

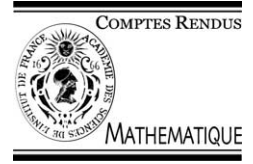


ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Acad. Sci. Paris, Ser. I 337 (2003) 473–476



Probabilités/Statistique

Algorithmes stochastiques à bruit dépendant

Paul Doukhan^a, Odile Brandière^b

^a LS CREST, timbre J340, 3, avenue Pierre Larousse, 92240 Malakoff, France

^b Université Paris 11, laboratoire de mathématiques, bât. 425, 91405 Orsay, France

Reçu le 20 mars 2003 ; accepté le 16 juillet 2003

Présenté par Paul Deheuvels

Résumé

La dépendance du bruit d'un algorithme stochastique est modélisée de différentes manières, de sorte que la méthode de l'équation différentielle ordinaire reste applicable. Ces techniques de dépendance faible sont illustrées ici par des applications à un algorithme de régression linéaire et à l'étude de tableaux triangulaires de variables aléatoires pondérées dépendantes. L'objectif est ici d'obtenir des conditions aisément vérifiables en pratique. *Pour citer cet article : P. Doukhan, O. Brandière, C. R. Acad. Sci. Paris, Ser. I 337 (2003).*

© 2003 Académie des sciences. Publié par Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

Abstract

Dependent noise for stochastic algorithms. We introduce different ways of modeling the dependency of the input noise of stochastic algorithms. We are aimed to prove that such innovations allow us to use the ODE (ordinary differential equation) method. Illustrations in the linear regression framework and in the law of the large number for triangular arrays of weighted dependent random variables are also given. We have aimed to provide results easy to check in practice. *To cite this article: P. Doukhan, O. Brandière, C. R. Acad. Sci. Paris, Ser. I 337 (2003).*

© 2003 Académie des sciences. Publié par Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

1. Introduction

Nous considérons les algorithmes stochastiques à valeurs dans \mathbb{R}^d , définis sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ par l'équation de récurrence ($n \geq 0$) :

$$Z_{n+1} = Z_n + c_n h(Z_n) + c_n (\varepsilon_{n+1} + r_{n+1}), \quad (1)$$

où la fonction h est continue d'un ouvert $G \subseteq \mathbb{R}^d$ dans \mathbb{R}^d , et (c_n) est une suite déterministe réelle décroissant vers zéro et telle que $\sum_n c_n = \infty$ et $\sum_n c_n^2 < \infty$; (ε_n) et (r_n) sont des vecteurs aléatoires de dimension d . La méthode de l'équation différentielle ordinaire (EDO), décrite par [9] et [8], permet d'étudier la convergence ou les valeurs d'adhérence [1] de (Z_n) si, *p.s.* (presque sûrement), les trois conditions suivantes sont réalisées : la suite (r_n) tend

Adresses e-mail : doukhan@ensae.fr (P. Doukhan), odile.brandiere@math.u-psud.fr (O. Brandière).

vers 0, la série $\sum_n c_n \varepsilon_{n+1}$ converge. La théorie classique des algorithmes porte sur un bruit (ε_n) dont les termes sont des accroissements de martingale. Nous remplaçons ici cette condition relative au bruit par une condition de dépendance faible.

La Section 2 donne plusieurs majorations relatives à des sommes pondérées de variables faiblement dépendantes qui conduisent notamment à la convergence *p.s.* de $\sum_n c_n \varepsilon_{n+1}$, utile dans l'étude des algorithmes.

La Section 3 montre sur un exemple comment appliquer la méthode de l'EDO avec un bruit faiblement dépendant.

2. Les résultats

Posons d'abord, comme Doukhan et Louhichi, [7],

$$\text{Lip}(h) = \sup_{(x_1, \dots, x_u) \neq (y_1, \dots, y_u)} \frac{|h(x_1, \dots, x_u) - h(y_1, \dots, y_u)|}{|x_1 - y_1| + \dots + |x_u - y_u|}, \quad \text{avec } h: \mathbb{R}^u \rightarrow \mathbb{R}, \quad \text{et}$$

$$\mathcal{L} = \{h, \text{ telle que pour } u \geq 0, h: \mathbb{R}^u \rightarrow \mathbb{R}, \text{ et } \text{Lip}(h) + \|h\|_\infty < \infty\}.$$

Considérons une fonction $C: \mathbb{N}^{*2} \rightarrow \mathbb{R}^+$: la suite (ε_n) est (θ, \mathcal{L}, C) -faiblement dépendante si la suite $\theta = (\theta_r)_{r \geq 0}$ décroît vers 0 à l'infini et vérifie

$$|\text{Cov}(h(\varepsilon_{t_1}, \dots, \varepsilon_{t_u}), k(\varepsilon_{t_{u+1}}, \dots, \varepsilon_{t_{u+v}}))| \leq C(u, v)(\text{Lip}(h) + \text{Lip}(k))\theta_r, \quad (2)$$

pour tous $(u+v)$ -uplets, (t_1, \dots, t_{u+v}) avec $t_1 \leq \dots \leq t_u < t_{u+r} \leq t_{u+1} \leq \dots \leq t_{u+v}$, Doukhan, [6], donne des exemples variés de telles situations.

La γ -faible dépendance définie par Dedecker et Doukhan dans [4] est aussi considérée, soit

$$\gamma_r = \sup_{k \geq 0} \|\mathbb{E}(\varepsilon_{k+r} | \sigma(\varepsilon_i, i \leq k)) - \mathbb{E}(\varepsilon_{k+r})\|_1,$$

alors, la suite (ε_n) est γ -faiblement dépendante si $\gamma_r \rightarrow 0$ à l'infini. Dans [6] on montre que la version causale de (θ, \mathcal{L}, C) -faible dépendance implique la γ -faible dépendance quand le membre de gauche de l'Éq. (2) est $\leq C(v) \text{Lip}(k)\theta_r$. Toutefois [4] exhibe un exemple de suite γ -faiblement dépendante et non (θ, \mathcal{L}, C) -faiblement dépendantes. Soit (ε_n) une suite centrée (θ, \mathcal{L}, C) -faiblement dépendante. Posant $C_q = \max_{u+v \leq q} C(u, v)$, nous supposons :

$$\text{sup} |\text{Cov}(\varepsilon_{t_1} \dots \varepsilon_{t_m}, \varepsilon_{t_{m+1}} \dots \varepsilon_{t_q})| \leq C_q q^\gamma M^{q-2} \theta_r, \quad (3)$$

où le sup est pris sur tous les $t_q \geq \dots \geq t_1 \geq 1$ tels que $t_{m+1} - t_m = r$ pour un $1 \leq m < q$. Désignons Q_X la fonction quantile de $|X|$ (inverse généralisée de la fonction $t \mapsto \mathbb{P}(|X| > t)$) et $M_q = \max(C_q, 2)$, une autre condition s'écrit

$$|\text{Cov}(\varepsilon_{t_1} \dots \varepsilon_{t_m}, \varepsilon_{t_{m+1}} \dots \varepsilon_{t_q})| \leq M_q \int_0^{\min(\theta_r, 1)} Q_{\varepsilon_{t_1}}(x) \dots Q_{\varepsilon_{t_q}}(x) dx. \quad (4)$$

La borne (3) s'applique plutôt aux fonctions bornées par M alors que (4) est réalisée pour des v.a.r. plus générales mais vérifiant des conditions de moments ou de queue. On trouvera beaucoup d'exemples dans [7]. Notons $\Sigma_n = \sum_{i=1}^n c_{i-1} \varepsilon_i$, de manière analogue à Doukhan et Louhichi, [7], nous obtenons

Proposition 2.1. Soient un entier $p \geq 2$ fixé et (ε_n) une suite de variables aléatoires réelles, centrées, (θ, \mathcal{L}, C) -faiblement dépendantes et vérifiant (3). Alors pour tout $n \geq 2$ on a,

$$|\mathbb{E} \Sigma_n^p| \leq \frac{(2p-2)!}{(p-1)!} \left\{ \left(C_p p^\gamma M^{p-2} \sum_{i=1}^n c_{i-1}^p \sum_{r=0}^{n-1} (r+1)^{p-2} \theta_r \right) \vee \left(C_2 2^\gamma \sum_{i=1}^n c_{i-1}^2 \sum_{r=0}^{n-1} \theta_r \right)^{p/2} \right\}. \quad (5)$$

Proposition 2.2. Soient $p \geq 2$, un entier fixé, et (ε_n) , une suite de variables aléatoires réelles, centrées (θ, \mathcal{L}, C) -faiblement dépendante. Supposons que pour tout $2 < q \leq p$, (ε_n) vérifie (4),

$$M_q \leq M_p^{(q-2)/(p-2)} M_2^{(p-q)/(p-2)}, \quad \text{et}, \tag{6}$$

$$\text{il existe une constante } c > 0 \text{ et } k > p \text{ tels que } \forall i \geq 0 : \mathbb{P}(|\varepsilon_i| > t) \leq \frac{c}{t^k}. \tag{7}$$

Alors pour tout $n \geq 2$,

$$|\mathbb{E} \Sigma_n^p| \leq \frac{(2p-2)!}{(p-1)!} \cdot c^{1/k} \left\{ \left(M_p \sum_{i=1}^n c_{i-1}^p \sum_{r=0}^{n-1} (r+1)^{p-2} \theta_r^{(k-p)/k} \right) \vee \left(M_2 \sum_{i=1}^n c_{i-1}^2 \sum_{r=0}^{n-1} \theta_r^{(k-2)/k} \right)^{p/2} \right\}. \tag{8}$$

Le résultat (5) ou (8) conduit à la convergence *p.s.* de (Σ_n) . Adaptant les démonstrations de Dedecker et Doukhan dans [4], on obtient

Proposition 2.3. Soit $p > 2$ et $(\varepsilon_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles intégrables de coefficients de γ -dépendance $(\gamma_r)_{r \geq 0}$ et vérifiant (7). Alors pour tout $n \geq 2$,

$$|\mathbb{E} \Sigma_n^p| \leq \left(2p K_1 \sum_{i=1}^n c_i^{2-2(k-p)/(p(k-1))} \sum_{j=0}^{n-1} \gamma_j^{2(k-p)/(p(k-1))} \right)^{p/2}, \tag{9}$$

où K_1 dépend de r, p et c .

Remarquons qu'ici $p \in \mathbb{R}$ n'est pas nécessairement entier. D'autre part, s'il existe p tel que (9) soit satisfaite et que $\sum_{i=1}^\infty c_i^{2-m} \sum_{j=0}^\infty \gamma_j^m < \infty$ avec $m = \frac{2(k-p)}{p(k-1)} < 1$, alors (Σ_n) converge *p.s.* Ces résultats s'étendent dans \mathbb{R}^d .

3. Applications

Les résultats précédents s'appliquent directement aux algorithmes de Robbins–Monro et de Kiefer–Wolfowitz.

Nous étendons les résultats de convergence complète que Chow, [3], prouve pour des variables indépendantes à des tableaux triangulaires de variables pondérées et faiblement dépendantes, $(\sum_{i=1}^n c_{n,i} \varepsilon_i)_{n \geq 0}$. Dans le cadre de la régression linéaire, observant une suite stationnaire bornée $(y_n, x_n) \in \mathbb{R} \times \mathbb{R}^d$, nous cherchons un vecteur Z^* tel que $Z^* = \arg \min_{Z \in \mathbb{R}^d} \mathbb{E}[(y_n - x_n^T Z)^2]$. Notons $U = \mathbb{E}(y_{n+1} x_{n+1})$ et $C = \mathbb{E}(x_{n+1} x_{n+1}^T)$. Le problème précédent nous conduit à étudier l'algorithme du gradient

$$Z_{n+1} = Z_n + c_n (y_{n+1} - x_{n+1}^T Z_n) x_{n+1} = Z_n + c_n h(Z_n) + c_n \eta_{n+1}, \tag{10}$$

où $c_n = g/n$, $h(Z) = U - CZ$ et $\eta_{n+1} = (y_{n+1} x_{n+1} - U) + (C - x_{n+1} x_{n+1}^T) Z_n$. Remarquons que l'EDO associée à (10) admet des trajectoires solutions qui convergent toutes vers l'unique zéro attractif de h , $C^{-1}U$. Introduisons les hypothèses H-lr :

C est inversible et pour tous $i, j \in \{1, \dots, d\}$, $(x_n^i x_n^j)$ et $(y_n x_n)$ sont γ -faiblement dépendantes, avec un coefficient de dépendance vérifiant $\gamma_r = \mathcal{O}(a^r)$ avec $a < 1$.

En posant $M = \sup_n \|x_n\|^2$, nous obtenons

Proposition 3.1. Sous H-lr :

(Z_n) est *p.s.* borné, la perturbation (η_n) se décompose en trois termes dont l'un est un reste qui tend vers 0 à l'infini et les deux autres sont γ -faiblement dépendants.

Ainsi la méthode de l'EDO assure la convergence p.s. de Z_n vers $C^{-1}U$. De plus, si $g < \frac{1}{2M}$, p.s. $\sqrt{n'}(Z_n - C^{-1}U) = \mathcal{O}(1)$.

Schéma de démonstration. On utilise la décomposition

$$\eta_{n+1} = (y_{n+1}x_{n+1} - U) + [(C - x_{n+1}x_{n+1}^T)Z_n - \mathbb{E}(C - x_{n+1}x_{n+1}^T)Z_n] + \mathbb{E}(C - x_{n+1}x_{n+1}^T)Z_n.$$

Le premier terme est γ -faiblement dépendant étant donnée H-lr. On montre que si (Z_n) est bornée, le dernier terme est un reste qui tend vers 0 et le deuxième γ -faiblement dépendant. Pour montrer que Z_n est bornée, on utilise la fonction de Lyapounov $V(Z) = Z^T C Z$ et on prouve que soit le théorème de Robbins–Sigmund, soit le théorème de Delyon, [5], s'applique.

Remarque. Signalons pour finir, que Chen [2] obtient un énoncé analogue en omettant l'utilisation de l'hypothèse (ii) grâce à l'utilisation d'une autre fonction de Lyapounov.

Références

- [1] M. Benaïm, A dynamical system approach to stochastic approximation, *SIAM J. Control Optim.* 34 (2) (1996) 437–472.
- [2] H.-F. Chen, *Stochastic Approximation and Its Applications*, Kluwer Academic, 2002.
- [3] Y.S. Chow, Some convergence theorems for independent random variables, *Ann. Math. Statist.* 37 (1966) 1482–1493; 36 (4) 1293–1314.
- [4] P. Dedecker, P. Doukhan, A new covariance inequality and applications, *Stochastic Process. Appl.* (2002), in press.
- [5] B. Delyon, General convergence result on stochastic approximation, *IEEE Trans. Automatic Control* 41 (9) (1996).
- [6] P. Doukhan, Models inequalities and limit theorems for stationary sequences, in: Doukhan, et al. (Eds.), *Theory and Applications of Long Range Dependence*, Birkhäuser, 2002, pp. 43–101.
- [7] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, *Stochastic Process. Appl.* 84 (1999) 313–342.
- [8] M. Duflo, *Algorithmes Stochastiques*, in: *Collect. Math. Appl.*, Vol. 23, Springer, 1996.
- [9] H.J. Kushner, D.S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, in: *Appl. Math. Sci.*, Vol. 26, Springer, 1978.