\mathcal{Q}_\eta}|\left(\frac{\partial p_k}{\partial u}(0,q)\right)^2-\left(\frac{\partial p_k}{\partial u}(0,q')\right)^2|\le\delta\cdot M_4 \tag{26}$$

with

$$M_4=\frac{1}{\eta}\cdot[2+2M_1+\frac{1+M_1}{q_\eta}+\frac{2(N^2+1)M_1}{\eta}].$$

Following the same lines, using (23) one obtains

$$E\sup_{|q-q'|\le\delta,q,q'\in\mathcal{Q}_\eta}|\frac{\partial^2 p_k}{\partial u^2}(0,q)-\frac{\partial^2 p_k}{\partial u^2}(0,q')|\le\frac{\delta}{q_\eta}\cdot[2+2M_1+2\frac{1+M_1}{q_\eta}+2N^2(\frac{M_1}{\eta}+M_4)]. \tag{27}$$

Using (25) and (27), it follows that for some constant M_5 one has

$$E\sup_{|q-q'|\le\delta,q,q'\in\mathcal{Q}_\eta}|\frac{\partial^2}{\partial u^2}S_n(0,q)-\frac{\partial^2}{\partial u^2}S_n(0,q')|\le\delta\cdot M_5$$

and Lemma 4.4 follows.

To finish the proof of Proposition 2.3 we shall prove:

Lemma 4.5 *Assume that (A2), (A3) and (A4) hold. Then, for any random $\tilde{u}$ converging to* 0 *uniformly in q,*

$$\frac{\partial^2}{\partial u^2} S_n(\tilde{u}, q) - \frac{\partial^2}{\partial u^2} S_n(0, q)$$

converges in probability to 0 *uniformly over $\mathcal{Q}_\eta$.*

Proof of Lemma 4.5.

For any non negative u, we have

$$\begin{aligned}\frac{\partial^2}{\partial u^2} S_n(u, q) - \frac{\partial^2}{\partial u^2} S_n(0, q) \quad = \quad & \frac{1}{n} \sum_{k=1}^{n} (W_k - 1)^2 [\frac{\partial p_k}{\partial u}(0, q)^2 - \frac{\partial p_k}{\partial u}(u, q)^2 \\ & + \frac{\partial p_k}{\partial u}(u, q)^2 (1 - \frac{1}{(1 - p_k(u, q) + p_k(u, q) W_k)^2})] \\ & + \frac{1}{n} \sum_{k=1}^{n} (W_k - 1)[-\frac{\partial^2 p_k}{\partial u^2}(0, q) + \frac{\partial^2 p_k}{\partial u^2}(u, q) \\ & + \frac{\partial^2 p_k}{\partial u^2}(u, q)(\frac{1}{(1 - p_k(u, q) + p_k(u, q) W_k)} - 1)]\end{aligned}$$

We shall use the following fact that holds as soon as $\overline{u}$ is small enough. For any integer k, any q and any $u \leq \overline{u}$:

$$\frac{qu}{1-u} \leq p_k(u, q) \leq 1 - q. \tag{28}$$

Also,

$$1 - \frac{1}{1 - p_k(u, q) + p_k(u, q) W_k} = \frac{p_k(u, q)(W_k - 1)}{1 - p_k(u, q) + p_k(u, q) W_k}$$

and

$$|\frac{1}{1 - p_k(u, q) + p_k(u, q) W_k}| \leq \frac{1}{q_\eta}$$

so that

$$\begin{aligned}E\left(\sup_{q, u \leq \overline{u}} |\frac{\partial^2}{\partial u^2} S_n(u, q) - \frac{\partial^2}{\partial u^2} S_n(0, q)|\right) \quad \leq \quad & \frac{1}{n} \sum_{k=1}^{n} [N^2 h_1(\overline{u}) + \frac{1 + q_\eta}{q_\eta^2} E|W_k - 1|^3 h_2(\overline{u}) \\ & + E|W_k - 1| h_3(\overline{u}) + \frac{1}{q_\eta} E|W_k - 1|^2 h_4(\overline{u})]\end{aligned}$$

with

$$h_1(\overline{u}) \geq E\left(\sup_{q, u \leq \overline{u}} |\frac{\partial p_k}{\partial u}(0, q)^2 - \frac{\partial p_k}{\partial u}(u, q)^2|\right),$$

$$h_2(\overline{u}) \geq E\left(\sup_{q, u \leq \overline{u}} |p_k(u, q) \frac{\partial p_k}{\partial u}(u, q)^2|\right),$$

$$h_3(\overline{u}) \geq E\left(\sup_{q, u \leq \overline{u}} |\frac{\partial^2 p_k}{\partial u^2}(0, q) - \frac{\partial^2 p_k}{\partial u^2}(u, q)|\right)$$

and

$$h_4(\overline{u}) \geq E\left(\sup_{q,u\leq\overline{u}} |p_k(u,q)\frac{\partial^2 p_k}{\partial u^2}(u,q)|\right).$$

Lemma 4.5 will thus follow from the fact that it is possible to choose h_1, h_2, h_3 and h_4 such that for any $i = 1,\ldots,4$, $\lim_{u\to 0} h_i(u) = 0$. From now on, Cte will designate a universal constant number. We shall use the fact **(T2)** that if the deterministic real valued sequence $(z_k)_{k\in\mathbb{N}}$ satisfies, for a positive number δ, for any integer k

$$z_{k+1} \leq M + (1-\delta)z_k$$

and also $z_1 \leq M$, then for any integer k

$$z_k \leq Cte \cdot M.$$

(T2) will often be used with M involving $\overline{u}$, and tending to 0 when $\overline{u}$ does.

Lemma 4.6 *As soon as $\overline{u}$ is small enough,*

$$E\left(\sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q)^2\right) \leq Cte < +\infty$$

Proof of Lemma 4.6.

Direct computation leads to, for any integer k, any q and any $u \leq \overline{u}$:

$$\begin{aligned}\frac{1}{(1-p_{k+1}(u,q))^4}\left(\frac{\partial p_{k+1}}{\partial u}(u,q)\right)^2 \quad \leq \quad & \frac{1}{q^2(1-\overline{u})^4} + \frac{2}{q^4(1-\overline{u})^2}(1-q)W_k\frac{\partial p_k}{\partial u}(u,q)\\ & \frac{(1-\frac{q}{1-u})^2}{(1-\frac{qu}{1-u})^4}W_k^2\frac{1}{(1-p_k(u,q))^4}\left(\frac{\partial p_k}{\partial u}(u,q)\right)^2.\end{aligned}$$

Now there exists $\overline{u}$ small enough such that:

$$\sup_{q,u\leq\overline{u}} \frac{(1-\frac{q}{1-u})^2}{(1-\frac{qu}{1-u})^4}E(W_k^2) \leq 1-\frac{\eta}{2}.$$

Taking the supremum over q and $u \leq \overline{u}$, then taking expectation, and then using **(T2)** one obtains that for any integer k:

$$E\left[\sup_{q,u\leq\overline{u}} \frac{1}{(1-p_k(u,q))^4}\left(\frac{\partial p_k}{\partial u}(u,q)\right)^2\right] \leq Cte < +\infty$$

and lemma 4.6 follows.

Lemma 4.7 *As soon as $\overline{u}$ is small enough,*

$$E\left(\sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q)^{2+\epsilon_*}\right) \leq Cte < +\infty,$$

in which ϵ_* is defined below.

Proof of Lemma 4.7.
Using (A4),

$$3^{\epsilon}\sup_{q\in\mathcal{Q}_\eta}(1-q)^{2+\epsilon}E(W_k)^{2+\epsilon}$$

is a finite, continuous function of (q,ϵ) for $\epsilon<\epsilon_0$, which is upper bounded by $1-\eta$ for $\epsilon=0$, so that there exists a positive ϵ_* such that

$$3^{\epsilon_*}\sup_{q\in\mathcal{Q}_\eta}(1-q)^{2+\epsilon_*}E(W_k)^{2+\epsilon_*}\leq 1-\eta/2. \tag{29}$$

We now have for any integer k, any q and any $u\leq\overline{u}$:

$$\begin{aligned}
&\left(\frac{1}{(1-p_{k+1}(u,q))^2}\frac{\partial p_{k+1}}{\partial u}(u,q)\right)^{2+\epsilon_*}\\
=\;&\left(\frac{1}{(1-p_{k+1}(u,q))^4}\frac{\partial p_{k+1}}{\partial u}(u,q)^2\right)^{1+\epsilon_*/2}\\
\leq\;&\Big(Cte+CteW_k\frac{\partial p_k}{\partial u}(u,q)\\
+\;&\frac{(1-\frac{q}{1-u})^2}{(1-\frac{qu}{1-u})^4}W_k^2\frac{1}{(1-p_k(u,q))^4}\frac{\partial p_k}{\partial u}(u,q)^2\Big)^{1+\epsilon_*/2}\\
\leq\;&3^{\epsilon_*/2}\Big(Cte+CteW_k^{1+\epsilon_*/2}\frac{\partial p_k}{\partial u}(u,q)^{1+\epsilon_*/2}\\
&+\frac{(1-\frac{q}{1-u})^{2+\epsilon_*}}{(1-\frac{qu}{1-u})^{4+2\epsilon_*}}W_k^{2+\epsilon_*}\left(\frac{1}{(1-p_k(u,q))^4}\frac{\partial p_k}{\partial u}(u,q)^2\right)^{1+\epsilon_*/2}\Big).
\end{aligned}$$

Now there exists $\overline{u}$ small enough such that:

$$\sup_{q,u\leq\overline{u}}3^{\epsilon_*}\frac{(1-\frac{q}{1-u})^{2+\epsilon_*}}{(1-\frac{qu}{1-u})^{4+2\epsilon_*}}E(W_k^{2+\epsilon_*})\leq 1-\frac{\eta}{4}.$$

Taking the supremum over q and $u\leq\overline{u}$, then taking expectation, and then using **(T2)** one obtains that for any integer k:

$$E\left[\sup_{q,u\leq\overline{u}}\frac{1}{(1-p_k(u,q))^{4+2\epsilon_*}}\left(\frac{\partial p_k}{\partial u}(u,q)\right)^{2+\epsilon_*}\right]\leq Cte<+\infty$$

and lemma 4.7 follows.

Lemma 4.8 *As soon as $\overline{u}$ is small enough,*

$$E\left(\sup_{q,u\leq\overline{u}}p_k(u,q)\right)\leq Cte\cdot\overline{u}.$$

Proof of Lemma 4.8.
Direct computation leads to, for any integer k, any q and any $u \le \overline{u}$:

$$\frac{p_{k+1}(u,q)}{1-p_{k+1}(u,q)} \le Cte \cdot \overline{u} + \frac{1-q_\eta}{1-\frac{u}{1-u}} W_k \frac{p_k(u,q)}{1-p_k(u,q)}.$$

Notice that $p_1(u,q) = u$. Taking the supremum over q and $u \le \overline{u}$, then taking expectation, and then using **(T2)** one obtains Lemma 4.8.

Lemma 4.9 *As soon as $\overline{u}$ is small enough, it is possible to choose h_2 such that*

$$\lim_{u \to 0} h_2(u) = 0.$$

Proof of Lemma 4.9.
Using Minkowsky's inequality, one may choose

$$h_2(\overline{u}) \le E\left((\sup_{q,u\le\overline{u}} p_k(u,q))^{1+2/\epsilon_*} \right)^{1/(1+2/\epsilon_*)} E\left(\sup_{q,u\le\overline{u}} \frac{\partial p_k}{\partial u}(u,q)^{2+\epsilon_*} \right)^{1/(1+\epsilon_*/2)}$$

Since the p_k are ≤ 1, we have for any integer k, any q and any u

$$p_k(u,q))^{1+2/\epsilon_*} \le p_k(u,q)$$

and using Lemma 4.7 and Lemma 4.8 one may choose

$$h_2(\overline{u}) = Cte \cdot (\overline{u})^{1/(1+2/\epsilon_*)}$$

Lemma 4.10 *As soon as $\overline{u}$ is small enough, it is possible to choose h_1 such that*

$$\lim_{u \to 0} h_1(u) = 0.$$

Proof of Lemma 4.10.
Direct computation leads to, for any integer k, any q and any $u \le \overline{u}$:

$$\begin{aligned} |\frac{\partial p_{k+1}}{\partial u}(u,q) - \frac{\partial p_{k+1}}{\partial u}(0,q)| \quad &\le \quad Cte \cdot \overline{u} + Cte \cdot W_k p_k(u,q) + Cte \cdot \overline{u} W_k \frac{\partial p_k}{\partial u}(u,q) \\ &\quad + Cte \cdot W_k^2 p_k(u,q) \frac{\partial p_k}{\partial u}(u,q) \\ &\quad + (1-q) W_k |\frac{\partial p_k}{\partial u}(u,q) - \frac{\partial p_k}{\partial u}(0,q)|. \end{aligned}$$

Taking the supremum over q and $u \le \overline{u}$, then taking expectation, and then using **(T2)**, Lemma 4.3 and Lemma 4.8, one obtains that for any integer k,

$$E\left(\sup_{q,u\le\overline{u}} |\frac{\partial p_k}{\partial u}(u,q) - \frac{\partial p_k}{\partial u}(0,q)| \right) \le h_5(\overline{u}) \tag{30}$$

with

$$h_5(u) = \frac{1}{q_\eta} \cdot Cte \cdot u. \tag{31}$$

Now, since

$$|\frac{\partial p_k}{\partial u}(u,q)^2 - \frac{\partial p_k}{\partial u}(0,q)^2| = |\frac{\partial p_k}{\partial u}(u,q) - \frac{\partial p_k}{\partial u}(0,q)| \cdot \left(\frac{\partial p_k}{\partial u}(u,q) + \frac{\partial p_k}{\partial u}(0,q)\right)$$

direct computation leads to, for any integer k, any q and any $u \leq \overline{u}$:

$$\begin{aligned}
|\frac{\partial p_{k+1}}{\partial u}(u,q)^2 - \frac{\partial p_{k+1}}{\partial u}(0,q)^2| \quad \leq \quad & Cte \cdot \overline{u} + Cte \cdot \overline{u} \sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q) \\
& + Cte \cdot W_k \sup_{q,u\leq\overline{u}} p_k(u,q) \sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q) \\
& + Cte \cdot \overline{u} W_k \sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q)^2 \\
& + Cte \cdot W_k^2 \sup_{q,u\leq\overline{u}} p_k(u,q) \frac{\partial p_k}{\partial u}(u,q)^2 \\
& + Cte \cdot W_k |\frac{\partial p_k}{\partial u}(u,q) - \frac{\partial p_k}{\partial u}(0,q)| \\
& + Cte \cdot \overline{u} W_k^2 \frac{\partial p_k}{\partial u}(u,q)^2 \\
& + Cte \cdot W_k^3 \sup_{q,u\leq\overline{u}} p_k(u,q) \sup_{q,u\leq\overline{u}} \frac{\partial p_k}{\partial u}(u,q)^2 \\
& + (1-q)^2 W_k^2 |\frac{\partial p_k}{\partial u}(u,q)^2 - \frac{\partial p_k}{\partial u}(0,q)^2|.
\end{aligned}$$

Taking the supremum over q and $u \leq \overline{u}$, then taking expectation, and then using **(T2)**, Lemma 4.6, Lemma 4.8, equation (30) and equation (31) one obtains Lemma 4.10.

Lemma 4.11 *As soon as $\overline{u}$ is small enough,*

$$E\left(\sup_{q,u\leq\overline{u}} |\frac{\partial^2 p_k}{\partial u^2}(u,q)|\right) \leq Cte < +\infty$$

Proof of Lemma 4.11.

First of all, let us give the recursive equations for the second derivative of the prediction filter:

$$\begin{aligned}
\frac{\partial^2 p_{k+1}}{\partial u^2}(u,q) \quad = \quad & \frac{2q}{(1-u)^3} \frac{(1-p_k)}{(1-p_k) + p_k W_k} \\
& - \frac{2q}{(1-u)^2} \frac{W_k}{[(1-p_k)+p_k W_k]^2} \frac{\partial p_k}{\partial u}(u,q) \\
-2\left(1 - \frac{q}{(1-u)}\right) \quad \cdot \quad & (W_k - 1) \frac{W_k}{[(1-p_k)+p_k W_k]^3} \left(\frac{\partial p_k}{\partial u}(u,q)\right)^2 \\
& + \left(1 - \frac{q}{(1-u)}\right) \frac{W_k}{[(1-p_k)+p_k W_k]^2} \frac{\partial^2 p_k}{\partial u^2}(u,q)
\end{aligned} \tag{32}$$

with initial value $\frac{\partial^2 p_1}{\partial u^2}(u,q)=0$. Direct computation leads to, for any integer k, any q and any $u \le \overline{u}$:

$$\begin{aligned}\frac{1}{(1-p_{k+1}(u,q))^2}\left|\frac{\partial^2 p_{k+1}}{\partial u^2}(u,q)\right| \;&\le\; Cte + Cte\cdot W_k\frac{\partial p_k}{\partial u}(u,q) + Cte\cdot W_k^2\frac{\partial p_k}{\partial u}(u,q)^2\\ &+\frac{(1-\frac{q}{1-u})}{(1-\frac{qu}{1-u})^2}W_k\frac{1}{(1-p_k(u,q))^2}\left|\frac{\partial^2 p_k}{\partial u^2}(u,q)\right|\end{aligned}$$

Now there exists $\overline{u}$ small enough such that:

$$\sup_{q,u\le\overline{u}}\frac{(1-\frac{q}{1-u})}{(1-\frac{qu}{1-u})^2}\le 1-\frac{\eta}{2}.$$

Taking the supremum over q and $u \le \overline{u}$, then taking expectation, and then using **(T2)** one obtains that for any integer k:

$$E\left[\sup_{q,u\le\overline{u}}\frac{1}{(1-p_k(u,q))^2}\left|\frac{\partial^2 p_k}{\partial u^2}(u,q)\right|\right]\le Cte<+\infty$$

and lemma 4.11 follows.

Lemma 4.12 *As soon as $\overline{u}$ is small enough, it is possible to choose h_4 such that*

$$\lim_{u\to 0} h_4(u)=0.$$

Proof of Lemma 4.12.
Follows from the recursive inequality:

$$\begin{aligned}\frac{p_{k+1}(u,q)}{(1-p_{k+1}(u,q))^3}\left|\frac{\partial^2 p_{k+1}}{\partial u^2}(u,q)\right| \;&\le\; Cte\cdot p_{k+1}(u,q) + Cte\cdot\overline{u}\cdot W_k\frac{\partial p_k}{\partial u}(u,q)\\ &+Cte\cdot\overline{u}W_k^2\frac{\partial p_k}{\partial u}(u,q)^2\\ &+Cte\cdot W_k^2 p_k(u,q)\frac{\partial p_k}{\partial u}(u,q)\\ &+Cte\cdot W_k^3 p_k(u,q)\frac{\partial p_k}{\partial u}(u,q)^2\\ &+Cte\cdot\overline{u}W_k\left|\frac{\partial^2 p_k}{\partial u^2}(u,q)\right|\\ &+\frac{(1-\frac{q}{1-u})^2}{(1-\frac{qu}{1-u})^3}W_k^2\frac{p_k(u,q)}{(1-p_k(u,q))^3}\left|\frac{\partial^2 p_k}{\partial u^2}(u,q)\right|.\end{aligned}$$

Lemma 4.13 *As soon as $\overline{u}$ is small enough, it is possible to choose h_3 such that*

$$\lim_{u\to 0} h_3(u)=0.$$

Proof of Lemma 4.13.
Follows from the recursive inequality:

$$\begin{aligned}
|\frac{\partial^2 p_{k+1}}{\partial u^2}(u,q)-\frac{\partial^2 p_{k+1}}{\partial u^2}(0,q)| \quad \leq \quad & Cte\cdot\overline{u}+Cte\cdot W_k p_k(u,q) \\
& +Cte\cdot W_k|\frac{\partial p_k}{\partial u}(u,q)-\frac{\partial p_k}{\partial u}(0,q)| \\
& +Cte\cdot\overline{u}W_k\frac{\partial p_k}{\partial u}(u,q)+Cte\cdot W_k^2 p_k(u,q)\frac{\partial p_k}{\partial u}(u,q) \\
& +Cte\cdot W_k^2|\frac{\partial p_k}{\partial u}(u,q)^2-\frac{\partial p_k}{\partial u}(0,q)^2| \\
& +Cte\cdot\overline{u}W_k^2\frac{\partial p_k}{\partial u}(u,q)^2 \\
& +Cte\cdot W_k^3 p_k(u,q)\frac{\partial p_k}{\partial u}(u,q)^2 \\
& +Cte\cdot\overline{u}W_k|\frac{\partial^2 p_k}{\partial u^2}(u,q)| \\
& +Cte\cdot W_k^2 p_k(u,q)|\frac{\partial^2 p_k}{\partial u^2}(u,q)| \\
& +(1-q)W_k|\frac{\partial^2 p_k}{\partial u^2}(u,q)-\frac{\partial^2 p_k}{\partial u^2}(0,q)|.
\end{aligned}$$

4.3 Proof of Proposition 2.4.

To prove Proposition 2.4, we shall prove the finite dimensional marginal convergence of the process and its tightness. We shall use:

Lemma 4.14 *Under the assumptions of Proposition 2.4, for any q in $\mathcal{Q}_\eta$,*

$$\left(\frac{\partial p_k}{\partial u}(0,q);W_k\right)_{k\in\mathbb{N}}$$

is a positive Harris Markov chain, g-uniformly ergodic for $g(x)=1+\parallel x\parallel_2^{2+\epsilon_*}$.

Proof of Lemma 4.14.
It is enough to prove the results of the Lemma for the Markov chain $\Phi=(\frac{\partial p_k}{\partial u}(0,q))_{k\in\mathbb{N}}$. Using (A2), and applying Proposition 4.2.1 of Meyn and Tweedie (1993), Φ is ψ-irreducible with ψ the Lebesgue measure on $[q,+\infty[$. Then, since Φ is a Feller chain by (15), applying Propositions 6.2.8 and 6.2.9 of Meyn and Tweedie (1993), all compact subsets of $[q,+\infty[$ are petite and Φ is a T-chain. Also, Φ is strongly aperiodic. Taking now $g(x)=1+\parallel x\parallel_2^{2+\epsilon_*}$, we have

$$E(V(\frac{\partial p_{k+1}}{\partial u}(0,q))/\frac{\partial p_k}{\partial u}(0,q)=p) \quad = \quad E\left((q+(1-q)\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y)p)^2\right)^{1+\epsilon_*/2}+1$$

$$
\begin{aligned}
&\leq 1+3^{\epsilon_*}\Big(q^{2+\epsilon_*}\\
&\quad +[2q(1-q)]^{1+\epsilon_*/2}E(\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y))^{1+\epsilon_*/2}p^{1+\epsilon_*/2}\\
&\quad +(1-q)^{2+\epsilon_*}E(\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y))^{2+\epsilon_*}(g(p)-1)\Big).
\end{aligned}
$$

The right side of the last inequality divided by $V(p)$ tends to

$$3^{\epsilon_*}(1-q)^{2+\epsilon_*}E(\frac{f_{\gamma_1}}{f_{\gamma_0}}(Y))^{2+\epsilon_*}$$

as p tends to ∞, which is less or equal to $1-\eta/2$, so that there exists a M such that, for $p \geq M$,

$$E(V(\frac{\partial p_{k+1}}{\partial u}(0,q))/\frac{\partial p_k}{\partial u}(0,q)=p) \leq (1-\eta/3)g(p),$$

and we have for some positive b,

$$E(g(\frac{\partial p_{k+1}}{\partial u}(0,q))/\frac{\partial p_k}{\partial u}(0,q)=p)-g(p) \leq -\eta/3\cdot g(p)+b1_{p\leq M}. \tag{33}$$

Consequently Φ is Harris positive by Theorem 11.3.4 and V-uniformly ergodic by Theorem 16.1.2. of Meyn and Tweedie (1993).

Lemma 4.15 *For any integer m and any $q_1,\ldots,q_m$ in $\mathcal{B}_\eta$, the vector*

$$\sqrt{n}(\frac{\partial}{\partial u}S_n(0,q_1),\ldots,\frac{\partial}{\partial u}S_n(0,q_m))$$

converges in distribution to the centered gaussian distribution with covariance $N^2C(q_i,q_j)$.

Proof of Lemma 4.15.
Define

$$M_n = n(\frac{\partial}{\partial u}S_n(0,q_1),\ldots,\frac{\partial}{\partial u}S_n(0,q_m)).$$

M_n is a martingale with respect to the σ-field $\mathcal{F}_n=\sigma(Y_1,\ldots,Y_n)$, with hook

$$\langle M_n\rangle_{i,j} = N^2\sum_{k=1}^n \frac{\partial p_k}{\partial u}(0,q_i)\frac{\partial p_k}{\partial u}(0,q_j).$$

We shall apply Theorem 2.8.41 of Dacunha-Castelle and Duflo (1993). The Lindeberg condition holds since, for any positive δ,

$$\frac{1}{n}\sum_{k=1}^n E[\| M_k-M_{k-1}\|^2 \, 1_{\|M_k-M_{k-1}\|\geq\delta\sqrt{n}}/\mathcal{F}_{k-1}] \leq \frac{Cte}{(\delta\sqrt{n})^{\epsilon_*}}\max_{i=1,\ldots,m}\frac{1}{n}\sum_{k=1}^n(\frac{\partial p_k}{\partial u}(0,q_i))^{2+3\epsilon_*}$$

which converges to 0 as n tends to infinity thanks to Lemma 4.14. Theorem 1.IV.24 of Duflo (1996) may be applied (with $a = 2$ and $\alpha = 1 - \eta$) to the iterative markovian model leading to $(\frac{\partial p_k}{\partial u}(0, q_1), \ldots, \frac{\partial p_k}{\partial u}(0, q_m))$, so that

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\partial p_k}{\partial u}(0, q_i) \frac{\partial p_k}{\partial u}(0, q_j)$$

converges, as n tends to infinity, to the covariance of $\frac{\partial p_k}{\partial u}(0, q_i)$ and $\frac{\partial p_k}{\partial u}(0, q_j)$ with respect to the invariant distribution of the Markov chain. This covariance with respect to the invariance distribution may be computed using equation (15). The obtained limiting value is

$$\frac{1 - (1 - q_1)(1 - q_2)}{1 - (1 - q_1)(1 - q_2)(N^2 + 1)}$$

so that

$$\lim_{n \to +\infty} \frac{1}{n} \langle M_n \rangle_{i,j} = N^2 C(q_i, q_j)$$

and Lemma 4.15 follows.

Lemma 4.16 *The process*

$$\left(\sqrt{n} \frac{\partial}{\partial u} S_n(0, q) \right)_{q \in \mathcal{Q}_\eta}$$

is uniformly tight.

Proof of Lemma 4.16.
For any q_1, q_2 and any positive integer k let

$$Z_k = \left(\frac{\partial p_k}{\partial u}(0, q_1) - \frac{\partial p_k}{\partial u}(0, q_2) \right) (W_k - 1).$$

Using the Rosenthal inequality for martingales (Hall and Heyde, 1980) one has for any $p \geq 2$:

$$\begin{aligned} E \mid \sqrt{n} (\frac{\partial}{\partial u} S_n(0, q_1) - \frac{\partial}{\partial u} S_n(0, q_2)) \mid^p \quad & \leq \quad Cte \cdot E \left[\left(\frac{1}{n} \sum_{k=1}^{n} E(Z_k^2 / Y_1, \ldots, Y_{k-1}) \right)^{p/2} \right] \\ & \quad + \frac{1}{n^{p/2}} \sum_{k=1}^{n} E \mid Z_k \mid^p . \end{aligned} \tag{34}$$

We have now

$$E(Z_k^2 / Y_1, \ldots, Y_{k-1}) = \left(\frac{\partial p_k}{\partial u}(0, q) - \frac{\partial p_k}{\partial u}(0, q_2) \right)^2 N^2.$$

By iterating equation (15), one obtains for any q:

$$\frac{\partial p_k}{\partial u}(0, q) = q + \sum_{l=1}^{k-2} q(1 - q)^l W_{k-1} \cdots W_{k-l} + (1 - q)^{k-1} W_{k-1} \cdots W_1$$

so that

$$\mid \frac{\partial p_k}{\partial u}(0,q_1) - \frac{\partial p_k}{\partial u}(0,q_2) \mid \leq \mid q_1 - q_2 \mid [1 + \sum_{l=1}^{k-1} (1-\tilde{q})^{l-1} \mid 1 - (l+1)\tilde{q} \mid W_{k-1} \cdots W_{k-l}]$$

where $\tilde{q}$ is some number between q_1 and q_2.
Define

$$\alpha = 1 - \frac{2}{2+\epsilon_*} \frac{\log 1 - \eta/4}{\log 1 - q_\eta}.$$

We have $0 < \alpha < 1$, and for big enough l one has

$$(l+1)(1-q_\eta)^{l-1} \leq (1-q_\eta)^{\alpha l},$$

so that:

$$\mid \frac{\partial p_k}{\partial u}(0,q_1) - \frac{\partial p_k}{\partial u}(0,q_2) \mid \leq Cte \mid q_1 - q_2 \mid [1 + \sum_{l=1}^{k-1} (1-q_\eta)^{\alpha l} W_{k-1} \cdots W_{k-l}]. \tag{35}$$

Define the Markov chain $(D_k)_{k \in \mathbb{N}}$ by $D_1 = 0$ and

$$D_{k+1} = 1 + (1-q_\eta)^\alpha W_k D_k.$$

Following the arguments of the proof of Lemma 4.14, it can be proved that this Markov chain is Harris positive and $1 + x^{2+\epsilon_*}$-uniformly ergodic. Now, equation (35) implies that

$$\mid \frac{\partial p_k}{\partial u}(0,q_1) - \frac{\partial p_k}{\partial u}(0,q_2) \mid \leq Cte \mid q_1 - q_2 \mid D_k. \tag{36}$$

We thus obtain

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_{k=1}^n E(Z_k^2/Y_1,\ldots,Y_{k-1})\right)^{p/2}\right] &\leq Cte \mid q_1 - q_2 \mid^p E\left[\left(\frac{1}{n}\sum_{k=1}^n D_k^2\right)^{p/2}\right] \\ &\leq Cte \mid q_1 - q_2 \mid^p E\left[\frac{1}{n}\sum_{k=1}^n D_k^p\right] \end{aligned} \tag{37}$$

where the last inequality follows by Jensen's inequality since $p/2 \geq 1$. But, as soon as $p \leq 2 + \epsilon_*$, $\frac{1}{n}\sum_{k=1}^n ED_k^p$ converges and we get:

$$E\left[\left(\frac{1}{n}\sum_{k=1}^n E(Z_k^2/Y_1,\ldots,Y_{k-1})\right)^{p/2}\right] \leq Cte \mid q_1 - q_2 \mid^p \tag{38}$$

for big enough n. Also, using (36), as soon as $p \leq 2 + \epsilon_*$:

$$\frac{1}{n^{p/2}} \sum_{k=1}^n E \mid Z_k \mid^p \leq Cte \mid q_1 - q_2 \mid^p \frac{1}{n^{p/2}} \sum_{k=1}^n ED_k^p \leq Cte \mid q_1 - q_2 \mid^p \tag{39}$$

for big enough n.
It follows from equations (34), (38) and (39) that for big enough n, for any q_1 and q_2,

$$E \mid \sqrt{n}(\frac{\partial}{\partial u}S_n(0,q_1) - \frac{\partial}{\partial u}S_n(0,q_2)) \mid^{2+\epsilon_*} \leq Cte \cdot (q_1 - q_2)^{2+\epsilon_*}, \tag{40}$$

which is the Kolmogorov criterion for the uniform tightness of a continuous process, and Lemma 4.16 follows.

References

[1] Atwood L.D., Wilson A.F., Bailey-Wilson J.E., Carruth J.N., Elston R.C. On the distribution of the likelihood ratio test statistic for a mixture of two normal distributions. *Commun. Statist.-Simula.*, 25(3):733–740, 1996.

[2] Baum, L.E. and Petrie, T. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37, 1554-1563, 1966.

[3] Bickel, P.J. and Ritov, Y. Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli*, 2, 199-228, 1996.

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

[5] Chuang R-J., Mendell N.R. The approximate null distribution of the likelihood ratio test for a mixture of two bivariate normal distributions with equal covariance. *Commun. Statist.-Simula.*, 26(2):631–648, 1997.

[6] Churchill, G.A. Stochastic models for heterogeneous DNA sequences. *Bulletin of Math. Biology*, 51, 79-94, 1989.

[7] Ciuperca, G. Sur le test de maximum de vraisemblance pour le mélange de populations *Note aux C.R.A.S.*, 328, série I, 4, 351-358, 1999.

[8] Dacunha-Castelle, D. and Duflo, M. *Probabilités et statistiques, tome 2.* Masson, 1993.

[9] Dacunha-Castelle, D. and Gassiat, E. Estimation of the number of components in a mixture *Bernoulli*, 3, 279-299, 1997a.

[10] Dacunha-Castelle, D. and Gassiat, E. Testing in locally conic models. *ESAIM Prob. and stat.*, 1, 1997b.

[11] Dacunha-Castelle, D. and Gassiat, E. Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Ann. Statist.*, (submitted) 1996.

[12] Dempster A.P., Laird N.M., Rubin D.B. Large Maximum-likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B*, 39:1–38, 1977.

[13] Duflo, M. *Algorithmes stochastiques.* Springer, 1996.

[14] Feng Z.D., McCulloch C.E. Using bootstrap Likelihood Ratio in Finite Mixture Models. *J. R. Statist. Soc. B*, 58(3):609–617, 1996.

[15] Finesso, L. *Consistent Estimation of the Order for Markov and Hidden Markov Chains* Ph. D. Thesis, University of Maryland, 1990.

[16] Fredkin, D.R. and Rice, J.A. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Royal Soc. London Ser. B*, 249, 125-132, 1992.

[17] Hall, P. and Heyde, C.C. *Martingale Limit Theory and Its Application.* Academic Press, 1980.

[18] Hartigan, J.A. A failure of likelihood ratio asymptotics for normal mixtures. *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, L.M. Le Cam and R.A. Olshen eds, 807-810, 1985.

[19] Henna, J. On estimating the number of constituents of a finite mixture of continuous distributions. *Ann. Inst. Statist. Math.*, 37, 235-240, 1985.

[20] Keribin, C. Consistent estimation of the Order of Mixture Models. *submitted*, 1997.

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

[22] Leroux, B.G. and Puterman, M.L. Maximum Penalized Likelihood Estimation for Independent and Markov-Dependent Mixture Models. *Biometrics*, 48, 545-558, 1992.

[23] Lindsay, B.G. *Mixture models: Theory, Geometry and Applications.* 1995.

[24] Mac Donald, I.L. and Zucchini, W. *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman and Hall, 1997.

[25] McLachlan G.J. On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture. *App. Statist.*, 36(3):318–324, 1987.

[26] Mevel, L. *Statistique asymptotique pour les modèles de Markov cachés* Thèse de l'Université de Rennes I, 1997.

[27] Mevel, L. and LeGland F Exponential forgetting and Geometric Ergodicity in Hidden Markov models. *Mathematics of Control, Signals, and Systems*, to appear.

[28] Meyn, S.P. and Tweedie, R.L. *Markov chains and stochastic stability* Springer-Verlag, 1993.

[29] Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77, 257-284, 1989.

[30] Ryden, T. Estimating the order of hidden Markov models. *Statistics*, 26, 345-354, 1995.

[31] Vandekerkhove, P. *Identification de l'ordre des processus ARMA stables. Contribution à l'étude statistique des chaines de Markov cachées.* Thèse de l'Université de Montpellier II, 1997.

[32] Van der Vaart, A. *Asymptotic Statistics.* Cambridge ed., 1999.