Consistency of Bayesian nonparametric Hidden Markov Models

Elodie Vernet[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 41

[1] Laboratoire de Mathématiques, Université Paris-Sud, Orsay, France
elodie.vernet@math.u-psud.fr

Abstract

We are interested in Bayesian nonparametric Hidden Markov Models. More precisely, we are going to prove the consistency of these models under appropriate conditions on the prior distribution and when the number of states of the Markov Chain is finite and known. Our approach is based on exponential forgetting and usual Bayesian consistency techniques.

Keywords: Bayesian nonparametrics Hidden Markov Model consistency.

1 Introduction

Hidden Markov Models are much used in practice as in econometrics, speech recognition, genomics (see [2] for some applications)... Frequentist methods to deal with hidden Markov chains are asymptotically understood [4] [3]. Conversely, asymptotic properties of Bayesian methods have not been much considered. The Bayesian parametric case has just been studied in [5] and [6]. While Yau & al. empirically noticed in [7] that using a nonparametric model may improve the estimations a lot, there is no asymptotic result in this case. In this paper we will show the posterior consistency of nonparametric Bayesian hidden Markov models under rather usual assumptions.

2 Hidden Markov models

Let $S_1, \ldots, S_T$ be a homogeneous Markov chain with a finite and known number of states k, an initial probability ν and a kxk transition matrix Q.

In hidden Markov models, we cannot access these previous Markov chain states (they are hidden). But we observe $Y_1, \dots, Y_T$ which are the noisy signals of the states of the chain. Given $S_1, \dots, S_T$; $Y_1, \dots, Y_T$ are independent and Y_t is equal to a parameter m_{S_t} depending on the corresponding state S_t plus a noise ϵ_t. We assume that $\epsilon_1, \dots, \epsilon_T$ are iid, distributed according to a probability F and independent of the Markov chain.

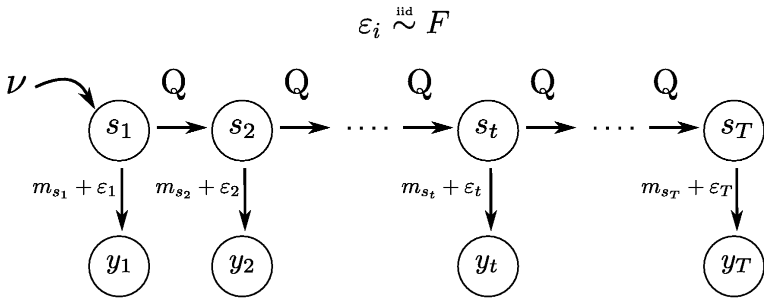

Figure 1: The model

The parameters of this model are ν, Q, $m = (m_1, \dots, m_k)$ and F. We assume that F has a density f with respect to a reference measure λ. Let $\theta = (\nu, Q, m, f)$ and p_θ^T be the associated joint density of $Y_1, \dots Y_T$ with respect to λ.

As usual in Bayesian statistics, we put a prior μ on the parameters. Then we take a "frequentist point of view" by wondering if the posterior asymptotically puts the mass on the neighborhood of the true density. In other words, we study the posterior consistency.

3 Consistency

Consistency is the first thing we may ask for an estimator. Here we will work with neighborhood with respect to the l_1 distance between two joint densities. For two parameters θ and θ' the pseudo-metric between the two of them will be $\int |p_\theta(y_1, \dots, y_l) - p_{\theta'}(y_1, \dots, y_l)| \lambda(dy_1) \dots \lambda(dy_l)$.

Theorem 1 *Under some assumptions on the set of parameters and the prior, if the true parameter $\theta^* = (\nu^*, Q^*, m^*, f^*)$ is such that the associated Markov chain mixes enough and the tail of f^* is small enough then the posterior is consistent with respect to the previously described pseudo-metric.*

This result is proved using Barron method [4]. That is to say we have to prove that the parameter set is not too big by proving the existence of the following tests. For all neighborhood A_n of the true parameter these tests enable to separate the true parameter from a set C_n such that the union of A_n and C_n has a small prior measure. This task can be achieved by the construction made in [5]. Secondly, we have to prove that the posterior puts enough mass in the

Kullback-Leibler neighborhood of the true density. For this purpose, we need to control a nasty Kullback-Leibler divergence by controlling the parameters. We did it thanks to existing results on hidden Markov chains [3] and [4].

The assumptions on the set of parameters are usual. We mostly ask that the Markov chain mixes enough. Notice that the construction of tests lead us to ask for priors which do not put mass on transition matrix Q such that $Q_{i,j} < \underline{q}$ for a constant $\underline{q} > 0$. The other assumptions on the prior consist on usual nonparametric assumptions. These last assumptions are checked for some Dirichlet mixtures on f.

4 Conclusion

Thanks to the previous results, we hope to identify the adaptive rate of convergence associated to this model.

References

[1] A.R. Barron. **The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions**. *Technical report*; April 1988.

[2] O. Cappé, E. Moulines, T. Rydén. **Inference in hidden Markov Models**. *Springer*; 2012.

[3] R. Douc, C. Matias. **Asymptotics of the maximum likelihood estimator for general hidden Markov models**. *Bernouilli*; 2001; 7; pp. 381–420.

[4] R. Douc, E. Moulines, T. Rydén. **Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime**. *The annals of statistics*; 2004; 32(5); pp. 2254-2304.

[5] E. Gassiat, J. Rousseau. **Non parametric finite translation mixtures with dependent regime**. *Submitted in 2013*; 2013.

[6] M. C. de Gunst, O. Shcherbakova. **Asymptotic behavior of Bayes estimators for hidden Markov models with application to ion channels**. *Mathematical Methods of Statistics*; 2008; 17(4); pp. 342-356.

[7] C. Yau, O. Papaspiliopoulos, G.O. Roberts, C. Holmes. **Bayesian nonparametric hidden Markov models with applications in genomics**. *Journal of the Royal Statistical Society*; 2011; 73; pp.37-57.