Estimation de l'ordre des modèles à chaîne de Markov cachée non paramétriques

Luc Lehéricy[1]

[1]*Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France*
luc.lehericy@math.u-psud.fr

Résumé. Jusqu'à récemment, la plupart des garanties théoriques sur l'estimation de modèles à chaîne de Markov cachée portaient sur le cadre paramétrique, c'est-à-dire lorsqu'on restreint les paramètres à un espace de dimension finie. Des travaux récents ont permis de démontrer des garanties dans le cadre non paramétrique, à condition que le nombre d'états cachés, supposé fini, soit connu. Des méthodes existent pour estimer ce nombre, mais elles nécessitent de supposer qu'il est borné par une constante connue (on parle alors de *borne a priori*), ce qui n'est pas toujours possible. Nous présentons deux méthodes pour l'estimer dans un cadre non paramétrique en se passant de borne a priori. La première méthode consiste à minimiser sur une famille de modèles un critère empirique des moindres carrés puis à choisir le nombre d'états par pénalisation du critère. La seconde repose sur l'analyse du spectre de tenseurs empiriques. Nous démontrons la consistance des estimateurs obtenus et comparons numériquement leurs performances.

Mots-clés. Modèles à chaîne de Markov cachée, estimation non paramétrique, sélection de modèle, méthodes spectrales, estimation d'ordre.

Abstract. Until recently, most theoretical guarantees on hidden Markov model estimation focused on the parametric case, that is to say when the parameters are assumed to belong to a finite-dimensional space. Recent works have proven guarantees in a nonparametric framework, provided that the number of hidden states, which we assume to be finite, is known. There are methods to estimate this number, but they need it to be bounded by a known constant (we call such a constant an *a priori bound*), which is not always possible. We present two methods to estimate it in a nonparametric framework without any a priori bound. The first one consists in minimizing an empirical least squares criterion on a family of models and then selecting one of these models by penalizing the criterion. The second one relies on the analysis of the spectrum of an empirical tensor. We show strong consistency of both estimators and numerically compare their ability to select the right number of states.

Keywords. Hidden Markov models, nonparametric estimation, model selection, spectral methods, order estimation.

1 Introduction et contexte

Les modèles à chaîne de Markov cachée (abrégé dans la suite en HMMs pour hidden Markov models) sont utilisés dans de nombreuses applications pour leur capacité à rendre compte d'une structure latente aux données. Comme pour les modèles de mélange, on suppose que les observations sont générées selon des lois dépendant d'un paramètre caché, par exemple la classe dans une population hétérogène. Toutefois, contrairement à un modèle de mélange où on suppose que la classe de chaque observation est tirée indépendamment des autres, les HMMs modélisent le paramètre caché par un processus markovien. Cela permet de prendre en compte une dépendance temporelle ou spatiale entre les observations.

Plus formellement, un HMM est un processus markovien $(X_t, Y_t)_t$. On appelle X_t l'état caché associé à l'observation Y_t. On suppose que $(X_t)_t$ est une chaîne de Markov et que les observations Y_t sont indépendantes conditionnellement à $(X_t)_t$ de loi dépendant uniquement de leur état caché. L'objectif est de retrouver les paramètres de la chaîne de Markov $(X_t)_t$ et la loi de Y_t conditionnellement à X_t uniquement à partir des observations $(Y_t)_t$.

Nous supposons que le processus caché est à valeurs dans un espace fini et que la loi de l'observation Y_t conditionnellement à l'évènement $\{X_t = k\}$ admet une densité par rapport à la mesure dominante, qu'on appellera *densité d'émission* de l'état k. On appelle le nombre d'états cachés l'*ordre* du modèle. Les paramètres du HMM sont alors l'ordre du modèle, la loi initiale et la matrice de transition de la chaîne cachée ainsi que les densités d'émission.

Gassiat et al. (2013) montrent qu'il suffit de connaître la loi de trois observations consécutives pour retrouver les paramètres du HMM lorsque la chaîne de Markov est ergodique, que la matrice de transition est inversible et que les densités d'émission forment une famille libre. C'est dans ce cas que nous nous plaçons.

2 Méthodes d'estimation

De Castro et al. (2016) supposent l'ordre du modèle connu et qu'ils disposent d'un estimateur auxiliaire des paramètres de la chaîne de Markov. Dans ce cadre, ils introduisent une méthode d'estimation non paramétrique des densités d'émission par minimisation d'un critère des moindres carrés pénalisé. Ils démontrent une inégalité oracle sur les estimateurs obtenus et montrent qu'ils convergent à la vitesse minimax de manière adaptative.

Nous généralisons leur approche pour pouvoir nous passer d'un estimateur auxiliaire et estimer d'un coup non seulement les densités d'émission mais également les paramètres de la chaîne de Markov et l'ordre du modèle. Outre sa capacité à estimer l'ordre du modèle, notre méthode reste adaptative et minimax et permet d'éviter les problèmes liés à l'estimateur auxiliaire. Notamment, les simulations montrent que notre méthode est nettement plus performante que celles fondées sur les estimateurs spectraux de de Castro et al. (2015) lorsque les densités d'émission sont presque liées.

Nous introduisons également une seconde méthode consistante d'estimation de l'ordre du modèle fondée sur l'analyse du spectre du tenseur obtenu en décomposant la densité de la loi de deux observations consécutives sur une base orthonormée : sous nos hypothèses, estimer l'ordre du modèle revient à estimer le rang de ce tenseur.

Ces deux méthodes non paramétriques sont détaillées dans Lehéricy (2016), qui montre également qu'elles sont toutes les deux fortement consistantes. C'est à notre connaissance le seul résultat de consistance dans un cadre général, paramétrique ou non, qui ne nécessite pas de borne a priori sur l'ordre du modèle.

3 Remarques bibliographiques

La plupart des résultats théoriques concernant l'estimation des paramètres portent sur le cadre paramétrique, c'est-à-dire lorsqu'on suppose que les densités d'émission sont dans un espace de dimension finie. Les seuls résultats théoriques dont nous ayons connaissance en non paramétrique sont de Castro et al. (2015) qui étudient une variante de la méthode spectrale de Anandkumar et al. (2012), et de Castro et al. (2016) et Lehéricy (2016) qui étudient des estimateurs des moindres carrés pénalisés. En particulier, aucune garantie non asymptotique n'a été démontrée concernant l'estimateur du maximum de vraisemblance pénalisé. Une autre limite des estimateurs adaptatifs existants est qu'ils ne sont pas capables de s'adapter à des régularités différentes pour chaque densité d'émission. Ces deux points font l'objet de travaux à venir.

A propos de l'estimation d'ordre, citons Gassiat et Boucheron (2003) et Gassiat (2002) qui montrent qu'un estimateur du maximum de vraisemblance pénalisé permet d'estimer l'ordre de façon consistante respectivement sans borne a priori lorsque les observations ne prennent qu'un nombre fini de valeurs, et dans un espace quelconque lorsqu'on dispose d'une borne a priori sur le modèle. Dans un cadre bayésien, Gassiat et al. (2014) et van Havre et al. (2016) obtiennent des résultats de consistance qui requièrent également une borne a priori.

Bibliographie

[1] Animashree Anandkumar, Daniel Hsu et Sham M. Kakade (2012), *A method of moments for mixture models and hidden Markov models*, COLT (Vol. 1, No. 4, p. 4).

[2] Yohann de Castro, Elisabeth Gassiat et Claire Lacour (2016), *Minimax adaptative estimation of nonparametric hidden Markov models*, Journal of Machine Learning Research, 17(111):1–43.

[3] Yohann de Castro, Elisabeth Gassiat et Sylvain Le Corff (2015), *Consistent estimation of the filtering and marginal smoothing distributions in nonparametric hidden Markov*

models, arXiv:1507.06510.
[4] Elisabeth Gassiat (2002), *Likelihood ratio inequalities with applications to various mixtures*, Annales de l'IHP Probabilités et statistiques, Vol. 38, No. 6, pp. 897–906.
[5] Elisabeth Gassiat, Alice Cleynen et Stéphane Robin (2013), *Finite state space non parametric hidden Markov models are in general identifiable*, arXiv:1306.4657.
[6] Elisabeth Gassiat et Judith Rousseau (2014), *About the posterior distribution in hidden Markov models with unknown number of states*, Bernoulli, 20(4), 2039-2075.
[7] Elisabeth Gassiat et Stéphane Boucheron (2003), *Optimal error exponents in hidden Markov models order estimation*, IEEE Transactions on Information Theory, 49(4), 964-980.
[8] Luc Lehéricy (2016), *Order estimation for non-parametric hidden Markov models*, arXiv:1606.00622.
[9] Zoé van Havre, Judith Rousseau, Nicole White et Kerrie Mengersen (2016), *Overfitting hidden Markov models with an unknown number of states*, arXiv:1602.02466.