Inégalités exponentielles et inégalités de concentration

Emmanuel Rio

Inégalités exponentielles et inégalités de concentration

Emmanuel Rio, Laboratoire de mathématiques, UMR 8100 CNRS,
Université de Versailles, Bâtiment Fermat, 45 Avenue des Etats Unis,
78035 VERSAILLES Cedex France

Ces notes proviennent d'un cours sur les inégalités de concentration fait à Bordeaux lors d'une année passée en délégation à l'INRIA Bordeaux Sud-Ouest dans l'équipe projet de Pierre Del Moral. Je tiens à remercier Pierre Del Moral, Bernard Bercu et François Caron pour leur accueil chaleureux lors de cette année 2008-2009.

1. Introduction

Le but de ce cours est, dans un premier temps, de donner des inégalités classiques pour les sommes de variables aléatoires bornées ou ayant des moments exponentiels, ainsi que pour les martingales à temps discret ayant des accroissements bornés. Nous donnons ensuit quelques applications des ces inégalités à la concentration de fonctions lipschitziennes de vecteurs indépendants autour de leur espérance, ainsi qu'une première inégalité de concentration, de type Hoeffding, pour la concentration du maximum d'un processus empirique autour de sa moyenne. Dans la dernière partie du cours nous donnons des constantes précises dans les inégalités de concentration de Talagrand (1996) sur la concentration du maximum d'un processus empirique autour de sa moyenne. Ces inégalités, plus fines que les inégalités de type Hoeffding, sont obtenues ici à l'aide d'inégalités de type log-Sobolev.

2. Les bornes de Chernoff

2.1. Dualité de Young

Nous rappelons ici quelques propriétés essentielles et bien connues de la transformée de Young. Considérons la classe de fonctions convexes

$$\bar{\Phi} = \{\phi : \mathbb{R}^+ \to \bar{\mathbb{R}}^+ : \ \phi \ \text{ convexe, croissante, continue à gauche, } \phi(0) = 0\}.$$

Pour ϕ dans $\bar{\Phi}$, soit D_ϕ son domaine de définition et

$$G_\phi = \{(x, y) \in D_\phi \times \mathbb{R} : y > \phi(x)\}$$

le sur-graphe de ϕ et $\bar{G}_\phi$ son adhérence. La duale de young ϕ^* de la fonction ϕ est définie par

$$\phi^*(\lambda) = \sup_{x \in D_\phi} (\lambda x - \phi(x)) \quad \text{pour} \quad \lambda \geq 0.$$

Par définition de ϕ^*,

$$\lambda x \leq \phi(x) + \phi^*(\lambda) \text{ pour tout } (x,\lambda) \in \mathbb{R}^+ \times \mathbb{R}^+.$$

Cette inégalité est appelée inégalité de Young.

Ainsi $z = \phi^*(\lambda)$ si et seulement si la droite d'équation $y = \lambda x - z$ soit intersecte $\bar{G}_\phi$ et n'intersecte pas G_ϕ, soit est asymptote à la courbe $y = \phi(x)$. Aussi

$$z \geq \phi^*(\lambda) \text{ si et seulement si } \{(x,y) \in \mathbb{R}^2 : y = \lambda x - z\} \cap G_\phi = \emptyset. \tag{1}$$

Nous allons maintenant déduire de (1) l'appartenance de ϕ^* à $\bar{\Phi}$. En effet $\phi(0) = 0$ car la droite $y = -z$ n'intersecte pas G_ϕ si et seulement si $z \geq 0$, ϕ est croissante car, pour $\lambda > \lambda'$, la droite d'équation $y = \lambda' x - \phi(\lambda)$ n'intersecte pas G_ϕ, ce qui montre que $\phi(\lambda) \geq \phi(\lambda')$ via (1). Pour montrer la convexité de ϕ^*, considérons le point d'intersection A des deux droites d'équations respectives $y = \lambda x - \phi(\lambda)$ et $y = \lambda' x - \phi(\lambda')$, où $\lambda' < \lambda$. Soit t dans $[0,1]$. La droite passant par A et ayant pour pente $t\lambda + (1-t)\lambda'$ n'intersecte pas G_ϕ et coupe l'axe $x = 0$ à la hauteur $t\phi(\lambda) + (1-t)\phi(\lambda')$. La convexité de ϕ^* découle donc de (1). Enfin, si λ_0 est la borne supérieure du domaine de définition de ϕ^* et si $\lim_{\lambda \nearrow \lambda_0} \phi^*(\lambda) = l$, la famille de droites $y = \lambda x - \phi(\lambda)$ a pour limite la droite $y = \lambda_0 x - l$ quand λ croit vers λ_0, et par conséquent cette droite n'intersecte pas G_ϕ. Donc $\phi(\lambda_0) = l$ par (1) et ϕ^* est continue à gauche.

Montrons maintenant que $\phi^{**} = \phi$. Comme ϕ est convexe

$$G_\phi = \bigcap_{(\lambda,z) : z \geq \phi^*(\lambda)} \{(x,y) \in \mathbb{R}_+ \times \mathbb{R} : y > \lambda x - z\}. \tag{2}$$

Or, à x fixé, $y > \phi^{**}(x)$ si et seulement si la droite de pente x passant par $(0,-y)$ n'intersecte pas $\bar{G}_{\phi^*}$, autrement dit si, pour tout couple (λ, z) de réels positif tel que $z \geq \phi^*(\lambda)$, $z > \lambda x - y$. Il résulte donc de (2) que $G_\phi = G_{\phi^{**}}$, et $\phi = \phi^{**}$.

2.2. Identités diverses

Montrons maintenant que les dérivées de ϕ est de ϕ^* sont liées par la relation suivante:

$$(\phi^*)'(\lambda + 0) = \phi'^{-1}(\lambda + 0) \text{ et } (\phi^*)'(\lambda - 0) = \phi'^{-1}(\lambda - 0) = \phi'^{-1}(\lambda). \tag{3}$$

Pour démontrer (3), considérons les points de contact de la droite $y = \lambda x - \phi^*(\lambda)$ avec $\bar{G}_\phi$. Puisque nous avons pris l'inverse continue à gauche, le point de contact le plus à droite a pour abcisse $x(\lambda) = \phi'^{-1}(\lambda + 0)$. Étudions la différentiabilité à droite de ϕ^*. Soit $\varepsilon > 0$ quelconque. Considérons la droite de pente $\lambda + \varepsilon$ passant par le point $(x(\lambda), \phi(x(\lambda)))$.

Cette droite coupe l'axe $x = 0$ à la hauteur $\phi^*(\lambda) + \varepsilon x(\lambda)$. Or, pour $x > x(\lambda)$, $\phi'(x) > \lambda$. Donc pour $x > x(\lambda)$ et ε assez petit, $\phi'(x) \geq \lambda + \varepsilon$, ce qui implique que

$$\phi^*(\lambda) + \varepsilon x(\lambda) \leq \phi^*(\lambda + \varepsilon x) \leq \phi^*(\lambda) + \varepsilon x.$$

De là, nous tirons la première partie de (3). Pour obtenir le seconde partie, il suffit de considérer le point de contact le plus à gauche.

Fonction réciproque de ϕ^*. La formule ci-dessous (Rio (2000), page 159) donne un calcul direct:

$$\phi^{*-1}(x) = \inf_{t \in D_\phi} t^{-1}(\phi(t) + x) \quad \text{pour tout } x > 0. \tag{4}$$

Pour montrer (4), on note que la pente de la droite joignant $(0, -x)$ à $(t, \phi(t))$ est égale à $t^{-1}(\phi(t)+x)$. Soit t_0 est le point où cette pente est minimale et λ_0 la pente correspondante. Alors $\phi^*(\lambda_0) = x$, car la droite partant de $(0, -x)$ ayant pour pente λ_0 n'intersecte pas G_ϕ et a au moins un point commun avec la courbe $y = \phi(t)$, ce qui établit (4).

2.3. Borne de Chernoff pour les martingales

On considère une filtration croissante $(\mathcal{F}_n)_{n\geq 0}$ et une suite de variables aléatoires réelles intégrables $(S_n)_n$, adaptée à cette filtration, telle que $S_0 = 0$.

Notation 1. Pour Z v.a.r positive ou intégrable, on note $E_i(Z)$ l'espérance conditionnelle de Z sachant $\mathcal{F}_i$.

On suppose que $(S_n)_n$ est une martingale: $E_{n-1}(S_n) = S_{n-1}$ pour tout $n > 0$. On pose $S_n^* = \max(0, S_1, \ldots, S_n)$. Rappelons tout d'abord la méthode de Crámer-Chernoff dans le lemme suivant.

Lemme 1. *Si γ est une fonction de $\bar{\Phi}$ telle que $\gamma(t) \geq \log \mathbb{E}(\exp(tS_n))$ pour tout $t > 0$, alors, pour tout λ positif,*

$$\log(\mathbb{P}(S_n^* \geq \lambda)) \leq -\gamma^*(\lambda).$$

Remarque 1. Comme $S_n \leq S_n^*$, l'inégalité ci-dessus est vraie avec S_n au lieu de S_n^*.

Preuve du lemme 1. Pour t dans le domaine de définition de γ, Posons $M_k(t) = \exp(tS_k)$. Pour t positif, $(M_k(t))_{k\geq 0}$ est une sous-martingale positive, par convexité de la fonction exponentielle, et intégrable car t est dans le domaine de définition de γ. Donc, d'après l'inégalité maximale de Doob,

$$\log \mathbb{P}(S_n^* \geq \lambda) \leq \log \mathbb{E}(\exp(tS_n - t\lambda)) \leq \gamma(t) - t\lambda \tag{5}$$

pour tout $t > 0$. Le lemme 1 s'obtient alors en optimisant (5) en t. ■

3. Les inégalités exponentielles de Hoeffding

Ces inégalités ont été obtenues par Hoeffding (1962), pour les sommes de variables aléatoires indépendantes, et étendues aux martingales par Azuma (1967). Elles s'appliquent aisément aux fonctions séparément lipschitziennes de variables aléatoires indépendantes, comme l'a montré Mc Diarmid (1989).

3.1. Inégalités de Hoeffding pour les martingales

Nous allons maintenant montrer l'inégalité exponentielle de Hoeffding-Azuma pour les martingales à accroissements bornés.

Théorème 1. - Inégalité de Hoeffding-Azuma - *Soit* $(S_n)_n$ *une martingale avec* $S_0 = 0$. *On suppose qu'il existe une suite adaptée* $(A_n)_n$ *de variables aléatoires réelles et une suite* $(c_n)_n$ *des constantes positives telles que*

$$A_{n-1} \leq S_n \leq A_{n-1} + c_n \text{ p.s.}$$

pour tout $n > 0$. *Alors, pour tout* t *positif,*

$$(a) \qquad \log \mathbb{E}(\exp(tS_n)) \leq (c_1^2 + c_2^2 + \cdots + c_n^2)t^2/8.$$

Par conséquent, pour tout x *positif,*

$$(b) \qquad \mathbb{P}(S_n^* \geq x) \leq \exp\Big(-2x^2/(c_1^2 + c_2^2 + \cdots + c_n^2)\Big)$$

(la même inégalité vaut donc pour S_n*).*

Preuve. L'inégalité (b) découle de (a) via le lemme 1. Montrons (a) par récurrence. La preuve est fondée sur le lemme ci-dessous.

Lemme 2. Soit $\mathcal{A}$ une tribu et Z variable aléatoire réelle à valeurs dans $[-m, m]$ p.s. Alors, pour tout t positif, p.s.

$$\log \mathbb{E}(\exp(tZ) \mid \mathcal{A}) \leq t\mathbb{E}(Z \mid \mathcal{A}) + t^2(m^2/2).$$

Preuve du lemme 2. Par une dilatation, on peut se ramener à $m = 1$. Si Z est p.s. dans $[-1, 1]$, alors, par convexité de la fonction exponentielle, p.s.

$$2\exp(tZ) \leq (1 - Z)\exp(-t) + (1 + Z)\exp(t)$$

Posons $M = \mathbb{E}(Z \mid \mathcal{A})$. En prenant l'espérance conditionnelle dans l'inégalité ci-dessus,

$$\mathbb{E}(\exp(tZ) \mid \mathcal{A}) \leq \cosh t + M \sinh t \text{ p.s.}$$

Par conséquent, en posant $f(t) = \cosh t + M \sinh t$, p.s.

$$\log \mathbb{E}(\exp(tZ) \mid \mathcal{A}) \leq \log f(t).$$

Comme $f'' = f$,

$$(\log f)'' = (f''/f) - (f'/f)^2 \leq 1,$$

et donc, comme $f'(0)/f(0) = M$, par la formule de Taylor, $\log f(t) \leq Mt + t^2/2$ p.s., ce qui achève la preuve du lemme 2. ■

Preuve de (a) du théorème 1. On procède par récurrence sur n. Pour $n = 0$, $S_0 = 0$, donc (a) est vraie avec 0 comme majorant. Si (a) est vraie au rang $n-1$, alors, au rang n, en appliquant le lemme 2 avec $Z = S_n - A_{n-1} - (c_n/2)$, $m = c_n/2$ et $\mathcal{A} = \mathcal{F}_{n-1}$, on obtient

$$\log E_{n-1}(\exp(tS_n)) \leq tS_{n-1} + c_n(t^2/8) \text{ p.s.}$$

Par conséquent

$$\mathbb{E}(\exp(tS_n)) = \mathbb{E}(E_{n-1}(\exp(tS_n))) \leq \exp(c_n t^2/8)\mathbb{E}(\exp(tS_{n-1}))$$

et donc, si (a) est vraie au rang $n-1$, alors (a) est vraie au rang n. Donc, par récurrence sur n, (a) est établie. ■

3.2. Application: fonctions lipschitziennes de v.a. indépendantes

Dans ce qui suit, $(\mathcal{X}, d)$ est un espace métrique de diamètre fini Δ et f une fonction de $\mathcal{X}^n$ dans $\mathbb{R}$. On suppose qu'il existe des constantes positives $c_1, c_2, \ldots c_n$ telles que

$$|f(x_1, x_2, \ldots, x_n) - f(y_1, y_2, \ldots, y_n)| \leq c_1 d(x_1, y_1) + c_2 d(x_2, y_2) + \cdots + c_n d(x_n, y_n)$$

pour tous les n-uplets $x = (x_1, x_2, \ldots, x_n)$ et $y = (y_1, y_2, \ldots, y_n)$ possibles.

Corollaire 1. - Mc-Diarmid - *Soit $\xi_1, \xi_2, \ldots, \xi_n$ une suite de v.a. indépendantes à valeurs dans $\mathcal{X}$ et $Z = f(\xi_1, \xi_2, \ldots, \xi_n)$. Alors, pour tout t positif,*

$$(a) \qquad \log \mathbb{E}(\exp(tZ)) \leq t\mathbb{E}(Z) + (c_1^2 + c_2^2 + \cdots + c_n^2)\Delta^2 t^2/8.$$

Par conséquent, pour tout x positif,

$$(b) \qquad \mathbb{P}(Z - \mathbb{E}(Z) \geq \Delta x) \leq \exp\big(-2x^2/(c_1^2 + c_2^2 + \cdots + c_n^2)\big).$$

Preuve pour les espaces polonais. (b) découle de (a). Il suffit de montrer (a) pour $\Delta = 1$. Quitte à retrancher l'espérance, on peut supposer que $\mathbb{E}(Z) = 0$. On considère la filtration finie $(\mathcal{F}_k)_k$ définie par $\mathcal{F}_k = \sigma(\xi_1, \ldots, \xi_k)$ pour k dans $[1, n]$ et par $\mathcal{F}_0 = \{\emptyset, \Omega\}$, de telle sorte que E_0 est l'opérateur d'espérance. Soit $S_k = E_k(Z)$. La suite $(S_k)_k$ est une martingale adaptée à la filtration $(\mathcal{F}_k)_k$ et telle que $S_0 = 0$. Posons

$$Z_k = \inf_{x \in \mathcal{X}} f(\xi_1, \ldots \xi_{k-1}, x, \xi_{k+1}, \ldots, \xi_n).$$

Comme $\mathcal{X}$ est de diamètre égal à 1 et f lipschitzienne de rapport c_k en la variable x_k, p.s.

$$Z_k \leq Z \leq Z_k + c_k.$$

Par conséquent, p.s.

$$E_k(Z_k) \leq S_k \leq E_k(Z_k) + c_k.$$

Mais Z_k est une fonction mesurable de $(\xi_i)_{i \neq k}$. Puisque les variables aléatoires $\xi_1, \xi_2, \ldots, \xi_n$ sont indépendantes, il en résulte que $E_k(Z_k)$ est $\mathcal{F}_{k-1}$-mesurable, et donc $E_k(Z_k) = E_{k-1}(Z_k)$. Donc, si $A_{k-1} = E_{k-1}(Z_k)$, on a $A_{k-1} \leq Z_k \leq A_{k-1} + c_k$ p.s., et par conséquent, les hypothèses du théorème 1 sont satisfaites, ce qui montre (a). ■

4. Les inégalités de type Bennett

Ces inégalités sont, à l'origine, des inégalités exponentielles pour les sommes de variables aléatoires indépendantes bornées de petite variance. Elles sont d'autant plus performantes que la variance de la somme est petite.

4.1. Inégalité de Bennett pour les martingales

Freedman (1975) a donné des extensions des inégalités de Bennett (1963) aux martingales à accroissements bornés.

Théorème 2. - Freedman (1975) - *Soit $(S_n)_n$ une martingale adaptée à la filtration $(\mathcal{F}_n)_n$ avec $S_0 = 0$. On pose $X_n = S_n - S_{n-1}$. On suppose qu'il existe une constante strictement positive M telle que*

$$S_n \leq S_{n-1} + M \text{ p.s.}$$

pour tout $n > 0$. Soit

$$V_\infty(n) = \Big\| \sum_{k=1}^{n} E_{k-1}(X_k^2) \Big\|_\infty.$$

Alors, pour tout t positif,

$$(a) \qquad \log \mathbb{E}(\exp(tS_n)) \leq M^{-2} V_\infty(n)(\exp(tM) - tM - 1).$$

Soit $h(x) = (1+x)\log(1+x) - x$. Si $V_\infty(n)$ est fini, alors, pour tout x positif,

$$(b) \qquad \mathbb{P}(S_n^* \geq x) \leq \exp\Big(-\frac{V_\infty(n)}{M^2} h\Big(\frac{xM}{V_\infty(n)}\Big)\Big) \leq \exp\Big(-\frac{x}{2M}\log\Big(1+\frac{xM}{V_\infty(n)}\Big)\Big).$$

(la même inégalité vaut donc pour S_n).

Preuve. Quitte à diviser par M, on peut supposer $M = 1$. Soit

$$Q_n = \sum_{k=1}^n E_{k-1}(X_k^2).$$

On considère, pour $t > 0$, la suite adaptée $(M_n(t))_n$ donnée par $M_0(t) = 1$ et

$$M_n(t) = \exp(tS_n - Q_n(e^t - t - 1)).$$

Montrons que la suite $(M_n(t))_n$ est une sur-martingale:

$$M_n(t) = M_{n-1}(t)\exp(tX_n)\exp(-E_{n-1}(X_n^2)(e^t - t - 1)).$$

Donc

$$E_{n-1}(M_n(t)) = M_{n-1}(t)\exp(\log(E_{n-1}\exp(tX_n)) - E_{n-1}(X_n^2)(e^t - t - 1))$$

Comme $\log x \leq x - 1$ et $E_{n-1}(X_n) = 0$ p.s.,

$$\log(E_{n-1}\exp(tX_n)) \leq E_{n-1}(\exp(tX_n) - tX_n - 1).$$

D'autre part, la fonction $x \to x^{-2}(e^x - x - 1)$ (prolongée par continuité en 0) est croissante sur $\mathbb{R}$ d'après la règle de l'Hospital pour la monotonie, dont on pourra trouver un énoncé dans Pinelis (2001). Comme $X_n \leq 1$ p.s., il en résulte que

$$\exp(tX_n) - tX_n - 1 \leq X_n^2(e^t - t - 1).$$

Par conséquent

$$\log(E_{n-1}\exp(tX_n)) \leq E_{n-1}(X_n^2)(e^t - t - 1),$$

et de là,

$$E_{n-1}(M_n(t)) \leq M_{n-1}(t),$$

ce qui montre la propriété de sur-martingale. On a donc $\mathbb{E}(M_n(t)) \leq \mathbb{E}(M_0(t)) = 1$. Comme

$$\exp(tS_n) \leq M_n(t)\exp(\|Q_n\|_\infty(e^t - t - 1)),$$

il en résulte que

$$\mathbb{E}(\exp(tS_n) \leq \exp(\|Q_n\|_\infty(e^t - t - 1)),$$

ce qui montre (a). Pour obtenir (b), on note que, si $\gamma(t) = e^t - t - 1$, alors $\gamma^* = h$, et on applique le lemme 1. ∎

4.2. Application: fonctions lipschitziennes de v.a. indépendantes

Dans ce qui suit, $(\mathcal{X}, d)$ est un espace métrique de diamètre fini Δ et f une fonction de $\mathcal{X}^n$ dans $\mathbb{R}$. On suppose que

$$|f(x_1, x_2, \dots, x_n) - f(y_1, y_2, \dots, y_n)| \leq d(x_1, y_1) + d(x_2, y_2) + \cdots + d(x_n, y_n)$$

pour tous les n-uplets $x = (x_1, x_2, \dots, x_n)$ et $y = (y_1, y_2, \dots, y_n)$ possibles.

Corollaire 2. *Soit $\xi_1, \xi_2, \dots, \xi_n$ une suite de v.a. indépendantes à valeurs dans $\mathcal{X}$ et $Z = f(\xi_1, \xi_2, \dots, \xi_n)$. Si $\xi'_1, \xi'_2, \dots, \xi'_n$ est une copie indépendante de la suite ci-dessus, on pose*

$$D_n = \frac{1}{2} \sum_{k=1}^{n} \mathbb{E}\big((d(\xi_k, \xi'_k))^2\big).$$

Alors, pour tout t positif,

$$\log \mathbb{E}(\exp(tZ)) \leq t\mathbb{E}(Z) + \Delta^{-2} D_n (e^{\Delta t} - \Delta t - 1). \tag{a}$$

Par conséquent, pour tout x positif,

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq x) \leq \exp\Big(-\Delta^{-2} D_n h(x\Delta / D_n)\Big). \tag{b}$$

Remarque 2. D_n est un majorant de la variance de Z. Cependant, D_n est souvent beaucoup trop grand par rapport à la variance de Z.

Preuve pour les espaces polonais. (b) découle de (a). Il suffit de montrer (a) pour $\Delta = 1$. Quitte à retrancher l'espérance, on peut supposer que $\mathbb{E}(Z) = 0$. Reprenons les notations de la section 3.2.

Soit donc $S_k = E_k(Z)$. Posons

$$f_k(x_1, \dots x_k) = \int_{\mathcal{X}^{n-k}} f(x_1, \dots x_k, x_{k+1}, \dots x_n) dP_{\xi_{k+1}}(x_{k+1}) \dots dP_{\xi_n}(x_n).$$

La fonction f_k est 1-lipschitzienne en chaque variable. Comme les v.a. $\xi_1, \xi_2, \dots, \xi_n$ sont indépendantes, $S_k = f_k(\xi_1, \cdots \xi_k)$. Soit maintenant

$$A_{k-1} = \inf_{x_k \in \mathcal{X}} f_k(\xi_1, \dots, \xi_{k-1}, x_k).$$

La v.a. A_{k-1} est $\mathcal{F}_{k-1}$-mesurable. De plus, comme f_k est 1-lipschitzienne en x_k, p.s.

$$A_{k-1} \leq S_k \leq A_{k-1} + 1,$$

car le diamètre de $\mathcal{X}$ vaut 1. En prenant l'espérance conditionnelle par rappor à $\mathcal{F}_{k-1}$, il vient, p.s.

$$A_{k-1} \leq S_{k-1} \leq A_{k-1} + 1$$

et, par conséquent

$$S_k \leq S_{k-1} + 1. \tag{6}$$

Pour appliquer le théorème 2, il reste à montrer que, p.s.

$$E_{k-1}((S_k - S_{k-1})^2) \leq \frac{1}{2}\mathbb{E}(d(\xi_k, \xi'_k))^2. \tag{7}$$

Soit $Y = (\xi_1, \ldots, \xi_{k-1})$. Pour montrer (7), on note d'abord que, pour tout y dans $\mathcal{X}^{k-1}$,

$$\mathbb{E}((S_k - S_{k-1})^2 \mid Y = y) = \mathbb{E}\big((f_k(y, \xi_k) - \mathbb{E}f_k(y, \xi_k))^2\big) = \frac{1}{2}\mathbb{E}\big((f_k(y, \xi_k) - f_k(y, \xi'_k))^2\big).$$

Mais f_k est 1-lipschitzienne p.r. x_k. Donc

$$(f_k(y, \xi_k) - f_k(y, \xi'_k))^2 \leq d^2(\xi_k, \xi'_k),$$

et donc, pour tout y dans $\mathcal{X}^{k-1}$,

$$\mathbb{E}((S_k - S_{k-1})^2 \mid Y = y) \leq \frac{1}{2}\mathbb{E}(d^2(\xi_k, \xi'_k)),$$

ce qui établit (7). Le théorème 2 s'applique donc, et donne le résultat.

5. Inégalités de Bernstein pour les sommes de v.a. indépendantes

Dans ce chapitre, nous allons donner des inégalités proches des inégalités de Bennett, dites inégalités de Bernstein.

Théorème 3. *Soient $X_1, \ldots, X_n$ une famille de variables aléatoires réelles indépendantes et centrées. Supposons qu'il existe des constantes positives V et K telles que, pour tout t positif,*

$$\log \mathbb{E} \exp(tS_n) \leq \gamma(t) = \frac{Vt^2}{2(1 - Kt)}. \tag{a}$$

Alors, pour tout x positif,

$$\mathbb{P}(S_n^* \geq \sqrt{2Vx} + Kx) \leq \exp(-x). \tag{b}$$

Remarque 3. L'inégalité de Bernstein usuelle est la suivante: sous (a),

$$\mathbb{P}(S_n^* \geq t) \leq \exp\Big(-\frac{t^2}{2(V+Kt)}\Big) := \exp(-\varphi(t)).$$

Pour comparer cette inégalité avec (b), on calcule la fonction réciproque φ^{-1}. Comme

$$\varphi^{-1}(x) = \sqrt{2Vx + (Kx)^2} + Kx.$$

Donc (b) est toujours meilleure: la perte asymptotique quand x tend vers l'infini est d'un facteur 2.

Exemple: variables de loi exponentielle. Soit $Z_1, Z_2, \ldots$ une suite de v.a.r. indépendantes de loi exponentielle de paramètre 1 et $X_k = Z_k - 1$. Alors, pour t positif,

$$\log \mathbb{E} \exp(tS_n) = n(-\log(1-t) - t) \leq nt^2/(2-2t).$$

Donc

$$\mathbb{P}(S_n^* \geq \sqrt{2nt} + t) \leq \exp(-t)$$

Preuve. Il suffit de montrer que

$$\gamma^{*-1}(x) = \sqrt{2Vx} + Kx. \tag{8}$$

D'après (4), on sait que

$$\gamma^{*-1}(x) = \inf_{t \in]0,1/K[} \Big(\frac{Vt}{2(1-Kt)} + \frac{x}{t}\Big).$$

On pose alors $s = -K + (1/t)$: avec ce changement de variable, s varie dans $]0, +\infty[$ et

$$\frac{Vt}{2(1-Kt)} + \frac{x}{t} = \frac{V}{2s} + xs + xK$$

L'expression en s atteint son minimum quand $s = (V/2x)^{1/2}$, ce qui donne (8). ■

La proposition suivante montre les propriétés particulières des fonctions convexes ci-dessus intervenant dans les inégalités de Bernstein. C'est un cas particulier d'un résultat plus général, à savoir le lemme 2.1 dans Rio (1994), dont la preuve est due à Bretagnolle.

Proposition 1. Merlevède, Peligrad et Rio (2011) *Soit $X_1, X_2, \ldots, X_n$ suite finie de variables aléatoires réelles centrées telles que, pour tout $t \geq 0$,*

$$\log \mathbb{E} \exp(tX_k) \leq (\sigma_k t)^2/(2 - 2c_k t) \quad \textit{tant que } c_k t < 1,$$

pour des constantes positives $\sigma_1, \sigma_2, \ldots, \sigma_n$ *et* $c_1, c_2, \ldots, c_n$. *Alors, pour tout* $t \geq 0$,

$$(a) \qquad \log \mathbb{E} \exp(tS_n) \leq (\sigma t)^2/(1-ct)$$

avec $c = c_1 + c_2 + \cdots + c_n$ *et* $\sigma = \sigma_1 + \sigma_2 + \cdots + \sigma_n$. *Si de plus les variables aléatoires sont indépendantes, alors (a) est vraie avec* $c = \max(c_1, c_2, \ldots, c_n)$ *et* $\sigma = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{1/2}$.

Preuve. Le cas indépendant est trivial. La proposition 1 se déduit du cas $n = 2$ par récurrence. Soit L la log-Laplace de $X_1 + X_2$. Definissons γ_k par

$$\gamma_k(t) = (\sigma_k t)^2/(1-c_k t) \text{ pour } t \in [0, 1/c_k[\text{ and } \gamma_k(t) = +\infty \text{ pour } t \geq 1/c_k.$$

Pour u dans $]0,1[$, soit $\gamma_u(t) = u\gamma_2(t/u) + (1-u)\gamma_1(t/(1-u))$. D'après l'inégalité de Hölder appliquée avec $p = 1/u$ and $q = 1/(1-u)$, on a $L(t) \leq \gamma_u(t)$ pour tout t positif. Pour t dans $[0, 1/c[$, on prend $u = (\sigma_2/\sigma)(1-ct) + c_2 t$. Pour ce choix de u,

$$1 - u = (\sigma_1/\sigma)(1-ct) + c_1 t,$$

et donc u est dans $]0,1[$. De plus $\gamma_u(t) = (\sigma t)^2/(1-ct)$, ce qui montre la proposition 1 pour $n = 2$. ■

Donnons maintenant une application du théorème 3 aux variables aléatoires bornées, prenant en compte les parties positives et négatives des variances.

Corollaire 3. *Soit* $X_1, X_2, \ldots$ *une suite de v.a. réelles centrées, indépendantes de carré intégrable et bornées p.s. supérieurement par une constante* $M > 0$. *On pose*

$$D_n = \sum_{k=1}^{n} \mathbb{E}(X_k^2) \text{ et } L_n = (MD_n)^{-1} \sum_{k=1}^{n} \mathbb{E}((\max(0, X_k))^3).$$

Avec les notations du théorème 1, pour tout x *positif,*

$$\mathbb{P}(S_n^* \geq \sqrt{2D_n x} + \max(1/4, L_n/3)Mx) \leq \exp(-x).$$

Remarque 4. Le facteur usuel devant x est $M/3$. (a) permet d'obtenir $M/4$ quand les variables X_k n'ont pas une loi très asymétrique. Par exemple, quand $X_k = Z_k - p$ et les variables Z_k sont indépendantes de loi $b(p)$, $L_n/D_n = 1 - p$ et $M = 1 - p$. Donc, pour $p \geq 1/4$,

$$\mathbb{P}(S_n^* \geq \sqrt{2np(1-p)x} + (1-p)(x/4)) \leq \exp(-x).$$

Pour p dans $[0, 1/4]$, le terme linéaire en x est $(1-p)^2(x/3)$.

Preuve. On peut supposer $M = 1$. Soit $C_n = \max(1/4, L_n/3)$. Nous allons montrer d'abord majorer la transformée de Laplace de S_n. Le corollaire 3 découle du théorème 3 via la majoration suivante de la transformée de Laplace de S_n:

$$\log \mathbb{E}(\exp(tS_n)) \leq \frac{D_n t^2}{2(1 - C_n t)} \quad \text{pour tout } t \in [0, 1/C_n[. \tag{9}$$

Pour montrer (9), on note d'abord que

$$\log \mathbb{E}(\exp(tS_n)) = \sum_{k=1}^{n} \log \mathbb{E}(\exp(tX_k)).$$

Comme $\log x \leq x - 1$ et $\mathbb{E}(X_k) = 0$, on en déduit que

$$\log \mathbb{E}(\exp(tS_n)) = \sum_{k=1}^{n} \mathbb{E}(\exp(tX_k) - tX_k - 1).$$

Soit $\gamma(x) = e^x - x - 1$. En dérivant deux fois, $\gamma(x) \leq x^2/2$ pour $x \leq 0$, et en développant en série entière, pour x positif,

$$\gamma(x) \leq \frac{x^2}{2} + \sum_{m \geq 3} \frac{x^m}{m!}.$$

Soit $X_{k+} = \max(0, X_k)$. pour t positif,

$$\mathbb{E}(e^{tX_k} - tX_k - 1) \leq \frac{t^2}{2}\mathbb{E}(X_k^2) + \sum_{m \geq 3} \frac{t^m}{m!}\mathbb{E}X_{k+}^m) \leq \frac{t^2}{2}\mathbb{E}(X_k^2) + \sum_{m \geq 3} \frac{t^m}{m!}\mathbb{E}(X_{k+}^3),$$

car les variables X_{k+} sont p.s. majorées par 1. Donc, en sommant en k,

$$\log \mathbb{E}(\exp(tS_n)) \leq \frac{D_n t^2}{2} + L_n D_n \sum_{m \geq 3} \frac{t^m}{m!} \leq \frac{D_n t^2}{2}\Big(1 + \frac{L_n t}{3(1 - t/4)}\Big).$$

Pour finir, on note que

$$1 + \frac{L_n t}{3(1 - t/4)} \leq 1/(1 - C_n t),$$

ce qui établit (9). On applique alors le théorème 3 pour conclure. ■

6. Application aux maxima de processus empiriques

Nous allons dans un premier temps définir les processus empiriques, puis appliquer les corollaires 1 et 2 pour obtenir des premières inégalités de concentration, et ensuite donner les limites de ces corollaires.

Dans la suite $\xi_1, \xi_2, \ldots$ est une suite de v.a. indépendantes à valeurs dans $\mathbb{R}^d$ muni de la tribu borélienne. La mesure empirique P_n est définie par

$$P_n(f) = \frac{1}{n} \sum_{k=1}^{n} f(\xi_k),$$

pour f fonction mesurable de $\mathbb{R}^d$ dans $\mathbb{R}$. Soit $\mathcal{F}$ une classe dénombrable de fonctions mesurables bornées de $\mathbb{R}^d$ dans $\mathbb{R}$. Pour μ mesure borélienne bornée on pose

$$Z_\mu = n \sup\{P_n(f) - \mu(f) : f \in \mathcal{F}\}.$$

Donnons une première application du corollaire 1.

Théorème 4. *Soit $\mathcal{F}$ une classe dénombrable de fonctions mesurables de $\mathbb{R}^d$ dans $[0, 1]$. Alors, pour tout x positif,*

$$\mathbb{P}(Z_\mu \geq \mathbb{E}(Z_\mu) + x) \leq \exp(-2x^2/n).$$

Preuve. Quitte à passer à la limite, on peut supposer $\mathcal{F}$ finie. On pose

$$Z(x_1, x_2, \ldots, x_n) = \sup\{f(x_1) + f(x_2) + \cdots + f(x_n) - n\mu(f) : f \in \mathcal{F}\}.$$

Soit f réalisant le maximum ci-dessus. Si $y = (y_1, \ldots, y_k)$ et $x = (x_1, \ldots, x_n)$, alors

$$Z(y) \geq \sum_{k=1}^{n} f(y_k) - n\mu(f) \geq Z(x) - \sum_{k=1}^{n} \mathbb{1}_{x_k \neq y_k}.$$

En renversant les rôles de x et y, on obtient

$$|Z(x) - Z(y)| \leq \sum_{k=1}^{n} \mathbb{1}_{x_k \neq y_k}.$$

Par conséquent, Z est 1-lipschitzienne en chaque variable pour la distance de Hamming $d(s, t) = \mathbb{1}_{s \neq t}$. Le corollaire 1 s'applique alors et donne le résultat.

7. Entropies, tensorisation de l'entropie

Pour f v.a.r. strictement positive et intégrable, on définit l'entropie $H(f)$ par

$$H(f) = \mathbb{E}(f \log(f/\mathbb{E}(f))) = \mathbb{E}(f \log f) - \mathbb{E}(f) \log \mathbb{E}(f)$$

l'entropie est une fonctionnelle convexe et positive. De plus $H(f) = 0$ si et seulement si f est p.s. constante. Plus généralement, pour $\mathcal{A}$ sous-tribu de $(\Omega, \mathcal{T}, \mathbb{P})$, nous définirons l'entropie relative $H_{\mathcal{A}}(f)$ par

$$H_{\mathcal{A}}(f) = \mathbb{E}(f \log(f/\mathbb{E}(f \mid \mathcal{A}))) = \mathbb{E}(f \log f) - \mathbb{E}(f \log \mathbb{E}(f \mid \mathcal{A})).$$

De même $H_{\mathcal{A}}(f) \geq 0$ et $H_{\mathcal{A}}(f) = 0$ seulement si f est p.s. égale à une v.a. $\mathcal{A}$-mesurable.

7.1. formule de dualité et formule variationnelle pour l'entropie

Donnons d'abord la formule de dualité pour l'entropie relative.

Lemme 3. *Soit f variable aléatoire p.s. strictement positive. Alors*

$$H_{\mathcal{A}}(f) = \max\{\mathbb{E}(fg) : g \text{ telles que } \mathbb{E}(\exp(g)|\mathcal{A}) = 1\}.$$

De plus le maximum est atteint par la variable aléatoire $g = \log(f/E(f|\mathcal{A}))$.

Preuve. D'abord, pour $g = \log(f/\mathbb{E}(f|\mathcal{A}))$, clairement $\mathbb{E}(fg) = H_{\mathcal{A}}(f)$. Il nous reste à montrer la majoration. Considérons la fonction ϕ définie par $\phi(x) = e^x - 1$. Pour $\lambda > 0$, soit

$$\phi^*(\lambda) = \sup_{x \in \mathbb{R}} (\lambda x - \phi(x))$$

Le maximum x de $\lambda x - \phi(x)$ est atteint quand $\phi'(x) = \lambda$, ce qui équivaut ici à $x = \log \lambda$. Donc

$$\phi^*(\lambda) = \lambda \log \lambda - \lambda + 1.$$

De plus

$$\lambda x \leq \phi^*(\lambda) + \phi(x) \quad \text{pour tout} \quad \lambda > 0 \quad \text{et tout } x \in \mathbb{R}.$$

En appliquant (10),

$$fg \leq \mathbb{E}(f|\mathcal{A})(\phi^*(f/\mathbb{E}(f|\mathcal{A})) + \phi(g)).$$

Comme $E(\phi(g)|\mathcal{A}) = 0$, il en résulte que

$$\mathbb{E}(fg|\mathcal{A}) \leq \mathbb{E}(f|\mathcal{A})\mathbb{E}(\phi^*(f/\mathbb{E}(f|\mathcal{A}))|\mathcal{A}).$$

Par conséquent

$$\mathbb{E}(\phi^*(f/\mathbb{E}(f|\mathcal{A}))|\mathcal{A}) = \mathbb{E}\Big(\frac{f}{\mathbb{E}(f|\mathcal{A})} \log\Big(\frac{f}{\mathbb{E}(f|\mathcal{A})}\Big) \Big| \mathcal{A}\Big),$$

et

$$\mathbb{E}(fg|\mathcal{A}) \leq \mathbb{E}(f \log(f/\mathbb{E}(f|\mathcal{A})) \mid \mathcal{A}),$$

ce qui implique le lemme 3. ■

Donnons maintenant la formule variationnelle.

Lemme 4. *Soit f variable aléatoire p.s. strictement positive. Alors*

$$H_{\mathcal{A}}(f) = \min\{\mathbb{E}(f\log(f/h) - f + h) : h > 0 \text{ p.s. et } \mathcal{A}\text{-mesurable }\}.$$

De plus le minimum est atteint par la variable aléatoire $h = \mathbb{E}(f|\mathcal{A})$.

Preuve. D'abord pour $h = \mathbb{E}(f|\mathcal{A})$, on a

$$\mathbb{E}(f\log(f/h) - f + h) = H_{\mathcal{A}}(f).$$

Il reste à montrer l'inégalité. Soit g une variable aléatoire réelle telle que $\mathbb{E}(\exp(g)|\mathcal{A}) = 1$. Alors

$$\mathbb{E}(fg|\mathcal{A}) = h\mathbb{E}(g(f/h)|\mathcal{A})).$$

L'inégalité de Young assure que

$$g(f/h) \leq e^g + (f/h)\log(f/h) - (f/h),$$

et donc, en prenant l'espérance conditionnelle,

$$\mathbb{E}(g(f/h)|\mathcal{A}) \leq \mathbb{E}((f/h)\log(f/h) - (f/h) + 1|\mathcal{A})$$

Or h est positive et $\mathcal{A}$-mesurable: donc on peut multiplier cette inégalité par h et rentrer h dans l'espérance conditionnelle, ce qui donne

$$\mathbb{E}(fg|\mathcal{A})) \leq \mathbb{E}(f\log(f/h) - f + h|\mathcal{A}).$$

Donc

$$\mathbb{E}(fg) \leq \mathbb{E}(f\log(f/h) - f + h).$$

On obtient alors le lemme 4 à partir du lemme 3 en prenant $g = \log(f/\mathbb{E}(f|\mathcal{A}))$ dans l'inégalité ci-dessus. ■.

7.2. Tensorisation de l'entropie: formule de Ledoux

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(\mathcal{F}_n)_{n\geq 0}$ une filtration croissante sur cet espace probabilisé, telle que $\mathcal{F}_0$ soit égale à la tribu triviale, de telle sorte que E_0 est l'opérateur d'espérance usuel. La formule qui suit permet de décomposer l'entropie suivant la filtration $(\mathcal{F}_n)_{n\geq 0}$.

Lemme 5. *Soit f une variable aléatoire $\mathcal{F}_n$-mesurable, strictement positive et intégrable. Alors*

$$H(f) = \sum_{k=1}^{n} H_{\mathcal{F}_{k-1}}(E_k f).$$

Preuve. Comme $f = E_n(f)$ et $E_0(f) = \mathbb{E}(f)$,

$$\log(f/\mathbb{E}(f)) = \sum_{k=1}^{n} \log(E_k(f)/E_{k-1}(f)).$$

Donc

$$H(f) = \sum_{k=1}^{n} \mathbb{E}(f \log(E_k(f)/E_{k-1}(f))) = \sum_{k=1}^{n} \mathbb{E}(E_k(f) \log(E_k(f)/E_{k-1}(f))),$$

ce qui montre le lemme 5. ■.

Théorème 5. - Tensorisation de l'entropie (Ledoux) - *Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes à valeurs dans un espace polonais $\mathcal{X}$. Soit $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$ et, pour k dans $[1, n]$, $\mathcal{F}_n^k$ la tribu engendrée par $(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Alors, pour toute variable aléatoire réelle f strictement positive, $\mathcal{F}_n$-mesurable et intégrable,*

$$H(f) \leq \sum_{k=1}^{n} H_{\mathcal{F}_n^k}(f).$$

Preuve. Nous allons montrer que, pour toute famille $(f_k)_{k \in [1,n]}$ de variables aléatoires réelles strictement positives et intégrables, respectivement $\mathcal{F}_n^k$-mesurables,

$$H(f) \leq \mathbb{E}(f_1 \phi^*(f/f_1)) + \cdots + \mathbb{E}(f_n \phi^*(f/f_n)),$$

avec $\phi^*(x) = x \log x - x + 1$. Le théorème 5 découle de (11) en prenant $f_k = E_n^k(f)$.

Pour montrer (11), il suffit de montrer que, pour tout k dans $[1, n]$,

$$H_{\mathcal{F}_{k-1}}(E_k f) \leq \mathbb{E}(f_k \phi^*(f/f_k)).$$

Posons $g_k = E_k(f)/E_{k-1}(f)$. Comme g_k est $\mathcal{F}_k$-mesurable,

$$H_{\mathcal{F}_{k-1}}(E_k f) = \mathbb{E}(f g_k) = \mathbb{E}(f_k (g_k f/f_k)) = \mathbb{E}(f_k \mathbb{E}(g_k (f/f_k) \mid \mathcal{F}_n^k)).$$

L'inégalité de Young assure que

$$g_k(f/f_k) \leq \phi^*(f/f_k) + \exp(g_k) - 1.$$

Or f_k est positive. Donc

$$E_0(fg_k) \leq \mathbb{E}(f_k \mathbb{E}(\phi^*(f/f_k) \mid \mathcal{F}_n^k)) + \mathbb{E}(f_k \mathbb{E}(\exp(g_k) - 1 \mid \mathcal{F}_n^k)).$$

Mais, puisque g_k est $\mathcal{F}_k$-mesurable, par indépendance de $X_1, \ldots, X_n$,

$$\mathbb{E}(\exp(g_k) - 1 \mid \mathcal{F}_n^k) = E_{k-1}(\exp(g_k) - 1) = 0.$$

Donc $E_0(fg_k) \leq E_0(f_k \phi^*(f/f_k))$, ce qui montre (11) et donc le théorème 5. ■

8. Une inégalité de Bernstein pour les maxima de processus empiriques

Dans cette partie, nous allons montrer comment le théorème 5 s'applique pour obtenir des majorations de la transformée de Laplace de certaines fonctions Z de v.a. indépendantes. La méthode est la suivante: pour t positif et $f = \exp(tZ)$,

$$H(f) = tF'(t) - F(t) \log F(t).$$

Le théorème 5 donne donc l'inéquation différentielle suivante:

$$tF'(t) - F(t) \log F(t) \leq \sum_{k=1}^{n} \mathbb{E}((tZ - \log E_n^k(e^{tZ}))e^{tZ}).$$

Le point délicat est de majorer au mieux les quantités à droite. C'est possible dans certains cas, comme les maxima de processus empiriques.

Soit $X_1, X_2, \ldots, X_n$ une suite de v.a. indépendantes à valeurs dans $\mathcal{X}$ espace polonais, et $\mathcal{S}$ une classe dénombrable de fonctions mesurables de $\mathcal{X}$ dans $[-1, 1]^n$. On suppose que $\mathbb{E}(s^k(X_k)) = 0$ pour tout $s = (s^1, \ldots, s^n)$ dans $\mathcal{S}$. On pose

$$S_n(s) = \sum_{k=1}^{n} s^k(X_k), \;\; V_n = \sup_{s \in \mathcal{S}} \operatorname{Var} S_n(s) \;\text{ et }\; Z = \sup_{s \in \mathcal{S}} S_n(s).$$

Talagrand (1996) a obtenu le premier des inégalités de type Bennett pour la concentration de Z au-dessus de sa moyenne. Ledoux (1996) en a donné une autre preuve. Le problème est alors d'obtenir les bonnes constantes dans ces inégalités. Le théorème ci-dessous, obtenu par Klein et Rio (2005) donne une inégalité de déviation de Z au-dessus de sa moyenne de type Bernstein avec le meilleur facteur variance connu.

Théorème 6. *Soit L le logarithme de la transformée de Laplace de Z. Soit Alors, pour tout t dans* $[0, 2/3[$,

$$(a) \qquad L(t) \leq t\mathbb{E}(Z) + \frac{(2\mathbb{E}(Z) + V_n)t^2}{2 - 3t}.$$

Par conséquent, si $v = 2\mathbb{E}(Z) + V_n$, pour tout x positif,

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + \sqrt{2vx} + (3x/2)) \leq \exp(-x). \tag{b}$$

Remarque 5. La même inégalité vaut pour $-Z$ (voir Klein et Rio (2005): la preuve est plus difficile), et fournit une inégalité de déviation de Z sous sa moyenne.

Preuve. On peut supposer $\mathcal{S}$ finie. Le cas dénombrable s'obtient par passage à la limite. Le premier pas est d'approcher les entropies relatives à $\mathcal{F}_n^k$ par des quantités plus faciles à majorer. C'est le but de la proposition ci-dessous.

Proposition 2. *Soit f une variable aléatoire strictement positive et $\mathcal{F}_n$-mesurable, telle que $\mathbb{E}(f \log f) < \infty$, et $g_1, g_2, \ldots, g_n$ une suite de variables aléatoires positives, intégrables et telles que $\mathbb{E}(g_k \log g_k) < \infty$. Alors*

$$\mathbb{E}(f \log(f/E_n^k f)) \leq \mathbb{E}(g_k \log(g_k/E_n^k g_k)) + \mathbb{E}((f - g_k) \log(f/E_n^k f)).$$

Preuve de la proposition 2. Posons $f_k = E_n^k f$.

$$\mathbb{E}(f \log(f/f_k)) = \mathbb{E}(g_k \log(f/f_k)) + \mathbb{E}((f - g_k) \log(f/f_k)).$$

Mais $E_n^k(f/f_k) = 1$, et donc

$$\mathbb{E}(g_k \log(f/f_k)) \leq \sup\{\mathbb{E}(g_k h) : \ h \ \mathcal{F}_n\text{-measurable}, \ E_n^k(e^h) = 1\}.$$

Donc, d'après le lemme 3,

$$\mathbb{E}(g_k \log(f/f_k)) \leq \mathbb{E}(g_k \log(g_k/E_n^k g_k)),$$

et la proposition 2 s'ensuit. ■

Pour montrer le théorème 6, on part de la proposition 2. Le point crucial est le choix de g_k. Nous allons choisir g_k en sorte que l'entropie relative de g_k soit facile à majorer et que $f - g_k$ soit positive ou nulle, afin de pouvoir contrôler le second terme. Comme $\mathcal{S}$ est finie, on peut l'écrire en extension, $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$. Nous allons maintenant définir g_k.

Definition 1. Soit τ le plus petit entier tel que $Z = S_n(s_\tau)$. Posons $f = \exp(tZ)$ et $f_k = E_n^k(f)$. Soit P_n^k la probabilité conditionnelle à $\mathcal{F}_n^k$. On pose

$$g_k = \sum_{i=1}^{m} P_n^k(\tau = i) \exp(tS_n(s_i)).$$

Soit F la transformée de Laplace de Z. Par le théorème 5 et la proposition 2,

$$tF'(t) - F(t)\log F(t) \leq \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k/E_n^k g_k)) + \sum_{k=1}^{n} \mathbb{E}((f-g_k)\log(f/f_k)). \tag{11}$$

Comme g_k est un barycentre des variables $\exp(tS_n(s_i))$, il est clair que $g_k \leq f$. Pour majorer le second terme dans (11), nous nous servirons du lemme suivant.

Lemme 6. *Avec les notations ci-dessus* $\log(f/f_k) \leq t$ *p.s.*

Preuve du lemme 6. Let $S_n^k(s) = S_n(s) - s^k(X_k)$. Soit $Z_k = \sup\{S_n^k(s) : s \in \mathcal{S}\}$. Alors Z_k est $\mathcal{F}_n^k$-mesurable et

$$Z \leq S_n^k(s_\tau) + 1 \leq Z_k + 1.$$

Donc

$$f \leq e^t \exp(tZ_k).$$

Soit maintenant τ_k le plus petit entier tel que $Z_k = S_n^k(s_{\tau_k})$. Alors τ_k est $\mathcal{F}_n^k$-mesurable, et donc $E_n^k(s_{\tau_k}^k(X_k)) = 0$ par indépendance des variables aléatoires et par l'hypothèse de centrage. Donc

$$Z_k = E_n^k(S_n(s_{\tau_k})) \leq E_n^k(Z).$$

En appliquant l'inégalité de Jensen conditionnelle avec la fonction convexe $x \to \exp(tx)$, il vient alors:

$$\exp(tZ_k) \leq E_n^k(\exp(tZ)) = f_k.$$

Donc $f \leq e^t f_k$, ce qui implique le lemme 6. ■

Comme $f - g_k \geq 0$, le lemme 6 assure que

$$\mathbb{E}((f-g_k)\log(f/f_k)) \leq t\mathbb{E}(f-g_k). \tag{12}$$

Dans (12), il nous reste à majorer les entropies relatives des fonctions g_k. Soit

$$h_k = \sum_{i=1}^{m} P_n^k(\tau = i)\exp(tS_n^k(s_i)).$$

La v.a. h_k est positive et $\mathcal{F}_n^k$-measurable. Donc, en appliquant le lemme 4,

$$\mathbb{E}(g_k \log(g_k/E_n^k g_k)) \leq \mathbb{E}(g_k \log(g_k/h_k) - g_k + h_k).$$

De (11) et (12), on tire alors

$$tF' - F\log F \leq \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k/h_k) + (1+t)(h_k - g_k)) + t\sum_{k=1}^{n} \mathbb{E}(f - h_k). \tag{13}$$

Pour majorer le dernier terme à droite, nous allons montrer le lemme suivant.

Lemma 7. *Pour les v.a. h_k définies ci-dessus*

$$\sum_{k=1}^{n} \mathbb{E}(f - h_k) \leq tF'(t).$$

Preuve du lemme 7. Comme les variables $S_n^k(s)$ sont $\mathcal{F}_n^k$-mesurables,

$$h_k = E_n^k\Big(\sum_{i=1}^{m} \mathbb{1}_{\tau=i} \exp(tS_n^k(s_i))\Big) = E_n^k\Big(\exp(tS_n^k(s_\tau))\Big).$$

Par conséquent, par convexité de l'exponentielle,

$$\sum_{k=1}^{n} \mathbb{E}(f - h_k) = n(f - n^{-1} \sum_{k=1}^{n} \mathbb{E}\Big(\exp(tS_n^k(s_\tau))\Big) \leq n(f - \mathbb{E}\Big(\exp(tn^{-1} \sum_{k=1}^{n} S_n^k(s_\tau))\Big)$$

et donc

$$\sum_{k=1}^{n} \mathbb{E}(f - h_k) \leq n(F(t) - F(t(n-1)/n)) \leq tF'(t)$$

par convexité de F, ce qui montre le lemme 7. ■

Definition 2. On pose $r(t, x) = x \log x + (1 + t)(1 - x)$.

Avec cette définition

$$g_k \log(g_k/h_k) + (1 + t)(h_k - g_k)) = h_k r(t, g_k/h_k).$$

En appliquant la convexité de r en x avec les poids $P_n^k(\tau = i) \exp(tS_n^k(s_i))$, il vient

$$h_k r(t, g_k/h_k) \leq \sum_i P_n^k(\tau = i) \exp(tS_n^k(s_i)) r(t, \exp(ts_i^k(X_k)),$$

ce qui implique, par indépendance de X_k avec $\mathcal{F}_n^k$, que

$$E_n^k(h_k r(t, g_k/h_k)) \leq \sum_i P_n^k(\tau = i) \exp(tS_n^k(s_i)) \mathbb{E}\Big(r(t, \exp(ts_i^k(X_k))\Big) \tag{14}$$

Pour finir la majoration, on se sert du lemme suivant.

Lemme 8. *Pour toute fonction s dans $\mathcal{S}$ et tout $t > 0$,*

$$\mathbb{E}\, r(t, \exp(ts^k(X_k))) \leq \frac{t^2}{2} \mathbb{E}(s^k(X_k))^2.$$

Proof. Soit $\eta(x) = r(t, e^{tx}) = txe^{tx} + (t+1)(1 - e^{tx})$. Montrons que, pour $x \leq 1$,

$$\eta(x) \leq x\eta'(0) + (tx)^2/2. \tag{15}$$

Soit $\delta(x) = \eta(x) - x\eta'(0) - (tx)^2/2$. Alors $\delta(0) = 0$ et $\delta'(x) = t^2(x-1)(e^{tx} - 1)$. Donc $\delta'(x)$ a le même signe as $x(x-1)$, ce qui implique (15). Comme les variables $s^k(X_k)$ sont centrées, pour $x = s^k(X_k)$, en prenant l'espérance, on obtient le lemme 8. ■

D'après le lemme 8, en partant de (14),

$$E_n^k(h_k r(t, g_k/h_k)) \leq \frac{t^2}{2} E_n^k\Big(\sum_i \mathbb{I}_{\tau=i} \exp(tS_n^k(s_i)) \mathbb{E}(s_i^k(X_k))^2\Big). \tag{15}$$

Or $\exp(tS_n^k(s_i)) \leq \exp(t + tS_n(s_i))$, et donc

$$\sum_{k=1}^n \mathbb{E}(h_k r(t, g_k/h_k)) \leq \frac{t^2 e^t}{2} \mathbb{E}\Big(\sum_i \mathbb{I}_{\tau=i} \exp(tS_n(s_i)) \sum_{k=1}^n \mathbb{E}(s_i^k(X_k))^2\Big).$$

Comme $\sum_k \mathbb{E}(s_i^k(X_k))^2 \leq V_n$, il en résulte que

$$\sum_{k=1}^n \mathbb{E}(h_k r(t, g_k/h_k)) \leq \frac{1}{2} t^2 e^t V_n F(t). \tag{16}$$

Soit $L = \log F$. (13), suivi du lemme 7 et de (16) conduisent à l'inéquation différentielle suivante: pour t dans $]0, 1[$,

$$t(1-t)L' - L \leq t^2 e^t (V_n/2).$$

En divisant par t^2, on obtient:

$$(L/t)' - L' \leq e^t (V_n/2)$$

Comme $L(0) = 0$, la limite en 0 de (L/t) vaut $L'(0) = \mathbb{E}(Z)$. Donc, en intégrant,

$$L(t)(1-t)/t \leq (e^t - 1)(V_n/2) + \mathbb{E}(Z).$$

De plus, une comparaison terme à terme des séries entières montre que

$$(e^t - 1)/2 \leq t/(2-t).$$

Donc

$$L(t) - t\mathbb{E}(Z) \leq \frac{V_n t}{((1-t)(2-t)} + \frac{\mathbb{E}(Z)t^2}{1-t} \ ,$$

ce qui montre (a). (b) découle de (a) via le théorème 3. ■

Remerciements. Je voudrais remercier Bernard Bercu, Jean-François Bony, François Caron et Pierre Del Moral pour leurs questions et commentaires avisés, qui m'ont permis d'améliorer la version initiale de ce cours.

Bibliographie

Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tôkohu Math. J.* **19**, 357-367.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* **57**, 33-45.

Freedman, D.A. (1975). On tail probabilites for martingales. *Annals of Probability* **3**, 100-118.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**, 13-30.

Klein, T. et Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, No 3, 1060-1077.

Mc Diarmid, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics,* 148-188. Cambridge University Press. Cambridge.

Ledoux, M. (1996). On Talagrand deviation inequalities for product measures. *ESAIM: Probability and Statistics* **1**, 63-87.

Merlevède, F., Peligrad, M. et Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Prob. Th. Rel. Fields* **151**, 435-474.

Rio, E. (1994). Local invariance principles and their

Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants. Ed. J.M. Ghidaglia et X. Guyon. Mathématiques et Applications 31. Springer. Heidelberg.

Talagrand, M. (1996). New concentration inequalities in product measures. *Invent. Math.* **126**, 505-563.