C. R. Acad. Sci. Paris, t. 325, Série I, p. 203-206, 1997
Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Statistique de chaînes de Markov cachées à espace d'états fini. Le cas non stationnaire

Dominique BAKRY, Xavier MILHAUD et Pierre VANDEKERKHOVE

D. B. et X. M. : Université Paul-Sabatier, Laboratoire de Statistique et Probabilités, URA CNRS 745, 118, route de Narbonne, 31062 Toulouse CEDEX, France.

P. V. : Université Montpellier-II, Laboratoire de Probabilités et Statistique, EA 720, place Eugène-Bataillon, 34095 Montpellier CEDEX 5, France.

Résumé. L'objet de cette Note est d'étudier, sous des conditions simples, la normalité asymptotique locale du modèle statistique associé à une chaîne de Markov cachée homogène, à espace fini d'états, de condition initiale déterministe, ainsi que la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance.

Statistics of hidden Markov chains, finite state space, nonstationary case

Abstract. *This Note is concerned with the inferential properties of nonstationary hidden Markov chain models. We prove, under mild conditions, local asymptotic normality of the statistical model, together with the consistency and the asymptotic normality of the maximum likelihood estimator.*

1. Introduction

Les chaînes de Markov cachées, qui sont des fonctions aléatoires de chaînes de Markov, ont été introduites par Baum et Petrie (*voir* [1]) à la fin des années soixante.

Il faut attendre environ une quinzaine d'années pour trouver leurs premières applications. Aujourd'hui ces modèles sont couramment utilisés, par exemple, en reconnaissance de la parole par Rabiner (*voir* [5]), pour modéliser des périodes d'excitation dans les canaux d'ions par Ball et Rice (*voir* [2]), ou encore par Churchill (*voir* [4]), pour modéliser l'hétérogénéité des séquences d'acides aminés dans l'ADN.

L'évolution actuelle, dans ce domaine, consiste essentiellement à proposer des algorithmes d'ajustement de ces modèles, alors que peu de travaux traitent réellement de leurs propriétés inférentielles. Citons les travaux de Baum et Petrie (*voir* [1]) (en bref B.P.) sur les chaînes de

Note présentée par Paul DEHEUVELS.

0764-4442/97/03250203 © Académie des Sciences/Elsevier, Paris

Markov cachées à espace fini d'états, stationnaires, et plus récemment ceux de Leroux (*voir* [7]), sur la convergence de l'estimateur du maximum de vraisemblance pour des chaînes de Markov cachées à espace d'états quelconques. Actuellement, Rydén (*voir* [9]) étudie des procédures d'estimation et de test basées sur la vraisemblance, tandis que Bickel et Ritov (*voir* [3]) s'intéressent à l'efficacité, et à la normalité asymptotique de l'estimateur du maximum de vraisemblance (EMV) pour des modèles de Markov cachés stationnaires.

Nous proposons dans cette Note de montrer la propriété de normalité asymptotique locale des structures statistiques associées aux chaînes de Markov cachées à espace fini d'états, de condition initiale déterministe, ainsi que la convergence et la normalité asymptotique de l'EMV dans ce cadre-là.

2. Préliminaires et résultats

Soit $((\mathbb{X}\times\mathbb{Y})^{\mathbb{N}},\ (\mathcal{X}\otimes\mathcal{Y})^{\otimes\mathbb{N}},\ (Z_n)_{n\geq 0}=(X_n,\ Y_n)_{n\geq 0},\ P_\theta,\ \nu_\theta,\ \theta\in\Theta)$, la structure statistique associée à la chaîne de Markov $Z=(Z_n)_{n\geq 0}$ de transition P_θ, de loi initiale ν_θ, et de paramètre θ appartenant à l'espace paramétrique Θ. Nous supposons que pour chaque $\theta\in\Theta$, nous avons :

- $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène, récurrente positive, à valeurs dans un espace d'états finis $\mathbb{X}$, de transition

$$\pi_\theta(x_n,\ x_{n+1})=P_\theta(X_{n+1}=x_{n+1}\,|\,X_n=x_n).$$

- Pour tout $n\geq 0$, les variables aléatoires $\mathbf{Y}_0^n=(Y_i,\ 0\leq i\leq n)$ sont indépendantes conditionnellement à $\mathbf{X}_0^n=(X_i,\ 0\leq i\leq n)$.
- La loi conditionnelle de Y_n à X_n est homogène dans le temps, et est donnée par la transition $\gamma_\theta(y|x)=P_\theta(Y_n=y\,|\,X_n=x)$.

Le processus $Y=(Y_n)_{n\geq 0}$ qui seul est observé, est appelé *Chaîne de Markov cachée* (CMC).

Précisons maintenant le cadre statistique des CMC que nous allons étudier.

Nous identifions le paramètre θ à la transition P_θ de la chaîne $Z=(Z_n)_{n\geq 0}$; θ est donc une matrice indexée par $(\mathbb{X}\times\mathbb{Y})^2$. L'espace d'état de la chaîne Z est noté $\mathbb{X}\times\mathbb{Y}$.

H1. Les ensembles $\mathbb{X}$ et $\mathbb{Y}$ sont supposés finis, et leurs éléments, éventuellement indexés, sont notés x et y. À l'instant $n=0$, nous avons $Y_0\equiv y_0$.

H2. L'espace des paramètres Θ est une partie compacte (pour la topologie usuelle) de l'ensemble des matrices stochastiques à entrées strictement positives indexées par $(\mathbb{X}\times\mathbb{Y})^2$. La « vraie valeur θ_0 du paramètre » appartient à l'intérieur Θ^{int}, supposé non vide, de Θ.

Pour chaque $\theta\in\Theta$, les chaînes de Markov Z et X sont apériodiques irréductibles. Les probabilités de transition P_θ admettent donc une loi invariante μ_θ, et nous avons de plus une convergence exponentielle en variation totale de $P_\theta^n(x,\ \cdot)$ vers μ_θ uniformément en x et en θ. D'après le théorème de Kolmogorov, il existe une chaîne $Z^*=(X^*,\ Y^*)$ stationnaire de transition $P_\theta(\cdot,\ \cdot)$ indexée par $\mathbb{Z}=\{0,\ \pm 1,\ \pm 2,\ \ldots\}$.

Nous définissons pour tout $k\geq 2$ et tout $y\in\mathbb{Y}^{\mathbb{Z}}$, les fonctions

$$g_k(\theta,\ y)=\log P_\theta(Y_0^*=y_0|\mathbf{Y}_{-k+1}^{*-1}=\mathbf{y}_{-k+1}^{-1}),$$

avec la notation $\mathbf{y}_l^k=(y_l,\ \ldots,\ y_k)'$. B.P. introduisent dans [1], pour la vraie valeur θ_0 du paramètre, une fonction $H(\theta)$, étroitement liée à ce que nous pouvons appeler l'*entropie*.

$$H(\theta)=\mathbb{E}_{\theta_0}(\lim_{k\to\infty} g_k(\theta,\ Y^*)),$$

où $\mathbb{E}_{\theta_0}$ est l'espérance pour la mesure invariante sur $\mathbb{Y}^{\mathbb{Z}}$.

De plus, $H(\theta)$ est 2 fois continûment différentiable (*voir* [1], corollaire 4.3), et l'information de Fisher, que nous noterons $I(\theta_0)$, est donnée par $-H^{(2)}(\theta_0)$, où $H^{(2)}$ désigne la matrice hessienne de la fonction H. Nous supposerons :

H3. $I(\theta_0)$ est inversible.

Nos résultats sont les suivants :

THÉORÈME 1. – *Sous les hypothèses précédentes, la structure statistique associée à la* CMC $(Y_n)_{n\geq 0}$ *est localement asymptotiquement normale (LAN).*

H4. (Identifiabilité) Nous supposons que $H(\theta) = H(\theta_0)$ si et seulement si $\theta = \theta_0$. Rappelons que $\theta \neq \theta_0$ implique alors la relation $H(\theta) < H(\theta_0)$.

THÉORÈME 2. – *Sous les hypothèses* H1-H4, *l'estimateur du maximum de vraisemblance* $\hat{\theta}_n$ *du paramètre est fortement convergent, asymptotiquement normalement distribué et efficace. Soit*

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\, I(\theta_0)^{-1}). \tag{1}$$

Pour démontrer ces théorèmes, nous couplons la chaîne de Markov Z à la chaîne stationnaire associée $Z^* = (X^*, Y^*)$ de même transition. Nous pourrons ainsi transférer certaines propriétés de Z^* à Z.

3. Couplage

Pour utiliser pleinement le cadre de B.P., la chaîne de Markov stationnaire Z^* est indexée par $\mathbb{Z}$. Pour la valeur θ du paramètre, Π_θ désignera la mesure induite par Z^* sur $(\mathbb{X} \times \mathbb{Y})^{\mathbb{Z}}$, invariante sous le shift.

Nous « prolongeons » alors la chaîne Z, indexée par $\mathbb{N}$ en une chaîne non-homogène indexée par $\mathbb{Z}$. Cette chaîne est égale à Z_0 pour $n \leq 0$, et à Z_n pour $n > 0$. Par abus, elle sera encore notée $Z = (Z_n)_{n\in\mathbb{Z}}$. Le mesure induite par Z sur $(\mathbb{X} \times \mathbb{Y})^{\mathbb{Z}}$ est notée Q_θ.

Nous considérons alors le modèle statistique suivant :

$$\mathcal{S} = ((\mathbb{X} \times \mathbb{Y})^2)^{\mathbb{Z}},\ ((\mathcal{X} \times \mathcal{Y})^{\otimes 2})^{\mathbb{Z}},\ (Z,\, Z^*),\ P_\theta^{\otimes 2},\ \mathbb{P}_\theta = Q_\theta \otimes \Pi_\theta,\ \theta \in \Theta).$$

Les chaînes Z et Z^* sont indépendantes. Nous utiliserons la filtration naturelle, et nous introduisons le temps d'arrêt T :

$$T(\omega) = \min\{k \in \mathbb{N} \mid Z_k(\omega) = Z_k^*(\omega)\}.$$

Sous l'hypothèse H2, on montre qu'il existe une constante $\rho\,(0 < \rho < 1)$ telle que pour tout $n \geq 0$ et tout $\theta \in \Theta$, on ait :

$$\mathbb{P}_\theta(T > n) \leq \rho^n. \tag{2}$$

Désignons par $\tilde{Z}$ la chaîne suivante :

$$\forall n \in \mathbb{N},\quad \tilde{Z}_n = \begin{cases} Z_n & \text{si } n < T, \\ Z_n^* & \text{si } n \geq T. \end{cases} \tag{3}$$

Remarque 2. – D'après la propriété de Markov forte, les chaînes Z, $\tilde{Z}$ et leurs fonctionnelles ont même loi. Grâce à l'inégalité (2), nous pourrons « récupérer » pour $\tilde{Z}$ certaines propriétés de la chaîne Z^*.

4. Démonstration des théorèmes

Étant donnée une fonction f de θ, $f^{(d)}$ désigne sa d-ième dérivée. Nous introduisons τ l'opérateur retard sur les suites, défini par $\tau(y)_n = y_{n+1}$. Pour une observation de la chaîne $\mathbf{y}_0^n$, la log-vraisemblance s'écrit :

$$\begin{aligned} L_n(\theta, y) &= \log p_\theta(y_1, \ldots, y_n) \\ &= \log \sum_{(x_1, \ldots, x_n) \in \mathbf{X}^n} \pi_\theta(x_0, x_1)\gamma_\theta(y_1|x_1) \prod_{i=1}^{n-1} \pi_\theta(x_i, x_{i+1})\gamma_\theta(y_{i+1}|x_{i+1}). \end{aligned}$$

Mais cette expression s'avère rapidement inutilisable pour montrer la propriété LAN. Nous lui préférons l'expression

$$L_n(\theta, y) = \sum_{i=1}^{n} \ell_i(\theta, y) \quad \text{où} \quad \ell_i(\theta, y) = \log P_\theta(Y_i = y_i | \mathbf{Y}_0^{i-1} = \mathbf{y}_0^{i-1}). \tag{4}$$

L'étude repose sur un développement de Taylor avec reste intégral de $L_n(\theta, y)$ et de sa dérivée $L_n^{(1)}(\theta, y)$. Grâce à [1], nous pouvons montrer que pour tous $k \geq 1$, θ, $y \in \mathbf{Y}^{\mathbf{Z}}$, nous avons :

$$|\ell_k^{(d)}(\theta, y) - g_{k+1}^{(d)}(\theta, \tau^k(y))| \leq \alpha_d(k), \tag{5}$$

où $\alpha_d(k)$, indépendant de θ et de y, est le terme général d'une série convergente. Utilisant le couplage des chaînes Z et Z^* et les résultats de [1], sections 3 et 4, on montre alors le lemme suivant :

LEMME 1. – *Sous les hypothèses* H1-H3 *pour la vraie valeur du paramètre* θ_0, *nous avons pour tous* $\theta \in \Theta$ *et* $\theta_0 \in \Theta^{\text{int}}$:

(i) $$\frac{1}{n} L_n^{(d)}(\theta, Y) \xrightarrow[n\to\infty]{} H^{(d)}(\theta), \quad P_{\theta_0}\text{-}p.s.$$

(ii) $$\frac{1}{n} L_n^{(2)}(\theta_0, Y) \xrightarrow[n\to\infty]{} -I(\theta_0), \quad P_{\theta_0}\text{-}p.s.$$

(iii) $$\frac{1}{\sqrt{n}} L_n^{(1)}(\theta_0, Y) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)).$$

Les théorèmes 1 et 2 se démontrent alors aisément.

Note remise le 23 décembre 1996, acceptée après révision le 20 mai 1997.

Références bibliographiques

[1] **Baum L. et Petrie T., 1966**. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37, p. 1554-1563.

[2] **Ball F. G. et Rice J. A., 1992**. Stochastic models for ion channels; introduction and bibliography, *Math. Biosci.*, 112, p. 189-206.

[3] **Bickel P. J. et Ritov Y., 1996**. Inference in hidden Markov models I. Local asymptotic normality in the stationary case. *Bernouilli*. 2, 3, p. 199-228.

[4] **Churchill G., 1992**. Hidden Markov chains and the analysis of genome structure. *Computers in Chemistry*.

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

[6] **LeCam L. et Yang G., 1990**. *Asymptotics in statistics. Some Basic Concepts*. Springer Verlag, New York.

[7] **Leroux B., 1992**. Maximum likelihood estimation for hidden Markov models. *Stoch. Processes and their Applications*, 40, p. 127-143.

[8] **Petrie T., 1969**. Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 40, p. 97-115.

[9] **Rydén T., 1994**. Consistent and asymptotically normal parameter estimates for hidden Markov models, *Ann. Statist.*, 22, p. 1884-1895.