Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 515–518
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique

Vitesses de convergence dans la loi forte des grands nombres et dans l'estimation de la densité pour des variables aléatoires associées

Lahcen Douge

L.S.T.A., Université Paris 6, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 20 juin 2006 ; accepté après révision le 26 février 2007

Disponible sur Internet le 6 avril 2007

Présenté par Paul Deheuvels

Résumé

Nous considérons un processus stationnaire associé $(X_i)_{i\in\mathbb{N}}$. Nous établissons une nouvelle inégalité exponentielle et nous déduisons une vitesse de convergence dans la loi forte des grands nombres. Pour ce type de processus une vitesse de convergence presque sûre uniforme sur les ensembles compacts de l'estimateur à noyau de la fonction de densité est également établie. ***Pour citer cet article : L. Douge, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Abstract

Convergence rates in the law of large numbers and for density estimation for associated random variables. Let $(X_i)_{i\in\mathbb{N}}$ be a stationary associated random process. We give a new exponential inequality and derive a rate of convergence for the law of large numbers. For this type of process, a uniform almost sure rate of convergence over compact sets for the kernel density estimator is also given. ***To cite this article: L. Douge, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

1. Introduction

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires stationnaires associées. Dans cette Note, sous une condition de décroissance géométrique de la covariance, nous démontrons une inégalité exponentielle qui conduit à une nouvelle vitesse de convergence dans la loi forte des grands nombres. En particulier, dans un cadre similaire, cette vitesse améliore celle obtenue avec l'inégalité de Roussas et Ioannides [6] de l'ordre $[(\log n)^2/n)]^{1/6}$. Ensuite, nous démontrons sous d'autres conditions de ρ régularité et de décroissance de la covariance que sur un compact I de $\mathbb{R}$

$$\sup_{x\in I}\left|f_n(x)-f(x)\right| = \mathcal{O}\left(\frac{\log^2 n}{n}\right)^{\rho/(2\rho+1)} \qquad \text{p.s.}$$

Une légère amélioration est alors observée par rapport aux résultats de Masry [5] d'ordre $n^{-\delta}$, $0<\delta$, et de Doukhan et Louhichi [3] d'ordre $(\log n)^{-2\rho/(2\rho+1)}$.

Adresse e-mail : douge@ccr.jussieu.fr.

1631-073X/$ – see front matter © 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.
doi:10.1016/j.crma.2007.02.017

Définition 1. Une suite $(X_n)_{n\in\mathbb{N}}$ est dite positivement associée, ou tout simplement associée, si pour tout sous-ensemble fini I de $\mathbb{N}$ et pour toutes fonctions réelles croissantes H_1 et H_2 définies sur $\mathbb{R}^I$,

$$\mathrm{Cov}\big(H_1(X_i, i\in I), H_2(X_i, i\in I)\big) \geqslant 0,$$

lorsque cette covariance existe.

La notion de l'association a été introduite par Esary, Proschan et Walkup [4], dont l'objectif était de trouver des applications en fiabilité et en statistique.

2. Inégalité exponentielle

Théorème 2.1. *Soit $(X_i)_{i\in\mathbb{N}}$ une suite stationnaire de variables aléatoires (v.a.) associées vérifiant les conditions suivantes*

(i) $|X_i| \leqslant M$, $\forall\, i \geqslant 0$, *où M est une constante* ;
(ii) $C(k) := \mathrm{Cov}(X_0, X_k) \leqslant \theta_0 \exp(-\theta k)$, *pour tout* $k \geqslant 0$, $\theta > 0$ *et* $\theta_0 > 0$.

Alors, pour tout $n \geqslant 2$ et tout $\varepsilon > \frac{6M}{\sqrt{n}}$, on obtient

$$P\left(\frac{1}{n}\left|\sum_{i=1}^{n}(X_i - \mathrm{E}X_i)\right| \geqslant \varepsilon\right) \leqslant 8C_0 \exp\left(-\frac{\theta\wedge 1}{12M}\sqrt{n}\,\varepsilon\right), \tag{1}$$

où $C_0 = \exp(\theta_0/(4M^2(1-\mathrm{e}^{-\theta})))$.

Démonstration. (Indications) Pour $n \geqslant 2$, on pose $p = p(n)$ et $r = r(n)$ tels que $2pr \leqslant n$. Soit $Y_i = X_i - \mathrm{E}X_i$, $i = 1, \ldots, n$, et $\overline{S}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Définissons les v.a. U_i, V_i, $i = 1, \ldots, r$, et W_n par

$$U_i = Y_{2(i-1)p+1} + \cdots + Y_{(2i-1)p}, \quad V_i = Y_{(2i-1)p+1} + \cdots + Y_{2ip}, \quad W_n = Y_{2pr+1} + \cdots + Y_n,$$

et $\overline{U}_n = \frac{1}{n}\sum_{i=1}^{r} U_i$, $\overline{V}_n = \frac{1}{n}\sum_{i=1}^{r} V_i$, $\overline{W}_n = \frac{W_n}{n}$. Pour tout $\varepsilon > 0$, nous constatons bien que

$$P(\overline{U}_n \geqslant \varepsilon) \leqslant \mathrm{e}^{-\lambda\varepsilon}\mathrm{E}\big(\mathrm{e}^{\lambda\overline{U}_n}\big) \leqslant \mathrm{e}^{-\lambda\varepsilon}\left[\left|\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}\sum_{i=1}^{r} U_i}\big) - \prod_{i=1}^{r}\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}U_i}\big)\right| + \prod_{i=1}^{r}\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}U_i}\big)\right]. \tag{2}$$

En appliquant l'inégalité $\mathrm{e}^x \leqslant 1 + x + x^2$ ($|x| \leqslant \frac{1}{2}$), pour $x = \frac{\lambda}{n}U_n$, où $\lambda = \frac{\theta\wedge 1}{4M}\frac{n}{p}$, nous obtenons

$$\prod_{i=1}^{r}\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}U_i}\big) \leqslant \prod_{i=1}^{r}\mathrm{E}\left(1 + \frac{\lambda}{n}U_i + \frac{\lambda^2}{n^2}U_i^2\right) \leqslant \exp\left(\frac{\theta_0}{1-\mathrm{e}^{-\theta}}\frac{\lambda^2}{n}\right). \tag{3}$$

D'autre part comme $U_1, \ldots, U_r$ sont des v.a. associées (propriété 4 de [4]), d'après le Lemme 3.1 de Dewan et Prakasa Rao [1], il ressort que

$$\left|\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}\sum_{i=1}^{r} U_i}\big) - \prod_{i=1}^{r}\mathrm{E}\big(\mathrm{e}^{\frac{\lambda}{n}U_i}\big)\right| \leqslant \frac{\lambda^2\theta_0}{2n(1-\mathrm{e}^{-\theta})}\exp(\lambda M - \theta p). \tag{4}$$

Les inégalités (3) et (4) sont vérifiées si on remplace U_i par $-U_i$. Pour un choix de $p = [\sqrt{n}]$, et d'après (2), il découle que

$$P\big(|\overline{U}_n| \geqslant \varepsilon\big) = P(\overline{U}_n \geqslant \varepsilon) + P(-\overline{U}_n \geqslant \varepsilon) \leqslant 4C_0\exp\left(-\frac{\theta\wedge 1}{4M}\sqrt{n}\,\varepsilon\right), \tag{5}$$

où $C_0 = \exp(\theta_0/(4M^2(1-\mathrm{e}^{-\theta})))$. Il est clair que $\overline{V}_n$ vérifie l'inégalité (5) et que $P(|\overline{W}_n| \geqslant \varepsilon) = 0$ pour tout $\varepsilon > 2M/\sqrt{n}$. Du fait que $\overline{S}_n = \overline{U}_n + \overline{V}_n + \overline{W}_n$, achève la démonstration. □

Remarque. D'après l'inégalité (1), on déduit que $\frac{1}{n}|\sum_{i=1}^{n}(X_i - \mathrm{E}X_i)| = \mathcal{O}(\frac{\log n}{\sqrt{n}})$ p.s.

Exemple *(Processus linéaire généré par des variables aléatoires associées).* Soit $(X_i, i \in \mathbb{N})$ un processus linéaire défini par : $X_j = \sum_{i=0}^{\infty} a_i \varepsilon_{j-i}$ où a_i est une suite de nombres réels telle que $|a_i| \leqslant r_0 r^i$, où $r_0 > 0$, $0 < r < 1$, et, $(\varepsilon_i, i \in \mathbb{Z})$ est une suite stationnaire de v.a. associées bornées telle que $\mathrm{E}(\varepsilon_i) = 0$ et $\mathrm{Cov}(\varepsilon_0, \varepsilon_n) \leqslant \rho_0 \rho^{|n|}$ où $\rho_0 > 0$ et $0 < \rho < 1$. Si $r < \rho$, alors $(X_i, i \in \mathbb{N})$ vérifie l'inégalité exponentielle (1).

3. Convergence uniforme de l'estimateur à noyau de la densité et vitesse de convergence

Soit $(X_i)_{i\in\mathbb{N}}$ une suite strictement stationnaire de v.a. associées telle que X_0 et (X_0, X_j) aient des densités, notées respectivement f et $f_{0,j}$. On considère l'estimateur à noyau de f, $f_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})$, où K est un noyau d'intégrale 1 et h_n une suite de nombres réels positifs telle que $h_n \to 0$ et $nh_n \to \infty$. Les hypothèses faites sur la suite de v.a. $(X_i)_{i\in\mathbb{N}}$ sont les suivantes :

Hypothèses.

(H1) $\sup_{j\geqslant 1} \| f_{0,j} - f \otimes f \|_\infty \leqslant C$ et $\|f\|_\infty \leqslant C$ où C est une constante ;
(H2) Le noyau K est borné, dérivable et sa dérivée K' est bornée ;
(H3) (i) $\sum_{k=0}^{n-1} (k+1)^l\, \theta_k^{1/4} = \mathcal{O}(n^{\gamma l/2})$, $\forall l \geqslant 0$, $0 < \gamma < 1$,
(ii) $\sum_{k=0}^{\infty} \theta_k^{1/4} < \infty$ où $\theta_k = \sup_{\{|i-j|\geqslant k\}} \mathrm{Cov}(X_i, X_j)$, $\forall k \geqslant 1$ et $\theta_0 = 1$.

Théorème 3.1. *Sous* (H1)–(H3)*, si $h_n^{-1} = \mathcal{O}(n^{1-\gamma})$, alors il existe $D > 0$ tel que pour tout $\lambda > 0$ et pour tout $x \in \mathbb{R}$*

$$P\big(\big|f_n(x) - \mathbb{E}f_n(x)\big| > \lambda\big) \leqslant \sqrt{2}\,\mathrm{e}^2 \exp\big(-D\lambda\sqrt{nh_n}\big). \tag{6}$$

Démonstration. (Indications) La preuve du théorème est similaire dans ses grandes lignes à celle de Doukhan et Louhichi [3] pour une classe de v.a. faiblement dépendantes qui inclut les variables associées. Notons qu'ici on remplace l'hypothèse de décroissance géométrique de la covariance dans [3] par (H3) ce qui nous permet d'avoir un facteur λ dans l'inégalité exponentielle au lieu de $\sqrt{\lambda}$.

Soient $S_n = \sum_{i=1}^{n} K(\frac{x-X_i}{h_n}) - \mathbb{E}K(\frac{x-X_i}{h_n})$ et $g(y) = K(\frac{x-y}{h_n}) - \mathbb{E}K(\frac{x-y}{h_n})$, on voit que

$$\big|\mathbb{E}S_n^q\big| \leqslant q! \sum_{1\leqslant t_1\leqslant\cdots\leqslant t_q\leqslant n} \big|\mathbb{E}g(X_{t_1})\cdots\mathbb{E}g(X_{t_q})\big| =: q!A_q(g).$$

Pour majorer $\mathbb{E}S_n^q$ il suffit donc de majorer $A_q(g)$. D'après les Lemmes 14 et 15 de [2], on a

$$A_q(g) \leqslant \sum_{m=1}^{q-1} A_m(g)A_{q-m}(g) + n\sum_{r=0}^{n-1}(r+1)^{q-2}\, C_{r,q}(g), \tag{7}$$

où

$$C_{r,q}(g) = \sup_{1\leqslant t_1\leqslant\cdots\leqslant t_q\leqslant n} \big|\mathrm{Cov}\big(g(X_{t_1})\cdots g(X_{t_m}), g(X_{t_{m+1}})\cdots g(X_{t_q})\big)\big|$$

et $r = t_{m+1} - t_m = \max_{1\leqslant i\leqslant q-1}(t_{i+1} - ti)$.

Pour tout $r \geqslant 1$, on distingue ensuite les étapes suivantes :

- Sous les hypothèses (H1) et (H2) : $|\mathrm{Cov}(g(X_{t_1})\cdots g(X_{t_m}), g(X_{t_{m+1}})\cdots g(X_{t_q}))| \leqslant 2^{q-2}\|K\|_\infty^{q-2} C h_n^2$;
- Par association : $|\mathrm{Cov}(g(X_{t_1})\cdots g(X_{t_m}), g(X_{t_{m+1}})\cdots g(X_{t_q}))| \leqslant 2^{q-2}\|K\|_\infty^{q-2}\|K'\|_\infty^2 \theta_r / h_n^2$;
- $C_{0,q}(g) = |\mathbb{E}g^q(X_1)| \leqslant 2^q \|K\|_\infty^{q-1}\, \|f\|_\infty h_n$.

Les trois bornes précédentes entraînent

$$C_{r,q}(g) \leqslant C_1^q \theta_r^{1/4} h_n \quad \forall r \geqslant 0, \tag{8}$$

où $C_1 = 2\max\{\|K\|_\infty, \|K'\|_\infty, C, 1\}$.

D'après (7) et (8) et sous (H3) avec la condition $h_n^{-1} = \mathcal{O}(n^{1-\gamma})$, il existe une constante $A > 0$ telle que

$$A_q(g) \leqslant \sum_{m=1}^{q-1} A_m(g)A_{q-m}(g) + A^q(nh_n)^{q/2} \quad \forall q \geqslant 2.$$

Un raisonnement par récurrence sur q, permet de conclure que

$$\left|\mathbb{E}S_n^q\right| \leqslant \frac{2q!}{q!}A^q(nh_n)^{q/2} \qquad \forall q \geqslant 0.$$

En optimisant cette inégalité par rapport à q et en utilisant l'inégalité de Markov nous obtenons (6). $\square$

Théorème 3.2. *Supposons que la fonction de densité f soit lipschitzienne d'ordre ρ et que les hypothèses* (H1)–(H3) *soient vérifiées pour $\gamma < 1 - \frac{1}{2\rho+1}$. Si de plus,* $\lim_{n\to\infty} \frac{nh_n}{\log^2 n} = \infty$ *et* $h_n = (\frac{\log^2 n}{n})^{1/(2\rho+1)}$ *et si* $\int |u|^\rho K(u)\,\mathrm{d}u < \infty$, *alors pour tout compact I de* $\mathbb{R}$ *on a*

$$\sup_{x\in I}\left|f_n(x) - f(x)\right| = \mathcal{O}\left(\frac{\log^2 n}{n}\right)^{\rho/(2\rho+1)} \qquad p.s.$$

Références

[1] I. Dewan, B.L.S. Prakasa Rao, A general method of density estimation for associated random variables, J. Nonparametric Statist. 10 (1999) 405–420.

[2] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, Stochastic Process. Appl. 84 (1999) 313–342.

[3] P. Doukhan, S. Louhichi, Functional estimation of a density under a new weak dependence condition, Scand. J. Statist. 28 (2001) 325–341.

[4] J.D. Esary, F. Proschan, D. Walkup, Association of random variables, with applications, Ann. Math. Statist. 38 (1967) 1466–1474.

[5] E. Masry, Multivariate probability density estimation for associated processes: Strong consistency and rates, Statist. Probab. Lett. 58 (2001) 205–219.

[6] G.G. Roussas, D.A. Ioannides, Exponential inequality for associated random variables, Statist. Probab. Lett. 42 (1999) 423–431.