

Vitesse de convergence en M-estimation de données markoviennes

Loïc Hervé, James Ledoux, Valentin Patilea

► **To cite this version:**

Loïc Hervé, James Ledoux, Valentin Patilea. Vitesse de convergence en M-estimation de données markoviennes. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494678>

HAL Id: inria-00494678

<https://hal.inria.fr/inria-00494678>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VITESSE DE CONVERGENCE EN M -ESTIMATION DE DONNÉES MARKOVIENNES

Loïc HERVÉ & James LEDOUX & Valentin PATILEA

Université Européenne de Bretagne, INSA IRMAR, UMR-CNRS 6625. Institut National des Sciences Appliquées de Rennes, 20, Avenue des Buttes de Cœsmes CS 14315, 35043 Rennes Cedex, France.

Résumé. Soit $\{X_n\}_{n \geq 0}$ une chaîne de Markov V -géométriquement ergodique, où V est une fonction non bornée à valeurs dans $[1, +\infty[$. Soit $M_n(\alpha) := n^{-1} \sum_{k=1}^n F(\alpha, X_{k-1}, X_k)$, avec $\alpha \in \mathcal{A} \subset \mathbb{R}$ et $F(\cdot, \cdot, \cdot)$ une fonction réelle. Le M -estimateur $\hat{\alpha}_n$ est défini par $M_n(\hat{\alpha}_n) \leq \min_{\alpha \in \mathcal{A}} M_n(\alpha) + c_n$, où $(c_n)_n$ est une suite décroissante de nombres réels positifs de limite nulle. Sous des hypothèses de régularité proches de celles du cas i.i.d., et sous la condition de moment

$$\left(\left| \frac{\partial F}{\partial \alpha}(\alpha, x, y) \right| + \left| \frac{\partial^2 F}{\partial \alpha^2}(\alpha, x, y) \right| \right)^{3+\varepsilon} \leq C(V(x) + V(y)),$$

avec $\varepsilon > 0$ et $C > 0$ indépendants de (x, y) , l'estimateur $\hat{\alpha}_n$ vérifie un théorème de Berry-Esseen de manière uniforme par rapport à la loi de la chaîne de Markov.

Abstract. Let $\{X_n\}_{n \geq 0}$ be a V -geometrically ergodic Markov chain with some fixed unbounded real-valued function $V \geq 1$ and consider $M_n(\alpha) := n^{-1} \sum_{k=1}^n F(\alpha, X_{k-1}, X_k)$ with $\alpha \in \mathcal{A} \subset \mathbb{R}$ and some real-valued functional $F(\cdot, \cdot, \cdot)$. Let us define the M -estimator $\hat{\alpha}_n$ by $M_n(\hat{\alpha}_n) \leq \min_{\alpha \in \mathcal{A}} M_n(\alpha) + c_n$ with $c_n, n \geq 1$ some sequence of real numbers decreasing to zero. Under some standard regularity assumptions, close to that of the i.i.d case, and under the moment assumption

$$\left(\left| \frac{\partial F}{\partial \alpha}(\alpha, x, y) \right| + \left| \frac{\partial^2 F}{\partial \alpha^2}(\alpha, x, y) \right| \right)^{3+\varepsilon} \leq C(V(x) + V(y))$$

for some constants $\varepsilon > 0$ and $C > 0$, the estimator $\hat{\alpha}_n$ satisfies a Berry-Esseen theorem uniformly with respect to the underlying probability distribution of the Markov chain.

Mots clés : Théorème de Berry-Esseen, chaîne de Markov géométriquement ergodique, M -estimateur, méthode spectrale

1. Hypothèses et énoncé du résultat.

Soit $\{X_n\}_{n \geq 0}$ une chaîne de Markov d'espace d'états (E, \mathcal{E}) et de probabilité de transition $\{Q_\theta(x, \cdot) : x \in E\}$, où θ appartient à un ensemble quelconque Θ de paramètres. La probabilité initiale notée μ peut, ou bien être fixée (par exemple égale à une distribution de Dirac), ou bien dépendre de θ . On désigne par $\mathbb{P}_{\theta, \mu}$ la loi de $\{X_n\}_{n \geq 0}$ (qui dépend de Q_θ et μ), et par $\mathbb{E}_{\theta, \mu}$ l'espérance associée.

Hypothèse (M). *Pour tout $\theta \in \Theta$, il existe une probabilité Q_θ -invariante, notée π_θ , et il existe une fonction $V : E \rightarrow [1, +\infty[$ non bornée telle que:*

$$\text{(VG1)} \quad b_1 := \sup_{\theta \in \Theta} \pi_\theta(V) < +\infty ;$$

(VG2) *Il existe $\kappa < 1$ et $C \geq 0$ tels que l'on ait, pour tous $\theta \in \Theta$, $n \geq 1$ et $x \in E$*

$$\sup \left\{ |Q_\theta^n f(x) - \pi_\theta(f)|, f : E \rightarrow \mathbb{C} \text{ mesurable, } |f| \leq V \right\} \leq C \kappa^n V(x).$$

À θ fixé, la propriété précédente exprime que $\{X_n\}_{n \geq 0}$ est V -géométriquement ergodique. Sous des conditions d'irréductibilité et d'apériodicité, cette propriété est équivalente à la condition de dérive $Q_\theta V \leq \rho V + b 1_S$, où $\rho < 1$, $b > 0$, et S est un ensemble petit, voir Meyn et Tweedie (1993) [Th. 16.0.1]. D'après Meyn et Tweedie (1994), pour obtenir l'uniformité en θ dans l'hypothèse (M), il suffit de vérifier que les constantes ρ , b , et celles intervenant pour définir S , ne dépendent pas de θ .

Décrivons maintenant notre cadre statistique. Le paramètre d'intérêt, $\alpha_0 = \alpha_0(\theta)$, est une fonction de la loi de la chaîne de Markov, et il est à valeurs dans un intervalle ouvert \mathcal{A} de \mathbb{R} . Pour estimer α_0 , nous introduisons la statistique

$$M_n(\alpha) = \frac{1}{n} \sum_{k=1}^n F(\alpha, X_{k-1}, X_k), \quad (1)$$

où $F(\cdot, \cdot, \cdot)$ est une fonction mesurable réelle définie sur $\mathcal{A} \times E^2$. Dans ce travail, le M -estimateur est défini comme une variable aléatoire $\hat{\alpha}_n$ qui dépend des observations (X_0, \dots, X_n) et vérifie la propriété suivante:

$$M_n(\hat{\alpha}_n) \leq \min_{\alpha \in \mathcal{A}} M_n(\alpha) + c_n,$$

où $(c_n)_{n \geq 1}$ est une suite de nombres réels positifs ou nuls, tendant vers 0 quand $n \rightarrow +\infty$. Supposons, d'une part que pour tout θ

$$M_\theta(\alpha) = \lim_{n \rightarrow \infty} \mathbb{E}_{\theta, \mu}[M_n(\alpha)]$$

soit défini sur \mathcal{A} et ne dépende pas de μ , d'autre part qu'il existe une unique "vraie" valeur α_0 du paramètre d'intérêt: $\forall \alpha \neq \alpha_0, M_\theta(\alpha_0) < M_\theta(\alpha)$, et enfin que pour tout $(x, y) \in E^2$, l'application $\alpha \mapsto F(\alpha, x, y)$ soit de classe \mathcal{C}^2 sur \mathcal{A} . On note $F'(\cdot, \cdot, \cdot)$ et $F''(\cdot, \cdot, \cdot)$ les dérivées partielles d'ordre 1 et 2 par rapport à α de $F(\cdot, \cdot, \cdot)$, et l'on pose

$$M'_n(\alpha) = \frac{1}{n} \sum_{k=1}^n F'(\alpha, X_{k-1}, X_k), \quad M''_n(\alpha) = \frac{1}{n} \sum_{k=1}^n F''(\alpha, X_{k-1}, X_k). \quad (2)$$

Étant donnés $m_0 \in \mathbb{N}$ et une fonction quelconque $\xi : E \times E \rightarrow \mathbb{R}$, nous ferons appel dans ce travail à la condition de moment suivante, qui est naturelle dans notre cadre :

Condition (D_{m_0}). *Il existe des constantes réelles $m > m_0$ et $C_\xi > 0$ telles que*

$$\forall \alpha \in \mathcal{A}, \forall (x, y) \in E^2, \quad |\xi(\alpha, x, y)|^m \leq C_\xi (V(x) + V(y)). \quad (3)$$

Hypothèses.

(V0) F' et F'' vérifient la condition (D_3);

(V1) $\forall \theta \in \Theta, \mathbb{E}_{\theta, \pi_\theta}[F'(\alpha_0, X_0, X_1)] = 0$ et α_0 est l'unique réel vérifiant cette condition;

(V2) $m(\theta) := \mathbb{E}_{\theta, \pi_\theta}[F''(\alpha_0, X_0, X_1)]$ est tel que $\inf_{\theta \in \Theta} m(\theta) > 0$;

(V3) $\forall n \geq 1$ il existe $r_n > 0$ tel que $\sqrt{n}r_n \rightarrow 0$ et $\sup_{\theta \in \Theta} \mathbb{P}_\theta (|M'_n(\hat{\alpha}_n)| \geq r_n) = O(n^{-1/2})$.

Les conditions (V0) (V1) assurent que $\sup_{\theta \in \Theta} m(\theta) < \infty$, que les variances asymptotiques

$$\begin{aligned} \sigma_1^2(\theta) &:= \lim_n \frac{1}{n} \mathbb{E}_{\theta, \mu} \left[\left(\sum_{k=1}^n F'(\alpha_0, X_{k-1}, X_k) \right)^2 \right] \\ \sigma_2^2(\theta) &:= \lim_n \frac{1}{n} \mathbb{E}_{\theta, \mu} \left[\left(\sum_{k=1}^n F''(\alpha_0, X_{k-1}, X_k) - n m(\theta) \right)^2 \right] \end{aligned}$$

sont bien définies, et enfin que $\sup_{\theta \in \Theta} \sigma_j(\theta) < \infty$ for $j = 1, 2$ (cf. Th. 2). Aux hypothèses (V0)-(V3), nous ajoutons les suivantes :

(V4) $\inf_{\theta \in \Theta} \sigma_j(\theta) > 0$ for $j = 1, 2$.

(V5) Il existe $\eta \in (0, 1/2)$ et $C > 0$ tels que

$$\forall (\alpha, \alpha') \in \mathcal{A}^2, \forall (x, y) \in E^2, \quad |F''(\alpha, x, y) - F''(\alpha', x, y)| \leq C |\alpha - \alpha'| (V(x) + V(y))^\eta.$$

(V6) Il existe $\gamma_n \rightarrow 0$ et une constante $d > 0$ tels que

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta \{ |\hat{\alpha}_n - \alpha_0| \geq d \} \leq \gamma_n.$$

Théorème 1 *Supposons que l'hypothèse (\mathcal{M}) soit satisfaite, que $\mu(V) < \infty$, que la fonctionnelle F vérifie la condition (D_1) , et que les conditions (V0) à (V6) soient satisfaites. Soit $\tau(\theta) := \sigma_1(\theta)/m(\theta)$. Alors il existe une constante positive C telle que l'on ait, en notant Φ la fonction de répartition de la loi normale centrée réduite :*

$$\forall n \geq 1, \quad \sup_{\theta \in \Theta} \sup_{u \in \mathbb{R}} \left| \mathbb{P}_{\theta, \mu} \left\{ \frac{\sqrt{n}}{\tau(\theta)} (\hat{\alpha}_n - \alpha_0) \leq u \right\} - \Phi(u) \right| \leq C \left(\frac{1}{\sqrt{n}} + \sqrt{n} r_n + \gamma_n \right). \quad (4)$$

Si μ dépend de θ , la condition $\mu(V) < \infty$ doit être remplacée par $\sup_{\theta \in \Theta} \mu_\theta(V) < \infty$, et la constante C de (4) est alors uniforme par rapport à μ_θ . Le Théorème 1 correspond au résultat de Michel et Pfanzagl (1971) du cas i.i.d., aux changements suivants près:

1. Les variances du cas i.i.d. (ie. $\mathbb{E}_\theta[F'(\theta, X_0)^2]$ et $\mathbb{E}_\theta[(F''(\theta, X_0) - m(\theta))^2]$ pour $\{X_n\}_{n \geq 0}$ i.i.d. et $F(\theta, x)$ donnée) sont ici remplacées par les variances asymptotiques $\sigma_1^2(\theta)$ et $\sigma_2^2(\theta)$ définies plus haut. Ceci est naturel dans notre contexte markovien.
2. La condition de moment d'ordre 3 du cas indépendant (ie. $\sup_{\theta \in \Theta} \mathbb{E}_\theta[|F'(\theta, X_0)|^3 + |F''(\theta, X_0)|^3] < \infty$) est remplacée par la condition de domination (D_3) . Ce type de condition est naturel sous l'hypothèse (\mathcal{M}) d'ergodicité V -géométrique. Sous cette même hypothèse, Dehay et Yao (2007) ont démontré un théorème limite central pour l'estimateur du maximum de vraisemblance sous une condition de domination d'ordre 2 sur F' et F'' qui correspond à l'inégalité (3) avec $m_0 = 2$.
3. Michel et Pfanzagl (1971) suppose que $M'_n(\hat{\alpha}_n) = 0$. La condition plus faible (V3) permet un traitement plus souple en pratique. Les nombres r_n de (V3) interviennent dans (4). Si $r_n = O(1/n)$ dans (V3) et $\gamma_n = O(1/\sqrt{n})$ dans (V6), on obtient alors la vitesse attendue en $O(1/\sqrt{n})$ dans (4).

À notre connaissance, le Théorème 1 est nouveau, et il complète le travail de Dehay et Yao (2007). Notre condition de moment (V0) est presque optimale (nous avons $m > 3$ dans (D_3) au lieu de l'ordre $m = 3$ du cas i.i.d.). La preuve du théorème repose sur la méthode de Pfanzagl et sur les résultats de Hervé et Pène (2008). Les détails sont présentés dans Hervé-Ledoux-Patilea (2009). Nous en rappelons les principaux éléments ci-dessous.

2. Méthode de Pfanzagl.

La méthode de Pfanzagl requiert, d'une part la propriété d'uniforme consistance (V6), d'autre part de disposer d'un théorème de Berry-Esseen pour une large classe de fonctionnelles du modèle étudié. Étant donnée une fonctionnelle centrée $\xi(\alpha_0, x, y)$, et en posant $S_n(\alpha_0) = \sum_{k=1}^n \xi(\alpha_0, X_{k-1}, X_k)$, nous devons ici disposer des résultats suivants :

- (a) *La suite $\{\mathbb{E}_{\theta, \mu}[S_n^2(\alpha_0)]/n\}_{n \geq 1}$ converge vers un nombre réel $\sigma^2(\theta)$.*
- (b) *Il existe une constante positive $B(\xi)$ telle que l'on ait pour tout $n \geq 1$*

$$\sup_{\theta \in \Theta} \sup_{u \in \mathbb{R}} \left| \mathbb{P}_{\theta, \mu} \left\{ \frac{S_n(\alpha_0)}{\sigma(\theta) \sqrt{n}} \leq u \right\} - \Phi(u) \right| \leq \frac{B(\xi)}{\sqrt{n}}.$$

Dans la méthode de Pfanzagl, les propriétés (a) et (b) sont appliquées à une famille de fonctionnelles définies comme combinaisons linéaires de F' et F'' , et à cet effet nous avons besoin d'un contrôle précis de la constante $B(\xi)$ ci-dessus en fonction de la taille de ξ . Dans le cas i.i.d., le théorème classique de Berry-Esseen fournit un tel contrôle, et l'énoncé correspondant pour les M -estimateurs hérite de cette condition de moment optimale (ie. d'ordre 3 pour F' et F''). La méthode de Pfanzagl a déjà été mise en oeuvre dans le contexte markovien, tout d'abord par Rao (1973) dans le cas des chaînes uniformément ergodiques, puis par Milhaud et Raugi (1989) pour les modèles autorégressifs linéaires. Dans ces deux articles, les conditions de moment requises pour la résolution de (b) sont assez fortes, et ces conditions se répercutent sur les fonctions F' et F'' lorsqu'on déduit le théorème de Berry-Esseen pour $\hat{\alpha}_n$. Sous l'hypothèse (\mathcal{M}) , les questions (a) et (b) peuvent être résolues grâce à la méthode spectrale développée par Hervé et Pène (2008), qui fournit en outre un contrôle de la constante $B(\xi)$. Nous présentons ci-dessous les énoncés précis que l'on peut déduire de Hervé et Pène (2008) pour les fonctionnelles de type $\xi(X_{k-1}, X_k)$.

3. Théorème de Berry-Esseen pour les fonctionnelles de (X_{k-1}, X_k) .

On suppose ci-dessous que l'hypothèse (\mathcal{M}) est satisfaite et que $\mu(V) < \infty$ (avec les précisions déjà mentionnées si μ dépend de θ). Soit $\xi(\alpha, x, y)$ une fonction réelle sur $\mathcal{A} \times E^2$ telle que la v.a. $\xi(\alpha_0, X_{k-1}, X_k)$ soit intégrable et centrée par rapport à la loi stationnaire π_θ (ie. $\mathbb{E}_{\theta, \pi_\theta}[\xi(\alpha_0, X_0, X_1)] = 0$). On pose : $S_n(\alpha) = \sum_{k=1}^n \xi(\alpha, X_{k-1}, X_k)$.

Théorème 2

1. Si ξ vérifie la condition (D_1) , alors on a $\sup_{\theta \in \Theta} \sup_{n \geq 1} |\mathbb{E}_{\theta, \mu}[S_n(\alpha_0)]| < \infty$, donc $\lim_n \mathbb{E}_{\theta, \mu}[S_n(\alpha_0)/n] = 0$ pour tout $\theta \in \Theta$.

2. Si ξ vérifie la condition (D_2) , alors pour tout $\theta \in \Theta$, le nombre réel positif

$$\sigma^2(\theta) := \lim_n \frac{\mathbb{E}_{\theta, \mu}[S_n(\alpha_0)^2]}{n}$$

est défini et ne dépend pas de μ , de plus $\sigma^2(\cdot)$ est bornée sur Θ et il existe une constante $C > 0$ dépendant uniquement de la constante C_ξ (dans (D_2)) et de $\mu(V)$ telle que

$$\forall \theta \in \Theta, \forall n \geq 1, \quad \left| \sigma^2(\theta) - \frac{\mathbb{E}_{\theta, \mu}[S_n(\alpha_0)^2]}{n} \right| \leq \frac{C}{n}.$$

3. Si ξ vérifie la condition (D_3) et si $\sigma_0^2 := \inf_{\theta \in \Theta} \sigma^2(\theta) > 0$, alors il existe une constante positive $B(\xi)$ telle que

$$\forall n \geq 1, \quad \sup_{\theta \in \Theta} \sup_{u \in \mathbb{R}} \left| \mathbb{P}_{\theta, \mu} \left\{ \frac{S_n(\alpha_0)}{\sigma(\theta)\sqrt{n}} \leq u \right\} - \Phi(u) \right| \leq \frac{B(\xi)}{\sqrt{n}}.$$

En outre la constante $B(\xi)$ dépend de la fonctionnelle ξ , mais uniquement comme fonction de σ_0 et de la constante C_ξ de la condition (D_3) .

En adaptant les techniques de transformées de Fourier du cas i.i.d., on peut voir que le théorème 2 résulte du lemme suivant qui a un intérêt propre : en effet il fournit, en relation avec la condition (3), un développement intéressant pour la fonction caractéristique de la v.a. $S_n(\alpha_0)$.

Lemma 1 *Si ξ vérifie la condition (D_{m_0}) avec $m_0 \in \mathbb{N}^*$, alors il existe $\beta > 0$ tel que*

$$\forall \theta \in \Theta, \forall n \geq 1, \forall t \in [-\beta, \beta], \quad \mathbb{E}_{\theta, \mu}[e^{itS_n(\alpha_0)}] = \lambda_\theta(t)^n (1 + L_\theta(t)) + r_{\theta, n}(t),$$

où $\lambda_\theta(\cdot)$, $L_\theta(\cdot)$ et $r_{\theta, n}(\cdot)$ sont des fonctions de classe \mathcal{C}^{m_0} de $[-\beta, \beta]$ dans \mathbb{C} satisfaisant $\lambda_\theta(0) = 1$, $\lambda'_\theta(0) = 0$, $L_\theta(0) = 0$ et $r_{\theta, n}(0) = 0$. En outre il existe $\rho \in [0, 1[$ tel que l'on ait pour $\ell = 0, \dots, m_0$:

$$G_\ell := \sup \{\rho^{-n} |r_{\theta, n}^{(\ell)}(t)|, |t| \leq \beta, \theta \in \Theta, n \geq 1\} < \infty.$$

Enfin les constantes β , ρ , G_ℓ , et les constantes suivantes (définies pour $\ell = 0, \dots, m_0$)

$$E_\ell := \sup\{|\lambda_\theta^{(\ell)}(t)|, |t| \leq \beta, \theta \in \Theta\} < \infty, \quad F_\ell := \sup\{|L_\theta^{(\ell)}(t)|, |t| \leq \beta, \theta \in \Theta\} < \infty,$$

dépendent de ξ , mais uniquement comme fonction de la constante C_ξ dans (D_{m_0}) .

Ce lemme est une conséquence des résultats de Hervé et Pène (2008).

Bibliographie

- [1] D. Dehay et J.-F. Yao (2007). On likelihood estimation for discretely observed Markov jump processes. *Aust. N. Z. J. Stat.*, 49:93–107.
- [2] L. Hervé et F. Pène (2008). Nagaev method via keller-liverani theorem. Prépublication Hal-00203408. À paraître dans *Bull. Soc. Math. France*.
- [3] L. Hervé, J. Ledoux, V. Patilea. A Berry-Esseen theorem on M -estimators for geometrically ergodic Markov chains. Preprint (2009).
- [4] S. P. Meyn et R. L. Tweedie (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Probab.*, 4:981–1011.
- [5] S.P. Meyn et R.L. Tweedie (1993). *Markov chains and stochastic stability*. Springer Verlag, New York, Heidelberg, Berlin.
- [6] R. Michel et J. Pfanzagl (1971). The accuracy of normal approximation. *Z. Wahrsch. Verw. Gebiete*, 18:73–84.
- [7] X. Milhaud et A. Raugi (1989). Étude de l'estimateur du maximum de vraisemblance dans le cas d'un processus autorégressif : convergence, normalité asymptotique, vitesse de convergence. *Ann. Inst. H. Poincaré Probab. Statist.*, 25:383–428.
- [8] B. L. S. Prakasa Rao (1973). On the rate of convergence of estimators for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 26:141–152.