La méthode de Leroux pour les modèles de Markov cachés généraux.

Valentine GENON-CATALOT* et Catherine LAREDO**

**Université René Descartes Paris 5*
U.F.R. de Mathématique et Informatique
45, rue des Saints-Pères, 75270 Paris Cedex 06, France, tel.: 00 33 (1) 42 86 21 14, fax: 00 33 (1) 44 55 35 35, e-mail: genon@math-info.univ-paris5.fr

***I.N.R.A.*
Laboratoire de Biométrie, 78350 Jouy-en-Josas, France, e-mail:cl@banian.jouy.inra.fr

Résumé

En 1992, B. Leroux a développé une méthode pour étudier la vraisemblance des modèles de Markov cachés lorsque la chaîne cachée a un espace d'état fini. Dans cet article, nous étudions l'extension de la méthode de Leroux au cas où la chaîne cachée a pour espace d'état un intervalle ouvert de la droite réelle. Sous des hypothèses qui nous semblent minimales, nous obtenons la convergence de la log-vraisemblance normalisée vers une limite. Nous identifions cette limite pour la vraie valeur du paramètre. Nous illustrons notre méthode, nos hypothèses et nos résultats sur le filtre de Kalman.

Abstract

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

Introduction.

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

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

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

- Problem (1): Estimation of the unobserved variable U_n (resp. U_{n+1}) from past observations $Z_n, \ldots, Z_1$. This is the problem of filtering (resp. prediction) in discrete time.

- Problem (2): Statistical inference based on $(Z_1, \ldots, Z_n)$ generally with the aim of estimating unknown parameters in the distribution of (U_n).

In the litterature interested in problem (1), it is generally assumed that the spate space $\mathcal{U}$ of (U_n) is a subset of an Euclidian space. In papers dealing with problem (2), it is more often assumed that the hidden chain (U_n) has a finite state space $\mathcal{U} = \{u_1, \ldots, u_m\}$ and one wants to estimate its transition probabilities.

In this paper, we are interested in problem (2), when the state variable (U_n) evolves in an open interval $\mathcal{U} = (l, r)$ of the real line, with $-\infty \leq l < r \leq +\infty$. Moreover, we assume below that the hidden chain $(U_n, n \in Z)$ is strictly stationary and ergodic, and that the conditional distribution of Z_n given $U_n = u$ does not depend on n (for instance, in (1), it is the distribution of $G(u, \varepsilon_1)$). Under these assumptions, it is well known that the joint process $((U_n, Z_n), n \in Z)$ is also strictly stationary and ergodic (see *e.g.* Leroux (1992)). We assume that we observe $Z_1, \ldots, Z_n$ extracted from the ergodic sequence $(Z_n, n \in Z)$.

In this set-up, our aim is to study parametric inference based on the exact likelihood of $Z_1, \ldots, Z_n$. Before giving details on the content of our paper, let us present the results and open problems in this domain.

In a seminal paper, Leroux (1992), assuming that the state space of (U_n) is a finite set, proves the convergence of the normalized log-likelihood of $(Z_1, \ldots, Z_n)$ and the consistency of the exact maximum likelihood estimator (MLE). The impressive feature of Leroux's paper is that his results are obtained under minimal assumptions.

Relying on the consistency result proved by Leroux, Bickel et al (1998) prove the asymptotic normality of the exact MLE. Then, these results are extended to the case where $\mathcal{U}$ is a compact set by Douc and Matias (2001).

For a general state space of (U_n), the asymptotic behaviour of the exact likelihood of $(Z_1, \ldots, Z_n)$ is still open, and consequently, the asymptotic behaviour of the exact MLE is not known.

However, there is a well known model which makes exception and is completely solved, namely the Kalman filter. In its simplest form, it may be described as follows. Let (U_n) be a one-dimensional Gaussian AR(1)- process:

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

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

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

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

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

Here is the outline the paper. We follow step by step Leroux's paper preserving its spirit in the sense of obtaining results under minimal assumptions, and we point out the analogies and the differences.

The framework is the following: the unobserved Markov chain (U_n) has state space $\mathcal{U} = (l, r)$ an open interval $(-\infty \leq l < r \leq +\infty)$. Its transition operator P_θ depends on an unknown parameter θ and transition probabilities $P_\theta(u, dv) = p(\theta, u, v)dv$ have densities with respect to the Lebesgue measure of $\mathcal{U}$ (denoted by dv). For simplicity, the conditional distribution of Z_n given $U_n = u$, say $F_u(dz)$, contains no additional unknown parameter. We assume that , when u is considered as a parameter, $F_u(dz) = f(z/u)\mu(dz)$ defines a standard dominated regular family of distributions with $f(z/u) > 0$ and, for all z, (μ-a.e.), $u \to f(z/u)$ continuous and bounded on $\mathcal{U}$. The exact likelihood of $(Z_1, \ldots, Z_n)$ may be computed by several classical formulae that we recall. One way to obtain it is to compute first the conditional density of $(Z_1, \ldots, Z_n)$ given $U_1 = u$, say $p_n(\theta, z_1, \ldots, z_n/u)$ and then integrate with respect to the distribution of U_1. More generally, for any probability density g on $\mathcal{U}$, we define the functions

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

and set $p_n^g(\theta) = p_n^g(\theta, Z_1, \ldots, Z_n)$. When g is the exact density of U_1, $p_n^g(\theta)$ is the exact likelihood function, that we simply denote below by $p_n(\theta)$. Otherwise, we call $p_n^g(\theta)$ a contrast process. As usual we denote by θ_0 the true value of the parameter. We prove that, for all positive and continuous density g on $\mathcal{U}$, $\frac{1}{n}\log p_n^g(\theta)$ converges, in $\mathbb{P}_{\theta_0}$-probability, to the same limit $H(\theta_0, \theta)$. This is obtained in two steps. First, we set, as in Leroux (1992),

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

and we call $q_n(\theta) = q_n(\theta, Z_1, \ldots, Z_n)$ the Leroux contrast. Since $\mathcal{U}$ is neither finite nor compact, we need an adequate assumption in order to prove that $q_n(\theta)$ is well defined for all n: this is reached by simply assuming that the transition operator P_θ of (U_n) is Feller, a property shared by all standard Markov chains on Euclidian spaces. Then, we are able to mimick Leroux's proof. Under a very weak moment assumption, we prove that $\frac{1}{n}\log q_n(\theta)$ converges, $\mathbb{P}_{\theta_0}$-a.s., to a limit $H(\theta_0, \theta) \in [-\infty, +\infty)$. Since this result is obtained by Kingman's ergodic theorem (Kingman (1976)), the limit is not given as the expectation of some random variable as for the classical ergodic theorem. We prove that $\frac{1}{n}\log q_n(\theta)$ and $\frac{1}{n}\log p_n^g(\theta)$ have the same limit in $\mathbb{P}_{\theta_0}$-probability, for all positive and continuous g. Whereas in Leroux's paper, this step is immediate, it is in our context harder and requires additional assumptions. The main new assumption is that the sequence of random variables

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

is $\mathbb{P}_{\theta_0}$-tight, for all θ. Then, the comparison of $q_n(\theta)$ and $p_n^g(\theta)$ is obtained by the Laplace method which roughly says that $p_n^g(\theta)$ (see 4) is logarithmically equivalent to $q_n(\theta)$.

Then, we identify the limit $H(\theta_0, \theta_0)$. This is done by obtaining the limit of $\frac{1}{n}\log p_n(\theta_0)$, the exact likelihood at θ_0, through another approach. The method uses conditional expectations and a martingale limit theorem which only holds at θ_0. It first requires a precise insight into the prediction algorithm which allows to compute recursively the successive conditional distributions of U_n given $Z_{n-1}, \ldots, Z_1$. Then, we study the conditional distributions, under $\mathbb{P}_{\theta_0}$, of U_n given the finite past $Z_{n-1}, \ldots, Z_{n-p}$ and the infinite past $\underline{Z}_{n-1} = (Z_{n-1}, Z_{n-2}, \ldots)$. We prove that the conditional distribution of U_n given $\underline{Z}_{n-1}$ (under $\mathbb{P}_{\theta_0}$) has a continuous density $\tilde{g}(\theta_0, u/\underline{Z}_{n-1})$ with respect to the Lebesgue measure on $\mathcal{U}$ and that the conditional distribution of U_n given $Z_{n-1}, \ldots, Z_{n-p}$ weakly converges, as p tends to infinity, $\mathbb{P}_{\theta_0}$-a.s., to the above distribution. Also, the process $((U_n, Z_n, \tilde{g}(\theta_0, u/\underline{Z}_{n-1}du), n \in Z)$ is a stationary version of the Markov process $(U_n, Z_n, \mathcal{L}_{\mathbb{P}_{\theta_0}}(U_n/Z_{n-1}, \ldots, Z_1), n \geq 1)$. Finally, we use these results to obtain the limit of $\frac{1}{n}\log p_n(\theta_0)$. This limit is linked with the entropy of the conditional distribution under $\mathbb{P}_{\theta_0}$ of Z_1 given the infinite past $\underline{Z}_0$.

Finally, we study in full details the Kalman filter model (see (2)-(3)). We prove that it satisfies all our assumptions. The computation of the limit $H(\theta_0, \theta)$ for all θ (not only for θ_0) is also explicit and obtained by using the limit of $\frac{1}{n}\log p_n^g(\theta)$ for a well chosen density

g. We recover the property that $H(\theta_0, \theta_0) - H(\theta_0, \theta)$ is the expectation of a Kullback information, a result obtained in a previous paper (see Genon-Catalot et al, 2003).

Bibliographie

[1] Bickel P.J., Ritov Y. and Ryden T., (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614-1635. *Bernoulli* **2** (3), 199-228.
[2] Douc R. and Matias C. (2001), Asymptotics of the maximum likelihood estimator for general hidden Markov models, *Bernoulli* **7**(3), 381-420.
[3] 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.
[4] Kingman J.F.C., 1976, Subadditive processes, in: P.L. Hennequin ed., Ecole d'été de probabilités de Saint-Flour V-1976, Lecture Notes in Math. 539 (Springer Berlin, 1976), 167-223.
[5] Leroux B.G. (1992). Maximum likelihood estimation for hidden Markov models. *Stoch. Proc. and their appl.* **40**, 127-143.