Ann. I. H. Poincaré – PR 38, 6 (2002) 1053–1057

S0246-0203(02)01122-6/FLA

UNE INÉGALITÉ DE BENNETT POUR LES MAXIMA DE PROCESSUS EMPIRIQUES

A BENNET TYPE INEQUALITY FOR MAXIMA OF EMPIRICAL PROCESSES

Emmanuel RIO

UMR 8100 CNRS, Université de Versailles Saint-Quentin en Yvelines, 45 avenue des Etats Unis, 78035 Versailles, France

Reçu le 1 avril 2001, revisé le 22 décembre 2001

RÉSUMÉ. – Nous donnons des constantes dans l'inégalité de concentration de Talagrand pour les processus empiriques indexés par des classes de fonctions, en partant de la méthode de Herbst. Le point nouveau est que la constante du facteur variance est exacte, ce qui répond à une conjecture de Massart.

ABSTRACT. – We give new constants in Talagrand's concentration inequality for maxima of empirical processes. Our approach is based on the Herbst method. The improvement we get concerns the constant in the variance factor, which is the one conjectured by Massart.

MSC: 60E15; 60F10

1. Introduction

Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes de loi commune P à valeurs dans $\mathcal{X}$ espace polonais muni de sa tribu borélienne. Soit $\mathcal{F}$ une famille dénombrable de fonctions mesurables de $\mathcal{X}$ dans $\mathbb{R}$, de carré intégrable sous P. On pose

$$S_n(f) = f(X_1) + \cdots + f(X_n). \tag{1.1}$$

Dans cet article nous regardons les propriétés de concentration de la variable aléatoire $Z = \sup\{S_n(f)\colon\ f \in \mathcal{F}\}$ autour de sa moyenne. Ce problème a été étudié dans une série de travaux successifs, en particulier par Talagrand [9] au moyen d'inégalités

E-mail address: Emmanuel.Rio@math.uvsq.fr (E. Rio).

isopérimétriques pour les mesures produits. Il a obtenu une inégalité de type Bennett pour la déviation de Z par rapport à sa moyenne. Ensuite Ledoux [3] a montré que ces inégalités pouvaient aussi être obtenues à partir d'inégalités de type log-Sobolev pour les mesures produit. Cette technique permet de majorer la transformée de Laplace de Z. Massart [5] a clairement montré l'intérêt des méthodes entropiques (ou log-Sobolev) pour le calcul des constantes dans les inégalités de Talagrand. En particulier, pour les classes de fonctions positives, Massart [5] obtient comme fonction de taux celle d'une variable de loi de Poisson de paramètre $\mathbb{E}(Z)$. Cependant, pour les fonctions de signe quelconque, il utilise comme Talagrand et Ledoux une méthode de symétrisation, suivie des théorèmes de comparaison de Ledoux et Talagrand [4]. Le coût de cette méthode est la perte d'un facteur 2 dans les constantes, qui intervient lors de la symétrisation.

Dans cet article, nous montrons comment éviter la symétrisation. Nous reprenons ici un argument introduit partiellement dans Rio [6] pour les classes de fonctions finies. Dans l'inégalité provenant de la tensorisation de l'entropie, la variable Z_k' approchant Z au pas k sera la suivante : si f_k est la fonction de $\mathcal{F}$ réalisant le maximum de

$$S_n^k(f) = f(X_1) + \cdots + f(X_{k-1}) + f(X_{k+1}) + \cdots + f(X_n), \tag{1.2}$$

on prend $Z_k' = S_n(f_k)$. La différence $Z - Z_k'$ est alors positive. De plus on montre aisément que $\mathbb{E}(Z - Z_k') \leqslant \mathbb{E}(Z)/n$, ce qui donne un ordre de grandeur de $(Z - Z_k')$ suffisant pour obtenir les inégalités suivantes.

THÉORÈME 1.1. – *Soit $\mathcal{F}$ classe dénombrable de fonctions de $\mathcal{X}$ dans $]-\infty, 1]$, mesurables, de carré intégrable et d'espérance nulle sous P. Supposons de plus l'existence d'une constante positive σ^2 telle que $\sigma^2 \geqslant P(f^2)$ pour toute f dans $\mathcal{F}$. Alors, pour tout λ positif,*

$$\log \mathbb{E} \exp\bigl(\lambda(Z - \mathbb{E}(Z))\bigr) \leqslant \bigl(n\sigma^2 + 2\mathbb{E}(Z)\bigr)\lambda\bigl(\mathrm{e}^\lambda - 1\bigr)/2. \tag{a}$$

Par conséquent, si $v_n = n\sigma^2 + 2\mathbb{E}(Z)$, pour tout x positif,

$$\mathbb{P}\bigl(Z \geqslant \mathbb{E}(Z) + x\bigr) \leqslant \exp\left(-\frac{x}{2}\log\left(1 + \frac{x}{v_n}\right)\right). \tag{b}$$

De plus, pour tout y positif

$$\mathbb{P}\bigl(Z \geqslant \mathbb{E}(Z) + \sqrt{2v_n y} + y/2\bigr) \leqslant \exp(-y). \tag{c}$$

Remarque 1.1. – Les observations sont ici équidistribuées, contrairement aux auteurs précédents. Cependant, dans le cas non équidistribué, un résultat de Bretagnolle [2] montre que la transformée de Laplace associée au maximum d'un processus empirique provenant d'observations indépendantes et non équidistribuées est majorée par celle du maximum du processus empirique associé à des variables indépendantes ayant pour loi commune la loi moyenne, ceci au moins pour les classes de parties (il semble raisonnable de penser que le résultat de Bretagnolle peut s'étendre aux classes de fonctions).

2. De la méthode de Herbst aux inégalités de Bennett

Dans cette section, nous montrons le théorème 1.1 à l'aide de l'inégalité de tensorisation de l'entropie proposée par Ledoux [3] dont nous rappelons ici un corollaire.

LEMME 2.1 (Rio [8]). – *Soient* $(X_1, \ldots, X_n)$ *des variables aléatoires indépendantes à valeurs dans un espace polonais* $\mathcal{X}$. *Soit* $\mathcal{F}_n^k$ *la tribu engendrée par* $(X_i)_{i \neq k}$. *Posons* $\phi(x) = x \log x - x + 1$. *Soit* h *une variable aléatoire fonction mesurable de* $(X_1, \ldots, X_n)$, *strictement positive et intégrable. Alors, pour toute famille* (h_k) *de variables aléatoires réelles strictement positives et intégrables, respectivement* $\mathcal{F}_n^k$-*mesurables,*

$$\mathbb{E}(h \log h) - \mathbb{E}(h) \log \mathbb{E}(h) \leqslant \mathbb{E}\big(h_1 \phi(h/h_1)\big) + \cdots + \mathbb{E}\big(h_n \phi(h/h_n)\big).$$

Preuve du théorème 1.1. – Nous pouvons supposer $\mathcal{F}$ finie et procéder ensuite par passage à la limite. Comme Ledoux [3] on applique le lemme 2.1 à $h = \exp(\lambda Z)$. Définissons $Z_k = \sup\{S_n^k(f)\colon f \in \mathcal{F}\}$. Puisque $\mathcal{F}$ est finie, on peut opérer une sélection $\mathcal{F}_n^k$-mesurable d'une fonction f_k telle que $Z_k = S_n^k(f_k)$. On pose alors $Z_k' = S_n(f_k)$. Soient $h = \exp(\lambda Z)$, $h_k = \exp(\lambda Z_k)$ et $g_k = \exp(\lambda Z_k')$. Par convexité de ϕ,

$$h_k \phi(h/h_k) \leqslant h_k \phi(g_k/h_k) + (h - g_k)\phi'(h/h_k).$$

Comme $\phi'(x) = \log x$, il en résulte que

$$h_k \phi(h/h_k) \leqslant h_k \phi(g_k/h_k) + \lambda(h - g_k)(Z - Z_k).$$

Soit f dans $\mathcal{F}$ telle que $S_n(f) = Z$. Alors $S_n^k(f) \leqslant Z_k$, et donc $Z - Z_k \leqslant 1$, car f est bornée par 1. D'autre part $Z \geqslant Z_k'$, ce qui assure que $\lambda(h - g_k) \geqslant 0$ pour tout réel λ. Par conséquent

$$h_k \phi(h/h_k) \leqslant h_k \phi(g_k/h_k) + \lambda(h - g_k).$$

Notons F la transformée de Laplace de Z. En partant du lemme 2.1 et en appliquant l'inégalité ci-dessus, on obtient :

$$\begin{aligned} &\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \\ &\quad \leqslant \lambda \sum_{k=1}^n \mathbb{E}(h - h_k) + \sum_{k=1}^n \mathbb{E}\big(h_k \phi(g_k/h_k) + \lambda(h_k - g_k)\big). \end{aligned} \tag{2.1}$$

Soit λ positif. Pour majorer la quantité $\sum_{k=1}^n \mathbb{E}(h - h_k)$, on applique à nouveau le lemme 2.1 avec $h = \exp(\lambda Z)$. On choisit pour famille de variables $\mathcal{F}_n^k$-mesurables les variables $h_k' = \exp(\lambda Z_k + L(\lambda)/n)$, avec $L(\lambda) = \log F(\lambda)$. Le lemme 2.1 donne alors, après simplification,

$$\lambda F'(\lambda) \leqslant \mathbb{E}\left(\sum_{k=1}^n \lambda(Z - Z_k)\mathrm{e}^{\lambda Z}\right) + \mathrm{e}^{L(\lambda)/n} \sum_{k=1}^n \mathbb{E}(h_k) - nF(\lambda). \tag{2.2}$$

Mais

$$(n-1)Z = \sup\left\{\sum_{k=1}^n S_n^k(f)\colon f \in \mathcal{F}\right\} \leqslant Z_1 + \cdots + Z_n,$$

et donc, pour λ positif,

$$\mathbb{E}\left(\sum_{k=1}^{n} \lambda(Z - Z_k)\mathrm{e}^{\lambda Z}\right) \leqslant \mathbb{E}\big(\lambda Z \mathrm{e}^{\lambda Z}\big).$$

En reportant cette majoration dans (2.2), il vient :

$$\mathrm{e}^{L(\lambda)/n} \sum_{k=1}^{n} \mathbb{E}(h_k) - nF(\lambda) \geqslant 0.$$

Cette majoration assure que

$$\sum_{k=1}^{n} \mathbb{E}(h - h_k) \leqslant nF(\lambda)\big(1 - \mathrm{e}^{-L(\lambda)/n}\big) \leqslant F(\lambda)\log F(\lambda), \tag{2.3}$$

puisque $L(\lambda) \geqslant \lambda\mathbb{E}(Z) \geqslant 0$ pour λ positif.

Regardons maintenant la seconde partie du majorant dans (2.1). Soit $\eta_k = f_k(X_k)$. Alors $(g_k/h_k) = \exp(\lambda\eta_k)$. Posons $\psi(x) = x\mathrm{e}^x - \mathrm{e}^x + 1$. En factorisant h_k, on obtient :

$$\mathbb{E}\big(h_k\phi(g_k/h_k) + \lambda(h_k - g_k)\big) = \mathbb{E}\big(h_k\mathbb{E}_n^k\big(\psi(\lambda\eta_k) + \lambda\big(1 - \mathrm{e}^{\lambda\eta_k}\big)\big)\big),$$

si $\mathbb{E}_n^k$ désigne l'espérance conditionnelle par rapport à $\mathcal{F}_n^k$. Comme la fonction f_k est sélectionnée en fonction des observations $(X_i)_{i\neq k}$,

$$\mathbb{E}_n^k(\eta_k) = 0 \quad \text{et} \quad \mathbb{E}_n^k\big(\eta_k^2\big) = P\big(f_k^2\big) \leqslant \sigma^2. \tag{2.4}$$

Donc

$$\mathbb{E}_n^k\big(1 - \mathrm{e}^{\lambda\eta_k}\big) = \mathbb{E}_n^k\big(1 + \lambda\eta_k - \mathrm{e}^{\lambda\eta_k}\big).$$

Soit η_k^+ la partie positive de η_k. Puisque $\eta_k \leqslant 1$ et $1 + x - \mathrm{e}^x \leqslant 0$, on a, pour λ positif,

$$\lambda\mathbb{E}_n^k\big(1 - \mathrm{e}^{\lambda\eta_k}\big) \leqslant \mathbb{E}_n^k\big(\lambda\eta_k^+\big(1 + \lambda\eta_k - \mathrm{e}^{\lambda\eta_k}\big)\big). \tag{2.5}$$

Soit alors $\gamma(x) = \psi(x) + x_+(1 + x - \mathrm{e}^x)$. En combinant (2.1), (2.3) et (2.5),

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leqslant \lambda^2 F'(\lambda) + \sum_{k=1}^{n} \mathbb{E}\big(h_k E_n^k\big(\gamma(\lambda\eta_k)\big)\big).$$

Si x est positif, $\gamma(x) = x^2 - (\mathrm{e}^x - 1 - x)$ et donc $\gamma(x) \leqslant (x^2/2)$; de même, pour x négatif, $\gamma(x) = \psi(x)$ et donc $\gamma(x) \leqslant (x^2/2)$. Donc, d'après (2.3) et (2.4)

$$\lambda F'(\lambda) - (1 + \lambda)F(\lambda)\log F(\lambda) \leqslant \frac{\lambda^2\sigma^2}{2} \sum_{k=1}^{n} \mathbb{E}(h_k). \tag{2.6}$$

En appliquant à la fonctionnelle convexe $\sup\{f(X_1) + \cdots + f(X_n)\colon\ f \in \mathcal{F}\}$ l'inégalité de Jensen conditionnellement à $\mathcal{F}_n^k$, on note alors que

$$\mathbb{E}\big(e^{\lambda Z}\big) \geqslant \mathbb{E}\big(e^{\lambda E_n^k Z}\big) \geqslant \mathbb{E}\big(e^{\lambda Z_k}\big) \quad \text{pour tout } \lambda > 0. \tag{2.7}$$

Par conséquent le logarithme L de la transformée de Laplace de Z vérifie l'inéquation différentielle suivante : pour tout λ positif,

$$\lambda L'(\lambda) - (1+\lambda)L(\lambda) \leqslant n\sigma^2\lambda^2/2. \tag{ED}$$

En divisant par $\lambda^2 e^{\lambda}$, on en déduit que

$$\left(\frac{2L}{\lambda e^{\lambda}}\right)' \leqslant n\sigma^2 \exp(-\lambda).$$

La majoration (a) s'obtient alors par intégration de cette inéquation différentielle. (b) s'obtient en partant de (a) et en appliquant l'inégalité de Markov à $\exp(\lambda Z)$ avec seuil $\exp(\lambda x)$ pour $\lambda = \log(1 + x/v_n)$. Enfin (c) se déduit de (a) à partir de la majoration $\lambda(e^{\lambda} - 1) \leqslant 2\lambda^2/(2-\lambda)$ et du calcul de la transformée de Legendre donné dans Rio [7, p. 153].

Nota. – Après la rédaction de cet article, Olivier Bousquet a montré comment améliorer la majoration (a) du théorème 1.1 pour obtenir la vraie inégalité de Bennett.

RÉFÉRENCES

[1] O. Bousquet, Thèse de doctorat, Centre de mathématiques appliquées de l'Ecole Polytechnique (en préparation).

[2] J. Bretagnolle, Statistique de Kolmogorov–Smirnov pour un échantillon non équiréparti, dans Aspects statistiques et aspects physiques des processus gaussiens, Éditions du centre national de la recherche scientifique, Paris, 1981.

[3] M. Ledoux, On Talagrand's deviation inequalities for product measures. European series in applied and industrial mathematics, Probab. Statist. 1 (1996) 63–87.

[4] M. Ledoux, M. Talagrand, Probability in Banach Spaces. Isoperimetry and Processes, Springer, Berlin, 1991.

[5] P. Massart, About the constants in Talagrand's concentration inequalities for empirical processes, Ann. Probab. 28 (2000) 863–884.

[6] E. Rio, Inégalités exponentielles pour les processus empiriques, C. R. Acad. Sci. Paris, Série I 330 (2000) 597–600.

[7] E. Rio, in : J.M. Ghidaglia, X. Guyon (Eds.), Théorie asymptotique des processus aléatoires faiblement dépendants, in : Mathématiques et Applications, Vol. 31, Springer, Berlin, 2000.

[8] E. Rio, Inégalités de concentration pour les processus empiriques de classes de parties, Probab. Theory Related Fields 119 (2001) 163–175.

[9] M. Talagrand, New concentration inequalities in product spaces, Invent. Math. 126 (1996) 503–563.