

CONCENTRÉ DE CONCENTRATION DE LA MESURE QUELQUES RECETTES SIMPLES POUR OBTENIR DES INÉGALITÉS DE CONCENTRATION

THIERRY KLEIN

TABLE DES MATIÈRES

1. Introduction	1
2. On commence par le dessert : le monde trop beau des variables gaussiennes	2
2.1. Fixons les notations	2
2.2. Une inégalité de concentration parfaite ?	2
2.3. De l'inégalité de Markov à une deuxième inégalité de concentration	3
2.4. Une troisième inégalité gaussienne ?	4
3. Plat de résistance la méthode de Cramér	4
3.1. Fixons le vocabulaire	4
3.2. La méthode	5
4. Trois inégalités Classiques	7
4.1. Variables sous Gaussiennes	7
4.2. Inégalité de Hoeffding	8
4.3. Inégalité de Bennett	9
4.4. Inégalité de Bernstein	10
5. Application aux indices de Sobol	12
5.1. Le contexte boîte noire	12
5.2. L'estimation S par une méthode de Monte Carlo : la méthode pick freeze	13
5.3. Inégalité de concentration pour $S_{N,Cl}^u$	13
Références	14

1. INTRODUCTION

Partons d'un des résultats les plus connus de la théorie des probabilités **LA LOI FORTE DES GRANDS NOMBRES**.

Théorème 1.1. *Soit $(X_n)_{n \geq 1}$ une de variables aléatoires indépendantes et identiquement distribuées admettant une espérance $\mathbb{E}(|X_n|) < +\infty$ alors*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}(X_1).$$

D'un point de vue pratique $\mathbb{E}(X_1)$ représente souvent une quantité inconnue que l'on désire estimer et $\frac{X_1 + \dots + X_n}{n}$ un estimateur naturel. Dans la vraie vie, $n \rightarrow \infty$ n'existe pas, on disposera d'un nombre d'expériences limité ($n = 100$, $n = 1000$). Il est donc naturel de se demander si lorsque n est fixé $\frac{X_1 + \dots + X_n}{n}$ est loin ou non de $\mathbb{E}(X_1)$. Autrement dit on peut se demander si $\frac{X_1 + \dots + X_n}{n}$ est un bon estimateur de $\mathbb{E}(X_1)$. A quelle vitesse converge-t-il ? etc...

La première réponse sur la vitesse de convergence est donnée par le théorème central limite.

Théorème 1.2. Soit $(X_n)_{n \geq 1}$ une de variables aléatoires indépendantes et identiquement distribuées admettant une variance σ^2 (i.e. $\mathbb{E}(X_n^2) < +\infty$) alors

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right) \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(0, \sigma^2).$$

Ce théorème assure grosso-modo que $\frac{X_1 + \dots + X_n}{n}$ converge à une vitesse de l'ordre de $\sqrt{n}$ vers $\mathbb{E}(X_1)$. Cependant, ce n'est qu'un résultat asymptotique et il ne dit rien si n est fixé (surtout si n est petit). L'objet de ce cours est de présenter quelques résultats élémentaires permettant de quantifier l'erreur commise entre $\frac{X_1 + \dots + X_n}{n}$ et $\mathbb{E}(X_1)$ lorsque n est fixé. La théorie de la concentration de la mesure et les inégalités de concentration représentent des domaines très actifs en recherche, nous en présenterons que les premiers résultats. Le lecteur pourra consulter les ouvrages de Michel Ledoux [4] et de Boucheron, Lugosi et Massart [1].

2. ON COMMENCE PAR LE DESSERT : LE MONDE TROP BEAU DES VARIABLES GAUSSIENNES

2.1. Fixons les notations.

Prenons des variables gaussiennes indépendantes et de même loi (on supposera qu'elles sont centrées réduites mais ce n'est pas nécessaire). Leur densité est bien connue elle vaut

$$\phi(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}.$$

On notera $\Phi(x)$ la fonction de répartition

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt.$$

2.2. Une inégalité de concentration parfaite ?

Fixons un n , je rappelle qu'on se demande si $\frac{X_1 + \dots + X_n}{n}$ est loin de $0 = \mathbb{E}(X_1)$. Voici une façon de traduire cette question mathématiquement. Prenons un $x > 0$ (il représente l'écart entre $\frac{X_1 + \dots + X_n}{n}$ et 0) et cherchons à estimer la probabilité que l'écart entre $\frac{X_1 + \dots + X_n}{n}$ et 0 soit supérieur à x

$$C(x) = \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right| \geq x \right)$$

Ici $\mathbb{E}(X_1) = 0$ et $\frac{X_1 + \dots + X_n}{n} \sim \mathcal{N}(0, 1/n)$, ainsi

$$C(x) = \mathbb{P}(|X_1| \geq x\sqrt{n}) = 2\mathbb{P}(X_1 \geq x\sqrt{n}) = \frac{2}{\sqrt{2\pi}} \int_{x\sqrt{n}}^{+\infty} e^{-t^2/2} dt.$$

On est donc ramener à contrôler le plus précisément possible l'intégrale $\int_x^{+\infty} e^{-t^2/2} dt$.

Lemme 2.1. Pour tout $x > 0$

$$(1) \quad \max \left(0, e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3} \right) \right) \leq \int_x^{+\infty} e^{-t^2/2} dt \leq e^{-x^2/2} \frac{1}{x}.$$

Preuve

Majoration

$$\int_x^{+\infty} e^{-t^2/2} dt = \int_x^{+\infty} \frac{t}{t} e^{-t^2/2} dt \leq \frac{1}{x} \int_x^{+\infty} t e^{-t^2/2} dt = \frac{1}{x} e^{-x^2/2}.$$

Minoration On fait une intégration par partie

$$\begin{aligned} \int_x^{+\infty} e^{-t^2/2} dt &= \int_x^{+\infty} \frac{t}{t} e^{-t^2/2} dt = \left[-\frac{e^{-t^2/2}}{t} \right]_x^{+\infty} - \int_x^{+\infty} \frac{1}{t^2} e^{-t^2/2} dt \\ &= \frac{e^{-x^2/2}}{x} - \int_x^{+\infty} \frac{t}{t^3} e^{-t^2/2} dt \geq \frac{e^{-x^2/2}}{x} - \frac{e^{-x^2/2}}{x^3}. \end{aligned}$$

■

Remarque 2.2. L'encadrement (1) peut-être rendu aussi précis que l'on veut en faisant des intégrations par parties successives par exemple la borne sup peut être remplacée par

$$e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} \right).$$

Théorème 2.1 (Concentration Gaussienne). Pour tout $n \geq 1$ et tout $x > 0$

$$(2) \quad 2\max \left(0, e^{-nx^2/2} \left(\frac{1}{\sqrt{n}x} - \frac{1}{n\sqrt{n}x^3} \right) \right) \leq \sqrt{2\pi} \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right| \geq x \right) \leq 2e^{-nx^2/2} \frac{1}{\sqrt{n}x}.$$

Remarque 2.3. On peut remarquer ici qu'on a un contrôle parfait de $C(x)$ puisque la majoration et la minoration sont du même ordre de grandeur.

Cette inégalité parfaite est dû à un coup de chance extraordinaire. Pour les gaussiennes, on sait faire à peu près tous les calculs que l'on veut, ce qui ne sera malheureusement plus le cas par la suite.

2.3. De l'inégalité de Markov à une deuxième inégalité de concentration. L'inégalité de Markov est un outil puissant qui nous permettra d'obtenir des inégalités de concentration.

Proposition 2.4. Soit Y une variable aléatoire positive et intégrable alors pour tout $t > 0$

$$\mathbb{P}(Y \geq t) \leq \frac{1}{t} \mathbb{E}(Y).$$

Si g est une fonction positive strictement croissante alors

$$\mathbb{P}(Y \geq t) = \mathbb{P}(g(Y) \geq g(t)) \leq \frac{\mathbb{E}(g(Y))}{g(t)}.$$

En choisissant $g(x) = (x - \mathbb{E}(Y))^2$ on obtient l'inégalité de Bienaymé-Tchebichev :

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq t) \leq \frac{\text{Var}(Y)}{t^2}.$$

Remarque 2.5. On a obtenu notre première inégalité de concentration.

Revenons maintenant à nos moutons gaussiens

$$C(x) = \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right| \geq x \right) = 2\mathbb{P}(X_1 \geq x\sqrt{n}).$$

Pour $t > 0$ considérons la fonction $g_t(y) = e^{ty}$ cette fonction est visiblement positive et croissante, on peut donc appliquer l'inégalité de Markov

$$(3) \quad C(x) \leq 2e^{-tx\sqrt{n}} \mathbb{E}(e^{tX_1}).$$

Lemme 2.6. Si X est une variable aléatoire gaussienne centrée réduite alors

$$\mathbb{E}(e^{tX_1}) = e^{t^2/2}.$$

Preuve

$$\mathbb{E}(e^{tX_1}) = \int_{\mathbb{R}} e^{tx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = e^{t^2/2} \int_{\mathbb{R}} e^{-(x-t)^2/2} \frac{dx}{\sqrt{2\pi}} = e^{t^2/2}.$$

■

Nous allons maintenant optimiser en $t > 0$ le terme de droite de (3)

$$C(x) \leq 2e^{-tx\sqrt{n}} \mathbb{E}(e^{tX_1}) = 2e^{-tx\sqrt{n}} e^{t^2/2}.$$

L'optimum est atteint en $t = x\sqrt{n}$, on obtient alors

Théorème 2.2 (Concentration Gaussienne 2). Pour tout $n \geq 1$ et tout $x > 0$

$$(4) \quad \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right| \geq x \right) \leq 2e^{-nx^2/2}.$$

Remarque 2.7. Comparons les deux bornes de concentration obtenues en (2) et (4) on a obtenu par la première méthode $2e^{-nx^2/2} \frac{1}{\sqrt{2\pi}\sqrt{nx}}$ et par la seconde $2e^{-nx^2/2}$. On remarque que en utilisant la seconde méthode on perd un facteur de l'ordre $\frac{1}{\sqrt{2\pi}\sqrt{nx}}$ par contre on obtient le “bon” facteur dominant $e^{-nx^2/2}$. La méthode utilisant l'inégalité de Markov a l'énorme avantage d'être “facilement” généralisable à d'autres variables aléatoires. On voit cependant que même dans un cas calculable, on perdra au minimum un facteur de l'ordre $\frac{1}{\sqrt{nx}}$.

2.4. Une troisième inégalité gaussienne ? On peut reprendre la démarche précédente en remplaçant les fonctions $g_t(y) = e^{ty}$ par les fonction $m_k(y) = |y|^k$ ici $k \in \mathbb{N}^*$. L'inégalité de Markov conduit naturellement à

$$(5) \quad C(x) \leq \frac{1}{\sqrt{n^k} x^k} \mathbb{E}(|X_1|^k).$$

Lemme 2.8 (Calcul des moments $M_k = \mathbb{E}(|X_1|^k)$).

(1) Si $k = 2p$ est pair alors

$$M_{2p} = \frac{(2p)!}{2^p p!}.$$

(2) Si $k = 2p + 1$ est impair

$$M_{2p+1} = 2^p p! \frac{2}{\sqrt{2\pi}}.$$

Preuve Notons $M_k = \mathbb{E}(|X_1|^k)$ alors (à l'aide d'une gentille intégration par parties)

$$M_k = 2 \int_0^\infty x^k e^{-x^2/2} dx / \sqrt{2/\pi} = 2(k-1) \int_0^\infty x^{k-2} e^{-x^2/2} dx / \sqrt{2/\pi} = (k-1) M_{k-2}.$$

(1) Si $k = 2p$ alors par récurrence en se rappelant que $M_2 = 1$ il est facile de voir que

$$M_{2p} = \frac{(2p)!}{2^p p!}.$$

(2) Si $k = 2p + 1$ alors par récurrence en se rappelant que $M_1 = 2/\sqrt{2\pi}$ il est facile de voir que

$$M_{2p+1} = 2^p p! \frac{2}{\sqrt{2\pi}}.$$

■

Maintenant il s'agirait d'optimiser (5) en k ce qu'il ne semble pas évident à faire.

3. PLAT DE RÉSISTANCE LA MÉTHODE DE CRAMÉR

3.1. Fixons le vocabulaire. Soit Z une variable aléatoire intégrable et $x > 0$ un nombre positif fixé.

– Inégalité de concentration

On appelle inégalité de concentration toutes inégalités donnant une majoration du type (f étant une fonction explicite)

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq x) \leq f(x).$$

– Inégalité de déviation à gauche

On appelle inégalité de déviation à gauche toutes inégalités donnant une majoration du type (f étant une fonction explicite)

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq x) \leq f(x).$$

– Inégalité de déviation à droite

On appelle inégalité de déviation à droite toutes inégalités donnant une majoration du type (f étant une fonction explicite)

$$\mathbb{P}(\mathbb{E}(Z) - Z \geq x) \leq f(x).$$

– L'objectif.

Comme nous l'avons déjà dit dans l'introduction notre objectif est de quantifier l'écart entre une moyenne empirique et une espérance. Nous allons donc vouloir donner des inégalités de concentration (ou de déviation) pour des variables Z du type $\frac{X_1 + \dots + X_n}{n}$. La fonction f majorante dépendra bien évidemment de n . Par exemple pour des gaussiennes on a vu que l'on pouvait prendre $f(x, n) = 2e^{-nx^2/2}$.

3.2. La méthode. La méthode basée sur l'inégalité de Markov à l'immense avantage de se généraliser. Nous allons maintenant exposer la méthode générale. Prenons une variable aléatoire Y et un nombre positif $t > 0$ alors l'inégalité de Markov assure que

$$(6) \quad \mathbb{P}(Y > x) \leq e^{-tx} \mathbb{E}(e^{tY}).$$

Nous noterons

$$\Psi_Y(t) = \log(\mathbb{E}(e^{tY})).$$

Ainsi (6) se réécrit

$$\mathbb{P}(Y > x) \leq e^{-tx + \Psi_Y(t)}.$$

Notons $\Psi_Y^*(x)$ la transformée de Legendre (Cramer de la loi de Y) de Ψ_Y

$$\Psi_Y^*(x) = \sup_{t \geq 0} \{tx - \Psi_Y(t)\}.$$

On obtient alors l'inégalité

$$\mathbb{P}(Y > x) \leq e^{-\Psi_Y^*(x)}.$$

Bien entendu nous allons appliquer ce processus à des moyennes empiriques de Y iid. Prenons $X_1, X_2, \dots, X_n$ i.i.d pour lesquelles il existe $t > 0$ telque $\mathbb{E}(e^{tX_1}) < +\infty$, on applique la méthode précédente à $Y = \frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1)$. Comme les variables sont iid il est évident que

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \geq x\right) \leq \exp\left(-t(x + \mathbb{E}(X_1) + n\Psi_{X_1}(\frac{t}{n}))\right).$$

Une fois arrivé là, il existe essentiellement deux situations

- (1) Soit la fonction $\Psi_Y^*(x)$ est bien connue (voir Exemple 1) et dans ce cas on obtient de belles inégalités.
- (2) Soit la fonction $\Psi_Y^*(x)$ n'est pas connue (incalculable ou loi de Y inconnue) et dans ce cas on va essayer de majorer (intelligemment de préférence) $e^{-\Psi_Y^*(x)}$.

Exemple 1.

- (1) *Cas des variables gaussiennes. Refaisons le calcul pour des gaussiennes générales $\mathcal{N}(\mu, \sigma^2)$ Soit $(X_i)_i$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ alors*

$$\mathbb{E}(e^{tX_1}) = e^{\frac{t^2\sigma^2}{2} + t\mu}.$$

On a alors

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \geq x\right) \leq \exp\left(-tx + \frac{t^2\sigma^2}{2n}\right).$$

Le terme de droite est optimisé pour $t = \frac{nx}{\sigma^2}$. On obtient alors la borne

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \geq x\right) \leq \exp\left(-\frac{nx^2}{2\sigma^2}\right).$$

On peut faire la même chose pour les déviations à droite et obtenir finalement,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1)\right| \geq x\right) \leq 2 \exp\left(-\frac{nx^2}{2\sigma^2}\right).$$

On peut noter que la borne est d'autant meilleure que σ^2 est petit, nx^2 grand (ce qui est moral).

(2) *Cas des variables de Poisson ($\mathcal{P}(\lambda)$).*

Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes de loi de poisson de paramètre λ .

Commençons par calculer la transformée de Laplace d'une variable $X \sim \mathcal{P}(\lambda)$. Soit $t \in \mathbb{R}$ on a

$$\psi_X(t) = \log(\mathbb{E}(e^{tX})) = \log\left(e^{-\lambda} \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!}\right) = \log(\exp(-\lambda + \lambda e^t)) = -\lambda + \lambda e^t.$$

Comme précédemment nous allons mettre en oeuvre la méthode de Cramér. Pour tout $x > 0$ et $t > 0$ on a

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \geq x\right) \leq \exp\left(-t(x + \lambda) + n\Psi_X\left(\frac{t}{n}\right)\right) = \exp\left(-t(x + \lambda) + n(-\lambda + \lambda e^{t/n})\right).$$

Le terme de gauche est minimal pour $t = n \log\left(\frac{x+\lambda}{\lambda}\right)$, on réinjecte cette valeur dans l'inégalité ci-dessus

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \geq x\right) &\leq \exp\left(-n \log\left(\frac{x+\lambda}{\lambda}\right)(x+\lambda) + n\left(-\lambda + \lambda \frac{x+\lambda}{\lambda}\right)\right) \\ &\leq \exp\left(-n \log\left(\frac{x+\lambda}{\lambda}\right)(x+\lambda) + nx\right) = \exp\left(-n\lambda \left[\left(1 + \frac{x}{\lambda}\right) \log\left(1 + \frac{x}{\lambda}\right) - \frac{x}{\lambda}\right]\right). \end{aligned}$$

Définition 3.1. *On définit pour tout $x \geq -1$ la fonction*

$$h(x) = (1+x) \log(1+x) - x.$$

On a

$$(7) \quad \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \geq x\right) \leq \exp\left(-n\lambda h\left(\frac{x}{\lambda}\right)\right).$$

Regardons maintenant les déviations de l'autre côté, comme les variables de poissons sont positives $\frac{X_1 + \dots + X_n}{n} - \lambda$ est toujours plus grand que $-\lambda$, prenons donc $0 < x < \lambda$ et $t > 0$

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \leq -x\right) &= \mathbb{P}\left(-t\left(\frac{X_1 + \dots + X_n}{n}\right) \geq -t(\lambda - x)\right) = \mathbb{P}\left(e^{-t\left(\frac{X_1 + \dots + X_n}{n}\right)} \geq e^{-t(\lambda - x)}\right) \\ &\leq \exp\left(t(\lambda - x) + n\Psi_X\left(-\frac{t}{n}\right)\right) = \exp\left(t(\lambda - x) + n(-\lambda + \lambda e^{-t/n})\right). \end{aligned}$$

Le terme de gauche est minimal pour $t = -n \log\left(\frac{\lambda-x}{\lambda}\right)$, on réinjecte cette valeur dans l'inégalité ci-dessus

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \leq -x\right) &\leq \exp\left(-n \log\left(\frac{\lambda-x}{\lambda}\right)(\lambda-x) + n\left(-\lambda + \lambda \frac{\lambda-x}{\lambda}\right)\right) \\ &\leq \exp\left(-n \log\left(\frac{\lambda-x}{\lambda}\right)(\lambda-x) - nx\right). \end{aligned}$$

Finalement, on a pour tout $0 < x < \lambda$

$$(8) \quad \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \leq -x\right) \leq \exp\left(-n\lambda h\left(-\frac{x}{\lambda}\right)\right).$$

Remarque 3.1. *On peut montrer que si $x \geq 0$ alors*

$$h(x) \geq \frac{x^2}{2 + 2x/3}.$$

Ainsi la borne de l'équation (7) est contrôlée par

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \lambda \geq x\right) \leq \exp\left(-n\lambda h\left(\frac{x}{\lambda}\right)\right) \leq \exp\left(-n \frac{x^2}{2\lambda + 2x/3}\right) \leq \exp\left(-\frac{nx^2}{2\lambda} \frac{1}{1 + 2x/(3\lambda)}\right).$$

On peut comparer cette borne à la borne gaussienne qui est $\exp\left(-\frac{nx^2}{2\lambda}\right)$.

(3) *Cas des variables Bernoulli ($\mathcal{B}(p)$).*

Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre p .

Commençons par calculer la transformée de Laplace d'une variable $X \sim \mathcal{B}(p)$. Soit $t \in \mathbb{R}$ on a

$$\psi_X(t) = \log(\mathbb{E}(e^{tX})) = \log(pe^t + (1-p)).$$

Comme précédemment nous allons mettre en oeuvre la méthode de Cramér. Comme les variables de Bernoulli sont toujours plus petites que 1 il suffit de considérer des x tels que $0 < x < 1-p$. Donc pour ces x et tout $t > 0$ on a

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - p \geq x\right) \leq \exp\left(-t(x+p) + n\psi_X\left(\frac{t}{n}\right)\right) = \exp\left[-t(x+p) + n \log(pe^t + (1-p))\right].$$

Notons $u = pe^t$ alors le terme de droite est optimisé pour $u = \frac{(1-p)(x+p)}{n-(x+p)}$ (i.e. $t = \log(u/p)$).

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - p \geq x\right) &\leq \exp\left[-\log(u/p)(x+p) + n \log\left(\frac{(1-p)(x+p)}{n-(x+p)} + (1-p)\right)\right] \\ &\leq \exp\left[-(x+p) \log\left(\frac{(1-p)(x+p)}{p(n-(x+p))}\right) + n \log\left(\frac{n(1-p)}{n-(x+p)}\right)\right] \\ &\leq \exp\left[-(n-(x+p)) \log\left(\frac{(n-(x+p))}{(1-p)}\right) - (x+p) \log\left(\frac{x+p}{p}\right) + n \log(n)\right]. \end{aligned}$$

Le fait d'obtenir ce genre de bornes explicites est un coup de chance. En général, on ne sait ni calculer explicitement les fonctions ψ , ni optimiser en t les bornes. On va donc procéder par comparaison, ceci est l'objet du prochain paragraphe.

4. TROIS INÉGALITÉS CLASSIQUES

4.1. Variables sous Gaussiennes.

Définition 4.1. Une variable centrée X est dite sous-gaussienne de facteur variance v si

$$\psi_X(t) := \log(\mathbb{E}(e^{tX})) \leq \frac{vt^2}{2}.$$

On notera par la suite $X \in \mathcal{G}(v)$.

Théorème 4.1 (Concentration sous-gaussienne). Soit $v > 0$, si $X_1, \dots, X_n$ sont i.i.d centrées appartenant à $\mathcal{G}(v)$. Alors pour tout $x > 0$,

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \geq x\right) \leq \exp\left(-\frac{nx^2}{2v}\right).$$

Preuve

On fait la méthode de Cramér, puis on majore ψ_X par son équivalent Gaussien et sa tombe tout seul. ■

Avant de donner des exemples importants de variables sous-gaussiennes, on peut se demander ce qui se passe pour des variables non sous-gaussiennes par exemple des variables exponentielles. Supposons donc que nous disposions de n variables exponentielles de paramètre $\lambda > 0$. Alors si $t < \lambda$ on a pour toute variable $X \sim \mathcal{E}(\lambda)$

$$\psi_X(t) = \log(\mathbb{E}(e^{tX})) = \log\left(\int_0^{+\infty} e^{ty} \lambda e^{-\lambda y} dy\right) = \log\left(\frac{\lambda}{\lambda - t}\right).$$

Ainsi si $t < n\lambda$ on obtient

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \frac{1}{\lambda} \geq x\right) \leq \exp\left(-t\left(x + \frac{1}{\lambda}\right) + n \log\left(\frac{\lambda}{\lambda - t/n}\right)\right).$$

La majoration est optimisée pour $t = \frac{nx\lambda^2}{1+x\lambda}$ (qui est soit dit en passant bien inférieur à $n\lambda$). On réinjecte le tout (on commence à en avoir l'habitude)

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} - \frac{1}{\lambda} \geq x\right) &\leq \exp\left(-\frac{nx\lambda^2}{1+x\lambda}\left(x + \frac{1}{\lambda}\right) + n \log\left(\frac{\lambda}{\lambda - \frac{x\lambda^2}{1+x\lambda}}\right)\right), \\ &\leq \exp(-nx\lambda + n \log(1+x\lambda)), \\ &\leq \exp(-n[x\lambda - \log(1+x\lambda)]). \end{aligned}$$

Bien évidemment cette borne est plus grande que la borne gaussienne mais est ce surprenant ?

4.2. Inégalité de Hoeffding. Les variables bornées forment une classe importante de variables aléatoires sous-gaussiennes. L'inégalité de Hoeffding va permettre de contrôler des déviations de variables bornées sans aucune hypothèse sur la loi (on essaiera de comprendre le prix à payer pour cela).

4.2.1. *Toutes les variables bornées sont sous-gaussiennes.*

Lemme 4.1. *Soit Z une variable aléatoire centrée ($\mathbb{E}(Z) = 0$) prenant ses valeurs dans $[a, b]$ ($a \leq Z \leq b$ presque sûrement). Alors $Z \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right)$.*

Preuve

Etape1 : On commence par remarquer que (la distance de Z au milieu de l'intervalle est plus petite que la longueur de l'intervalle)

$$\left|Z - \frac{b+a}{2}\right| \leq \frac{b-a}{2}.$$

Ainsi

$$\text{Var}(Z) \leq \frac{(b-a)^2}{4}.$$

Etape2 : Calculons gentiment (Y étant bornée sa transformée de Laplace est tellement gentille qu'on peut dériver comme on veut)

$$\begin{aligned} \psi_Z(t) &= \log(\mathbb{E}(e^{tZ})), \\ \psi'_Z(t) &= \frac{\mathbb{E}(Ze^{tZ})}{\mathbb{E}(e^{tZ})}, \\ \psi''_Z(t) &= \frac{\mathbb{E}(Z^2e^{tZ})}{\mathbb{E}(e^{tZ})} - \frac{(\mathbb{E}(Ze^{tZ}))^2}{(\mathbb{E}(e^{tZ}))^2} = \mathbb{E}\left(Z^2e^{tZ}e^{-\psi_Z(t)}\right) - \left(\mathbb{E}\left(Ze^{tZ}e^{-\psi_Z(t)}\right)\right)^2. \end{aligned}$$

Le dernier terme ressemble beaucoup à une variance, remarquons $\mathbb{E}(e^{tZ}e^{-\psi_Z(t)}) = 1$ ainsi la fonction $u \mapsto e^{tu}e^{-\psi_Z(t)}$ est une densité de probabilité par rapport à la loi de Z , ainsi si U est une variable aléatoire admettant cette densité par rapport à la loi¹ de Z on a

$$\text{Var}(U) = \mathbb{E}\left(Z^2e^{tZ}e^{-\psi_Z(t)}\right) - \left(\mathbb{E}\left(Ze^{tZ}e^{-\psi_Z(t)}\right)\right)^2.$$

Comme $a \leq Z \leq b$ il en va de même pour U , le résultat de l'étape 1 reste valable pour U on en déduit donc que

$$\psi''_Z(t) \leq \frac{(b-a)^2}{4}.$$

Maintenant il est facile de voir que $\psi_Z(0) = 0$ ($\log(1) = 0$) et que $\psi'_Z(0) = 0$ (la variable Z est centrée), on fait un développement de Taylor à l'ordre 2

$$\psi_Z(t) = \frac{t^2}{2}\psi''_Z(\theta) \leq \frac{t^2(b-a)^2}{8}.$$

■

Remarque 4.2. Prenons $b = -a = 1$ et des variables aléatoires de Rademacher i.e. $\mathbb{P}(Z = -1) = \mathbb{P}(Z = 1) = \frac{1}{2}$ alors $\text{Var}(Z) = 1 = \frac{(b-a)^2}{4}$. Donc la borne contrôlant la variance est atteinte pour les variables aléatoires de Rademacher qui sont les pires variables bornées!!!
Finalement, la technique consiste à comparer nos variables aux pires variables possible.

1. C'est un changement de probabilité classique que l'on retrouve en théorie des grandes déviations

4.2.2. *Inégalité de Hoeffding.* En concaténant les deux paragraphes précédent on obtient sans trop se fatiguer

Théorème 4.2. Soit $X_1, \dots, X_n$ n variables aléatoires indépendantes (non nécessairement de même loi). On suppose que pour tout indice i , $X_i \in [a_i, b_i]$. Soit

$$S = \sum_{i=1}^n (X_i - \mathbb{E}(X_i)).$$

Alors pour $t > 0$,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Remarque 4.3. Si on s'intéresse comme d'habitude aux moyennes empirique il faut diviser S par n ce qui revient à multiplier t par n et on obtient

$$\mathbb{P}\left(\frac{S}{n} \geq t\right) \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Attention, ici il n'y a pas de miracle ! on ne fait pas mieux que de la concentration gaussienne le n^2 est trompeur car moralement il y a un n au dénominateur à cause de la somme.

Remarque 4.4. On voit que la borne obtenue ne dépend absolument pas de la variance des variables aléatoires considérées (seulement d'un majorant universel de cette variance). Ainsi on obtient la même borne de concentration pour deux variables ayant le même support borné et ceci même si une a une variance cent fois inférieure à l'autre. C'est cette constatation (non raisonnable) qui motive l'introduction des inégalités de Bennett et Bernstein.

4.3. Inégalité de Bennett.

Théorème 4.3 (Inégalité de Bennett). Soit $X_1, \dots, X_n$ n variables aléatoires indépendantes (non nécessairement de même loi), de variances finies et tel que pour tout indice i , $X_i \leq b$. Soit

$$S = \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$$

et

$$v = \sum_{i=1}^n \mathbb{E}(X_i^2).$$

Notons pour $u \in \mathbb{R}$, $\phi(u) = e^u - u - 1$ et pour $u \geq -1$, $h(u) = (1+u) \log(1+u) - u$ (on l'a déjà vue celle-ci !!) Alors

(1) Pour $t > 0$

$$\psi_S(t) := \log(\mathbb{E}(e^{tS})) \leq n \log\left(1 + \frac{v}{nb^2 \phi(bt)}\right) \leq \frac{v}{b^2} \phi(bt).$$

(2) Pour $x > 0$,

$$\mathbb{P}(S \geq x) \leq \exp\left(-\frac{v}{b^2} h\left(\frac{bx}{v}\right)\right).$$

Preuve

(1)

Etape1 : Quitte à remplacer chaque variable X_i par X_i/b on peut supposer que $b = 1$ (les formules sont bien homogènes en b).

Etape2 : On commence par remarquer² que $u \mapsto \frac{\phi(u)}{u^2}$ est une fonction croissante sur $\mathbb{R}$. Alors comme $X_i \leq 1$ il est évident que

$$\begin{aligned} \phi(tX_i) &\leq t^2 X_i^2 \phi(t) = X_i^2 (e^t - t - 1), \\ e^{tX_i} &\leq tX_i + 1 + X_i^2 (e^t - t - 1). \end{aligned}$$

2. Faire une gentille étude de fonction

Etape3 : On calcule tranquillement $\psi_S(t)$ et on utilise l'étape 2.

$$\begin{aligned}\psi_S(t) &= \sum_{i=1}^n \log \left(\mathbb{E} \left[e^{t(X_i - \mathbb{E}(X_i))} \right] \right) = \sum_{i=1}^n (\log (\mathbb{E} [e^{tX_i}]) - t\mathbb{E}(X_i)) \\ &\leq \sum_{i=1}^n (\log (1 + t\mathbb{E}(X_i) + \mathbb{E}(X_i^2) (e^t - t - 1)) - t\mathbb{E}(X_i)).\end{aligned}$$

On multiplie et divise par n et on utilise la concavité de la fonction $u \mapsto \log(1 + u)$

$$\begin{aligned}\psi_S(t) &\leq n \left(\log \left(1 + t \frac{\sum_{i=1}^n \mathbb{E}(X_i)}{n} + \frac{v}{n} (e^t - t - 1) \right) - t \frac{\sum_{i=1}^n \mathbb{E}(X_i)}{n} \right) \\ &\leq v (e^t - t - 1).\end{aligned}$$

Ce qui prouve le premier point.

(2) On met en route la machinerie de Cramér

$$\mathbb{P}(S \geq x) \leq e^{-tx + \psi_S(t)} \leq e^{-tx + v(e^t - t - 1)}.$$

Le terme de droite est optimal pour $t = \log(1 + \frac{x}{v})$, on obtient alors

$$\begin{aligned}\mathbb{P}(S \geq x) &\leq e^{-\log(1 + \frac{x}{v})x + v(\frac{x}{v} - \log(1 + \frac{x}{v}))} \\ &\leq e^{-v[(1 + \frac{x}{v})\log(1 + \frac{x}{v}) - \frac{x}{v}]}.\end{aligned}$$

Ce qui est exactement le deuxième point du Théorème. ■

Remarque 4.5. Il est facile de voir que

$$h(u) \geq \frac{u^2}{2(1 + u/3)}.$$

On peut donc affaiblir le résultat ci-dessus par

$$\mathbb{P}(S \geq x) \leq e^{-\frac{x^2}{2(v + bx/3)}}.$$

Ce résultat est moins précis que celui du théorème alors pourquoi le présenter ?

4.4. Inégalité de Bernstein. Le gros défaut de l'inégalité précédente et qu'elle suppose les variables bornées par b , l'inégalité qui suit va permettre de relâcher cette hypothèse. En contre partie, on va être obligé de faire des hypothèses de moments et on va perdre un peu au niveau de la borne.

Théorème 4.4 (Inégalité de Bernstein). *Soit $X_1, \dots, X_n$ n variables aléatoires indépendantes (non nécessairement de même loi), pour lesquelles il existe de nombres positifs v et c vérifiant $\sum_{i=1}^n \mathbb{E}(X_i^2) \leq v$ et*

$$\sum_{i=1}^n \mathbb{E}[(X_i)_+^q] \leq \frac{q!}{2} v c^{q-2}, \quad \forall q \geq 3,$$

avec $x_+ = \max(x, 0)$. Soit

$$S = \sum_{i=1}^n (X_i - \mathbb{E}(X_i)).$$

Notons pour $u > 0$, $h_1(u) = 1 + u - \sqrt{1 + 2u}$ Alors

(1) Pour $0 < t < 1/c$

$$\psi_S(t) := \log(\mathbb{E}(e^{tS})) \leq \frac{vt^2}{2(1 - ct)}$$

(2) pour $x > 0$,

$$\mathbb{P}(S \geq \sqrt{2vx} + cx) \leq \exp(-x).$$

Preuve

- (1) Comme précédemment en considère la fonction $\phi(u) = e^u - u - 1$. Cette fonction est contrôlée par $u^2/2$ pour les $u \leq 0$

$$\phi(u) \leq \frac{u^2}{2} \quad \forall u \leq 0.$$

Soit $t > 0$ alors

$$\begin{aligned} \phi(tX_i) &= \sum_{q=2}^{+\infty} \frac{t^q X_i^q}{q!} \leq \frac{t^2 X_i^2}{2} + \sum_{q=3}^{+\infty} \frac{t^q (X_i)_+^q}{q!} \\ \mathbb{E}(\phi(tX_i)) &\leq \frac{t^2 \mathbb{E}(X_i^2)}{2} + \sum_{q=3}^{+\infty} \frac{t^q \mathbb{E}[(X_i)_+^q]}{q!} \\ \sum_{i=1}^n \mathbb{E}(\phi(tX_i)) &\leq \frac{t^2 v}{2} + \sum_{q=3}^{+\infty} \frac{t^q v c^{q-2}}{2} \\ \sum_{i=1}^n \mathbb{E}(\phi(tX_i)) &\leq \frac{v}{2} \sum_{q=2}^{+\infty} t^q c^{q-2}. \end{aligned}$$

On voit que la série est convergente si et seulement si $tc < 1$ c'est à dire $t \leq 1/c$. On va maintenant contrôler $\psi_S(t)$ (on utilisera une fois de plus que $\log(u) \leq u - 1$ pour $u > 0$).

$$\begin{aligned} \psi_S(t) &= \log[\mathbb{E}(e^{tS})] = \sum_{i=1}^n \left(\log \left[\mathbb{E} \left(e^{t \sum_{i=1}^n X_i} \right) \right] \right) - t \sum_{i=1}^n \mathbb{E}(X_i) \\ &\leq \sum_{i=1}^n \left(\mathbb{E} \left(e^{t \sum_{i=1}^n X_i} \right) - 1 - t \mathbb{E}(X_i) \right) = \sum_{i=1}^n \mathbb{E}(\phi(tX_i)) \\ &\leq \frac{v}{2} \sum_{q=2}^{+\infty} t^q c^{q-2} \\ &\leq \frac{vt^2}{2} \frac{1}{1-tc}. \end{aligned}$$

Ce qui prouve le premier point.

- (2) Soit $y > 0$ et $0 < t < 1/c$

$$(9) \quad \mathbb{P}(S \geq y) \leq e^{-ty + \psi_S(y)} \leq e^{-ty + \frac{vt^2}{2} \frac{1}{1-tc}}.$$

Posons $h(t) = ty - \frac{vt^2}{2} \frac{1}{1-tc}$ il s'agit d'optimiser en $t \in [0, 1/c]$ cette expression. Ce n'est pas difficile, mais il faut bien s'y prendre. Posons $u = \frac{cy}{v}$ (ie $y = \frac{uv}{c}$) La fonction h se réécrit simplement

$$h(t) = \frac{uvt}{c} - \frac{vt^2}{2(1-ct)}.$$

On divise par v et on cherche les zéros de la dérivé par rapport à t , on tombe sur l'équation

$$\frac{u}{c} - \frac{t}{1-ct} - \frac{ct^2}{2(1-ct)^2} = 0.$$

On multiplie tout le monde par $2(1-ct)^2$ pour obtenir après simplification

$$t^2 - \frac{2t}{c} + \frac{2u}{c^2(1+2u)} = 0.$$

Il est facile de voir que la racine du discriminant vaut

$$\sqrt{\Delta} = \frac{2}{c\sqrt{1+2u}}.$$

Ainsi la seule racine inférieure à $1/c$ est

$$t^* = \frac{1}{c} \left(1 - \frac{1}{\sqrt{1+2u}} \right).$$

On réinjecte dans h et on trouve

$$\begin{aligned} h(t^*) &= v \left[\frac{u}{c^2} \left(1 - \frac{1}{\sqrt{1+2u}} \right) - \frac{\sqrt{1+2u} \left(1 - \frac{1}{\sqrt{1+2u}} \right)^2}{2c^2} \right] \\ &= \frac{v}{c^2} h_1(u). \end{aligned}$$

La fonction h_1 est définie pour $u > 0$ par

$$h_1(u) = 1 + u - \sqrt{1+2u}.$$

On réinjecte tout ça dans l'équation (9) pour obtenir

$$\mathbb{P}(S \geq y) \leq \exp \left(-\frac{v}{c^2} h_1 \left(\frac{cy}{v} \right) \right).$$

L'inégalité

$$\mathbb{P}(S \geq \sqrt{2vx} + cx) \leq \exp(-x)$$

s'obtient en inversant la fonction $h_1(u)$. Son inverse est simplement $h_1^{-1}(t) = t + \sqrt{t}$. ■

Remarque 4.6. (1) On trouve parfois l'inégalité de Bernstein sous la forme équivalente

$$\mathbb{P}(S \geq y) \leq \exp \left(-\frac{v}{c^2} h_1 \left(\frac{cy}{v} \right) \right).$$

(2) On peut voir que pour $u > 0$,

$$h_1(u) \leq \frac{u^2}{2(1+u)}.$$

Sous les mêmes hypothèses que le théorème précédent on obtient alors

$$\mathbb{P}(S \geq y) \leq \exp \left(-\frac{t^2}{2(v+ct)} \right).$$
■

5. APPLICATION AUX INDICES DE SOBOL

Les résultats de cette partie sont issus du travail de Gamboa, Janon, Klein, Lagnoux et Prieur [2].

5.1. Le contexte boîte noire. On considère un modèle de régression de la forme

$$(10) \quad Y = f(X) := f(X_1, \dots, X_p).$$

Ici $Y \in \mathbb{R}$ et les X_i pour $i = 1, \dots, p$, sont des objets aléatoires indépendants (variables vecteurs, processus, tout ce qu'on veut). On notera $\mathcal{X}_i$ l'espace dans lequel vit X_i . De plus f est supposée être une fonction déterministe et Y être de carré intégrable non presque sûrement constante ($\text{Var} Y \neq 0$). Un des objectifs de l'analyse de sensibilité est de déterminer les entrées qui ont de l'influence. Les indices de Sobol permettent de quantifier cette influence.

Nous rappelons (sans rentrer dans les détails) leurs définitions. Donnons nous un sous ensembles³ de $I_p := \{1, \dots, p\} : \mathbf{u}$. L'indice de Sobol fermé (voir [5]) est défini par

$$S_{\text{Cl}}^{\mathbf{u}} := \frac{\text{Var}(\mathbb{E}(Y|X_i, i \in \mathbf{u}))}{\text{Var}(Y)}.$$

3. Prenons par exemple $p = 4$ $\mathbf{u} := (\{2, 3\})$

5.2. L'estimation S par une méthode de Monte Carlo : la méthode pick freeze. En pratique la fonction f n'est pas connue et on ne dispose donc pas de valeurs pour les indices de Sobol, il est donc naturel de vouloir les estimer. Pour X et un sous-ensemble $\mathbf{v}$ de I_p on pose $X^{\mathbf{v}}$ le vecteur tel que $X_i^{\mathbf{v}} = X_i$ if $i \in \mathbf{v}$ et $X_i^{\mathbf{v}} = X'_i$ si $i \notin \mathbf{v}$ où X'_i est une copie indépendante de X_i . On pose alors

$$Y^{\mathbf{v}} := f(X^{\mathbf{v}}).$$

Le lemme suivant permet de réécrire l'espérance de la variance conditionnelle en terme de covariance (voir [3, Lemma 1.2]).

Lemma 5.1. *Pour tout $\mathbf{u} \subset \mathbf{I}_p$, on a*

$$(11) \quad \text{Var}(\mathbb{E}(Y|X_i, i \in \mathbf{u})) = \text{Cov}(\mathbf{Y}, \mathbf{Y}^{\mathbf{u}}).$$

Cette représentation conduit à l'estimateur de type moment

$$(12) \quad S_{N,\text{Cl}}^{\mathbf{u}} = \frac{\frac{1}{N} \sum Y_i Y_i^{\mathbf{u}} - \left(\frac{1}{N} \sum Y_i\right) \left(\frac{1}{N} \sum Y_i^{\mathbf{u}}\right)}{\frac{1}{N} \sum Y_i^2 - \left(\frac{1}{N} \sum Y_i\right)^2}.$$

Remarque 5.2. *Il existe un estimateur du même style un peu meilleur présenté dans [3]. Nous ne le présenterons pas ici car la méthode pour obtenir les inégalités de concentration est la même que celle que l'on va présenter (les calculs sont juste un peu plus lourds).*

Il est clair que cet estimateur converge presque sûrement vers l'indice de Sobol. Comme dans le monde réel, on ne peut faire qu'un nombre fini d'expérience, il est raisonnable de vouloir quantifier la distance entre l'estimateur et le vrai indice de Sobol.

Les différences avec les paragraphes précédents sont les suivantes

- (1) L'estimateur considéré n'est pas une somme de variables indépendantes.
- (2) L'estimateur est biaisé.

Notation

On notera $V := \text{Var}(Y)$. On rappelle que la fonction h est définie pour tout $x > -1$ par

$$h(x) = (1+x) \ln(1+x) - x.$$

5.3. Inégalité de concentration pour $S_{N,\text{Cl}}^{\mathbf{u}}$. On introduit les variables aléatoires

$$U_i^{\pm} = Y_i Y_i^{\mathbf{u}} - (S_{\text{Cl}}^{\mathbf{u}} \pm y)(Y_i)^2 \text{ et } J_i^{\pm} = (S_{\text{Cl}}^{\mathbf{u}} \pm y)Y_i - Y_i^{\mathbf{u}}$$

on note V_U^+ (resp. V_U^- , V_J^+ and V_J^-) le moment d'ordre 2 des variables U_i^+ (resp. U_i^- , J_i^+ et J_i^-).

Théorème 5.1. *Soit $b > 0$ et $y > 0$. On suppose que toutes les variables Y_i et $Y_i^{\mathbf{u}}$ appartiennent à $[-b, b]$. Alors*

$$(13) \quad \mathbb{P}(S_{N,\text{Cl}}^{\mathbf{u}} \geq S_{\text{Cl}}^{\mathbf{u}} + y) \leq M_1 + 2M_2 + 2M_3,$$

$$(14) \quad \mathbb{P}(S_{N,\text{Cl}}^{\mathbf{u}} \leq S_{\text{Cl}}^{\mathbf{u}} - y) \leq M_4 + 2M_2 + 2M_5,$$

où

$$\begin{aligned} M_1 &= \exp \left\{ -\frac{NV_U^+}{b_U^2} h \left(\frac{b_U}{V_U^+} \frac{yV}{2} \right) \right\} & M_3 &= \exp \left\{ -\frac{NV_J^+ b^2}{b_U^2} h \left(\frac{b_U}{bV_J^+} \sqrt{\frac{yV}{2}} \right) \right\} \\ M_2 &= \exp \left\{ -\frac{NV}{b^2} h \left(\frac{b}{V} \sqrt{\frac{yV}{2}} \right) \right\} & M_4 &= \exp \left\{ -\frac{NV_U^-}{b_U^2} h \left(\frac{b_U}{V_U^-} \frac{yV}{2} \right) \right\} \\ M_5 &= \exp \left\{ -\frac{NV_J^- b^2}{b_U^2} h \left(\frac{b_U}{bV_J^-} \sqrt{\frac{yV}{2}} \right) \right\} \end{aligned}$$

et $b_U = b^2(1 + S_{\text{Cl}}^{\mathbf{u}} + y)$.

Démonstration. On commence par remarquer que $S_{\text{Cl}}^{\mathbf{u}}$ et $S_{N,\text{Cl}}^{\mathbf{u}}$ sont des quantités invariantes lorsque l'on translate les variables Y et $Y^{\mathbf{u}}$ par des constantes, on peut donc supposer (sans pertes de généralité) que Y est centrée.

(1) Il est clair que U_i^+ et U_i^- sont majorées par b_U , J_i^+ et J_i^- par b_U/b , de plus

$$\begin{aligned}\mathbb{E}(U_i^+) &= -yV & \mathbb{E}(J_i^+) &= 0 \\ \mathbb{E}(U_i^-) &= yV & \mathbb{E}(J_i^-) &= 0\end{aligned}$$

et

$$V_U^\pm = \text{Var}(YY^{\mathbf{u}}) + (S_{\text{Cl}}^{\mathbf{u}} + y)^2 \text{Var}(Y^2) - 2(S_{\text{Cl}}^{\mathbf{u}} \pm y) \text{Cov}(YY^{\mathbf{u}}, Y^2) + y^2 V^2$$

$$V_J^\pm = ((S_{\text{Cl}}^{\mathbf{u}} \pm y)^2 + 1)V - 2(S_{\text{Cl}}^{\mathbf{u}} \pm y)C_u.$$

(2) Preuve de (13). Comme

$$\{a + b \geq c\} \subset \{a \geq c/2\} \cup \{b \geq c/2\} \quad \text{et} \quad \{ab \geq c\} \subset \{|a| \geq \sqrt{c}\} \cup \{|b| \geq \sqrt{c}\}$$

on a

$$\begin{aligned}\mathbb{P}(S_{N,\text{Cl}}^{\mathbf{u}} \geq S_{\text{Cl}}^{\mathbf{u}} + y) &= \mathbb{P}\left(\frac{\frac{1}{N} \sum_{i=1}^N Y_i Y_i^{\mathbf{u}} - \bar{Y}_N \bar{Y}_N^{\mathbf{u}}}{\frac{1}{N} \sum_{i=1}^N (Y_i)^2 - (\bar{Y}_N)^2} \geq S_{\text{Cl}}^{\mathbf{u}} + y\right) \\ &= \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^N (U_i^+ - \mathbb{E}(U^+)) + \bar{Y}_N \bar{J}_N^+ \geq yV\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^N (U_i^+ - \mathbb{E}(U^+)) \geq N \frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^N Y_i \geq N \sqrt{\frac{yV}{2}}\right) \\ &\quad + \mathbb{P}\left(\sum_{i=1}^N (-Y_i) \geq N \sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^N J_i^+ \geq N \sqrt{\frac{yV}{2}}\right) \\ &\quad + \mathbb{P}\left(\sum_{i=1}^N (-J_i^+) \geq N \sqrt{\frac{yV}{2}}\right).\end{aligned}$$

L'inégalité (13) est alors une quintuple application directe de l'inégalité de Bennett.

(3) Preuve (14). De la même manière on a

$$\begin{aligned}\mathbb{P}(S_{N,\text{Cl}}^{\mathbf{u}} \leq S_{\text{Cl}}^{\mathbf{u}} - y) &= \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^N (-U_i^- + \mathbb{E}(U^-)) + (-\bar{Y}_N) \bar{J}_N^- \geq yV\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^N (-U_i^- + \mathbb{E}(U^-)) \geq N \frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^N Y_i \geq N \sqrt{\frac{yV}{2}}\right) \\ &\quad + \mathbb{P}\left(\sum_{i=1}^N (-Y_i) \geq N \sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^N J_i^- \geq N \sqrt{\frac{yV}{2}}\right) \\ &\quad + \mathbb{P}\left(\sum_{i=1}^N (-J_i^-) \geq N \sqrt{\frac{yV}{2}}\right).\end{aligned}$$

L'inégalité (14) est alors une quintuple application directe de l'inégalité de Bennett. $\square$

RÉFÉRENCES

- [1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities : A Non-asymptotic Theory of Independence*. OUP Oxford, 2013.
- [2] Fabrice Gamboa, Alexandre Janon, Thierry Klein, Agnès Lagnoux-Renaudie, and Clémentine Prieur. Statistical inference for Sobol pick freeze monte carlo method. Preprint available at <http://hal.inria.fr/hal-00804668/en>, 2013.
- [3] Alexandre Janon, Thierry Klein, Agnès Lagnoux, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two Sobol index estimators.
- [4] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

- [5] A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000.

E-mail address: `thierry.klein@math.univ-toulouse.fr`