



Probabilités

Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques

About the rate function in Talagrand's inequality for empirical processes

Emmanuel Rio

UMR 8100 CNRS, laboratoire de mathématiques de Versailles, bâtiment Fermat, 45, avenue des États Unis, 78035 Versailles cedex, France

INFO ARTICLE

Historique de l'article :

Reçu le 17 janvier 2012

Accepté après révision le 14 février 2012

Disponible sur Internet le 3 mars 2012

Présenté par le Comité de rédaction

RÉSUMÉ

Dans cette Note, nous donnons une nouvelle version de l'inégalité de Talagrand pour les processus empiriques avec une meilleure fonction de taux de grandes déviations.

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

ABSTRACT

In this Note we obtain a new Talagrand type concentration inequality for the maximal deviation of empirical processes associated to independent observations, with an improved rate function.

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction et résultat principal

Soit X_1, X_2, \dots une suite de variables aléatoires indépendantes, à valeurs dans \mathcal{X} espace polonais muni de sa tribu borélienne, et $P_n = n^{-1}(\delta_{x_1} + \dots + \delta_{x_n})$ la mesure empirique associée. Soit \mathcal{G} une famille dénombrable de fonctions mesurables de \mathcal{X} dans \mathbb{R} , de carré intégrable sous P .

Dans cette Note nous regardons les propriétés de concentration de la variable aléatoire $Z_n = \sup\{nP_n(g) : g \in \mathcal{G}\}$ autour de sa moyenne. Pour des classes de fonctions uniformément bornées, Talagrand [7] a obtenu une inégalité de type Bennett, pour la déviation de Z_n autour de sa moyenne. Ledoux [4] a donné une méthode fondée sur l'entropie permettant de retrouver l'inégalité de Talagrand. Rio [5,6], Klein [3] et Bousquet [2] ont obtenu les constantes suivantes, presque optimales, dans les inégalités de Talagrand.

Soit $E = \mathbb{E}(Z_n)/n$ (noter que E dépend de n) et $\sigma^2 = n \sup\{\text{Var} P_n(g) : g \in \mathcal{G}\}$. Quand \mathcal{G} est une classe de fonctions à valeurs dans $[0, 1]$, si $E \leq 1/2$, alors Rio [5] a montré que, pour t positif,

$$n^{-1} \log \mathbb{E}(\exp(tZ_n)) \leq tE + E(1-E)^{-1}(e^{t(1-E)} - t(1-E) - 1) \quad (1)$$

et a donné une borne de type Hoeffding (fonction de taux binomiale) pour t négatif. L'inégalité (1) montre que $\text{Var} Z_n \leq nE(1-E)$: cette borne sur la variance ne peut pas être améliorée. Quand les variables X_i sont de même loi P , les fonctions g centrées sous P et majorées par 1, Rio [6] a montré que, pour t positif,

$$n^{-1} \log \mathbb{E}(\exp(tZ_n)) \leq tE + (\sigma^2 + 2E)b^{-2}(e^{bt} - bt - 1), \quad (2)$$

Adresse e-mail : Emmanuel.Rio@math.uvsq.fr.

avec $b = 3/2$. L'inégalité (2) montre en particulier que $\text{Var } Z_n \leq n(\sigma^2 + 2E)$. Plus tard Bousquet [2] a obtenu (2) avec $b = 1$. Enfin Klein [3] a été le premier à obtenir des bornes similaires à (2) pour t négatif. L'inégalité (2) n'est cependant pas satisfaisante par rapport à (1) : d'une part, comme $(Z_n - E(Z_n))/n$ est bornée par $1 - E$ il semblerait plus conforme d'obtenir (2) avec $b = 1 - E$, et d'autre part, dans [2], en adaptant un argument déjà présent dans [5], il est montré que, sous les hypothèses de (2), $\text{Var } Z_n \leq n(\sigma^2 + 2E - E^2)$, et donc le facteur variance dans (2) n'est pas optimal.

Le résultat principal de cette Note, ci-dessous, comble ces lacunes :

Théorème 1. Soit X_1, X_2, \dots une suite de variables indépendantes de loi P sur \mathcal{X} , et \mathcal{G} une classe dénombrable de fonction mesurables de \mathcal{X} dans $]-\infty, 1]$, centrées sous P , de variance finie sous P et majorée par σ^2 . Soient $Z_n = \sup\{g(X_1) + g(X_2) + \dots + g(X_n) : g \in \mathcal{G}\}$ et $E = \mathbb{E}(Z_n)/n$. Alors, pour tout t positif,

$$n^{-1} \log \mathbb{E}(\exp(tZ_n)) \leq tE + (\sigma^2 + 2E - E^2)(1 - E)^{-2}(e^{(1-E)t} - (1 - E)t - 1). \quad (3)$$

Par conséquent, si $V = \sigma^2 + 2E - E^2$ et $h(u) = (1 + u) \log(1 + u) - u$, alors, pour tout x positif,

$$\mathbb{P}(Z_n - \mathbb{E}(Z_n) \geq nx) \leq \exp(-nV(1 - E)^{-2}h(x(1 - E)/V)).$$

Remarque. La seconde partie du théorème découle de (3) par le calcul de Cramér–Chernov usuel. La borne obtenue améliore les bornes antérieures.

2. Preuve du résultat principal

Pour montrer le théorème 1, on suppose d'abord que \mathcal{G} est finie et on passe ensuite à la limite par convergence monotone pour obtenir le résultat pour \mathcal{G} dénombrable. Soit $L_n(t) = \log \mathbb{E}(\exp(tZ_n))$. Par convention, on pose $L_0(t) = 0$. Le point essentiel de la preuve est de montrer que

$$(n - 1)L_n(t) \leq nL_{n-1}(t) \quad \text{pour tout } t > 0 \text{ et tout entier } n > 0. \quad (4)$$

Cette inégalité découle immédiatement du lemme ci-dessous, nouveau à ma connaissance :

Lemme 1. Soit Y_1, Y_2, \dots une suite de vecteurs aléatoires indépendants et équidistribués, à valeurs dans un espace vectoriel E , et $S_n = Y_1 + Y_2 + \dots + Y_n$. Soit φ une fonction convexe et mesurable de E dans \mathbb{R} , positivement homogène de degré 1. Supposons que $\mathbb{E} \exp(\varphi(Y_1)) < \infty$. Alors $n^{-1} \log \mathbb{E} \exp(\varphi(S_n)) \leq (n - 1)^{-1} \log \mathbb{E} \exp(\varphi(S_{n-1}))$ pour tout entier $n \geq 2$.

Preuve du Lemme 1. D'abord, comme φ est sous-additive, et comme les variables Y_1, Y_2, \dots, Y_n sont indépendantes, $\mathbb{E} \exp(\varphi(S_n)) \leq (\mathbb{E} \exp(\varphi(Y_1)))^n < \infty$.

Posons $S_n^k = S_n - Y_k$. Les variables S_n^k ont pour loi commune la loi de S_{n-1} . Soit f une variable aléatoire strictement positive et intégrable, mesurable pour la tribu \mathcal{B}_n engendrée par (Y_1, Y_2, \dots, Y_n) . Alors, le lemme 2.2 dans [5] (ou bien le lemme 2.2 dans [1]), qui est une conséquence immédiate de l'inégalité de tensorisation de l'entropie donnée dans [4], montre que, pour toute famille f_1, f_2, \dots, f_n de variables aléatoires \mathcal{B}_n -mesurables, intégrables, strictement positives et telles que f_k soit indépendante de Y_k pour chaque entier k ,

$$\text{Ent}(f) := \mathbb{E}(f \log f) - \mathbb{E}(f) \log \mathbb{E}(f) \leq \sum_{k=1}^n \mathbb{E}(f \log(f/f_k) + f_k - f). \quad (5)$$

Soit $T_n = \varphi(S_n)$ et $T_n^k = \varphi(S_n^k)$. Posons $F_n = \mathbb{E}(\exp(T_n))$ et $L_n = \log F_n$. Appliquons (5) avec $f = \exp(T_n)$ et $f_k = (F_n/F_{n-1}) \exp(T_n^k)$. Comme les variables T_n^k ont pour loi commune la loi de T_{n-1} , $\mathbb{E}(f_k) = \mathbb{E}(f)$ pour ce choix de f_k , et donc (5) assure que

$$\mathbb{E}(T_n \exp(T_n)) - L_n F_n \leq \mathbb{E} \left(\exp(T_n) \sum_{k=1}^n (T_n - T_n^k - L_n + L_{n-1}) \right). \quad (6)$$

Majorons le terme de droite dans (6). Comme φ est convexe,

$$T_n^1 + T_n^2 + \dots + T_n^n \geq n\varphi((S_n^1 + S_n^2 + \dots + S_n^n)/n) = n\varphi(S_n(n-1)/n) = (n-1)T_n$$

en se servant de l'homogénéité de φ . Donc $\sum_{k=1}^n (T_n - T_n^k) \leq T_n$, et (6) donne alors

$$\mathbb{E}(T_n \exp(T_n)) - L_n F_n \leq \mathbb{E}(T_n \exp(T_n)) + n(L_{n-1} - L_n) F_n,$$

ce qui conduit au lemme 1 après simplification. \square

Fin de la preuve du Théorème 1. Soit $Z_n^k = \sup\{nP_n(g) - g(X_k) : g \in \mathcal{G}\}$. On applique de nouveau le lemme 2.2 dans [5] en prenant cette fois ci $f_k = \exp(tZ_n^k + tE)$. Posons $F_n(t) = \mathbb{E}(\exp(tZ_n))$ et $D_k = Z_n - Z_n^k - E$. Alors, en appliquant le lemme 2.2 dans [5], et en se servant de l'équidistribution des variables,

$$tF'_n - F_n L_n \leq n\mathbb{E}(f_n(tD_n \exp(tD_n) - \exp(tD_n) + 1)). \tag{7}$$

Pour majorer les termes de droite, on se sert alors de la majoration suivante, obtenue dans [2] : pour tout $y \geq 0$ et tout $x \leq y$,

$$xe^x - e^x + 1 \leq \phi(y)(y(e^x - x - 1) + x^2/2) \quad \text{avec } \phi(y) = (ye^y - e^y + 1)/(y(e^y - 1 - y/2)) \tag{8}$$

(cette majoration provient de la croissance en x de $(xe^x - e^x + 1)/(y(e^x - x - 1) + x^2/2)$). Posons $b = 1 - E$. En prenant $x = tD_n$ et $y = tb$ dans (8), en multipliant par f_n l'inégalité obtenue et en remarquant que $Z_n^n = Z_{n-1}$, on obtient alors, en partant de (7), après calculs,

$$tF'_n - F_n L_n \leq n\phi(tb)(tb(F_n - e^{tE}F_{n-1}) + t^2\mathbb{E}((D_n^2/2 - bD_n)e^{tZ_{n-1}+tE})). \tag{9}$$

On note alors, comme dans [6], que si h est une fonction sélectionnée de façon mesurable en fonction de X_1, \dots, X_{n-1} de sorte que $Z_{n-1} = (n-1)P_{n-1}(h)$, alors $Z_n - Z_{n-1} \geq \eta_n = h(X_n)$. Donc $\eta_n - E \leq D_n \leq b = 1 - E$. Comme la fonction $x \rightarrow x^2/2 - bx$ est décroissante sur $]-\infty, b]$, il en résulte que

$$\mathbb{E}((D_n^2/2 - bD_n)e^{tZ_{n-1}+tE}) \leq \mathbb{E}(((\eta_n - E)^2/2 - b\eta_n + bE)e^{tZ_{n-1}+tE}).$$

Or $\mathbb{E}(\eta_n | X_1, \dots, X_{n-1}) = 0$ et $\mathbb{E}(\eta_n^2 | X_1, \dots, X_{n-1}) \leq \sigma^2$ (voir [6], inégalité (2.4), p. 1056). Donc, si $V = \sigma^2 + 2E - E^2$, la quantité ci-dessus est majorée par $V e^{tE} F_{n-1}/2$. Soit $l_n(t) = L_n(t)/n$. En reportant cette majoration dans (9) et en divisant par nF_n , il vient :

$$t'_n - l_n \leq t\phi(t(1-E))(1-E + (Vt/2 - 1 + E)e^{tE+L_{n-1}-L_n}). \tag{10}$$

Pour finir la majoration, on considère deux cas. Si $t \leq t_0 := 2(1-E)/V$, $Vt/2 - 1 + E \leq 0$. En appliquant (4), on a

$$\exp(tE + L_{n-1} - L_n) \geq \exp(tE - l_n) \geq 1 + (tE - l_n)$$

(comme $L_0 = 0$, cette minoration est encore vraie pour $n = 1$). Par conséquent, pour $t \leq t_0$,

$$t'_n - l_n \leq t\phi(t(1-E))(1-E + (Vt/2 - 1 + E)(1 + tE - l_n)). \tag{11}$$

Soit $z = (l_n/t) - E$. En divisant (11) par t^2 , il vient

$$z' \leq \phi(t(1-E))(V/2 + (1-E - Vt/2)z) \leq \phi(t(1-E))(V/2 + (1-E)z),$$

et donc $(1-E)z'/(V/2 + (1-E)z) \leq (1-E)\phi(t(1-E))$. Pour finir, on remarque que ϕ est la dérivée de la fonction $t \rightarrow \log(2e^t - 2 - t) - \log t$ (voir [2]). Donc en intégrant l'inéquation différentielle ci-dessus et en notant que $z(0) = 0$, on obtient

$$\log(V/2 + (1-E)z) - \log(V/2) \leq \log(2e^{(1-E)t} - 2 - (1-E)t) - \log((1-E)t),$$

ce qui établit (3) pour $t \leq t_0$.

Enfin, si $t \geq t_0 := 2(1-E)/V$, la majoration (3) est triviale : en effet,

$$(1-E)^{-2}(e^{(1-E)t} - (1-E)t - 1) \geq t^2/2 \geq (1-E)t/V,$$

et donc

$$tE + V(1-E)^{-2}(e^{(1-E)t} - (1-E)t - 1) \geq tE + t(1-E) \geq l_n(t)$$

puisque $Z_n \leq n$, ce qui garantit que $l_n(t) \leq t$. La preuve du théorème 1 est donc complète. \square

Remerciements

Je tiens à remercier le rapporteur pour une lecture attentive du manuscrit initial.

Références

[1] S. Boucheron, G. Lugosi, P. Massart, A sharp concentration inequality with applications, *Random Structures Algorithms* 3 (2000) 277–292.
 [2] O. Bousquet, Concentration inequalities for sub-additive functions using the entropy method, in: *Stochastic Inequalities and Applications*, in: *Progr. Probab.*, vol. 56, Birkhäuser, Basel, 2003, pp. 213–247.
 [3] T. Klein, Une inégalité de concentration à gauche pour les processus empiriques, *C. R. Acad. Sci. Paris, Sér. I* 334 (2002) 495–500.
 [4] M. Ledoux, On Talagrand's deviation inequalities for product measures, *ESAIM, P&S* 1 (1996) 63–87.
 [5] E. Rio, Inégalités de concentration pour les processus empiriques de classes de parties, *Probab. Theory Related Fields* 119 (2001) 163–175.
 [6] E. Rio, Une inégalité de Bennett pour les maxima de processus empiriques, *Ann. Inst. Henri Poincaré, Probab. Stat.* 38 (2002) 1053–1057.
 [7] M. Talagrand, New concentration inequalities in product spaces, *Invent. Math.* 126 (1996) 503–563.