Chapitre 7

Principes de base et inégalités exponentielles classiques

Après la brève introduction théorique et historique du chapitre précédent, nous allons nous attarder sur des exemples classiques. A ce titre, nous débuterons par les inégalités de concentration les plus simples : celles portant sur la somme de variables aléatoires indépendantes. Dans ce qui suit, Z désignera une variable aléatoire intégrable :

$$Z \,:\, (\Omega, \mathcal{A}, \mathbb{P}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$$

et nous souhaitons obtenir une majoration pertinente des probabilités suivantes :

$$\mathbb{P}\big(Z - \mathbb{E}[Z] \geq t\big) \leq \theta_1(t) \quad \text{et} \quad \mathbb{P}\big(Z - \mathbb{E}[Z] \leq -t\big) \leq \theta_2(t) \quad \text{pour tout } t \geq 0$$

avec $t \mapsto \theta_i(t) \,:\, \mathbb{R}_+ \to \mathbb{R}_+$, pour $i = 1, 2$, des fonctions positives qui tendent vers 0 à l'infini.

Comme nous l'avons mentionné dans l'introduction, les fonctions $(\theta_i)_{i=1,2}$ sont typiquement des fonctions de la forme $t \mapsto Ce^{-ct^p}$ où $p > 0$ avec $c, C > 0$ désignent des constantes numériques. Les inégalités précédentes sont désignées sous l'appelation *déviation (par rapport à la moyenne) à droite (resp. à gauche) de* Z et, lorsque de telles inégalités sont établies, elles permettent d'en déduire une inégalité de concentration qui prend la forme suivante

$$\mathbb{P}\big(|Z - \mathbb{E}[Z]| \geq t\big) \leq \max\big(\theta_1(t), \theta_2(t)\big) \quad \text{pour tout } t \geq 0.$$

7.1 Inégalité de Markov et ses conséquences

L'inégalité de Markov fait partie des inégalités de déviation les plus célèbres et ses conséquences sont des points essentiels de la théorie.

Proposition 7.1.1. *Soit Y une variable aléatoire réelle et positive. Alors, pour tout $t > 0$, nous avons*

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

Démonstration. Soit $t > 0$, il suffit d'observer que l'inégalité suivante est satisfaite de manière évidente :

$$1_{\{Y \geq t\}} \leq \frac{Y}{t} 1_{\{Y \geq t\}}$$

puis d'intégrer celle-ci par rapport à $d\mathbb{P}$. □

Remarque. 1. Le procédé décrit plus haut se généralise aisément. En effet, considérons

$$\phi : I \to \mathbb{R}$$

une fonction positive, croissante, définie sur un sous ensemble $I \subset \mathbb{R}$; si $Y \in I$ p.s. alors nous avons

$$\mathbb{P}(Y \geq t) \leq \frac{1}{\phi(t)} \mathbb{E}\big[\phi(Y)\big] \quad \text{pour tout} \quad t \in I.$$

Bien entendu, cette majoration est pertinente uniquement si le membre de droite (de l'inégalité précédente) est inférieur ou égal à 1. En général, il est coutume de choisir $\phi(t) = t^q$ avec $q > 1$ ou $\phi(t) = e^{\lambda t}$ avec $\lambda > 0$.

2. Bien entendu, les inégalités de concentration sont étroitement liées aux moments d'une variable aléatoire. En effet, les résultats suivants sont satisfaits :

   - S'il existe $q > 0$ et $C > 0$ tel que

   $$\mathbb{P}(|Z| \geq t) \leq \frac{C}{t^q} \quad \text{pour tout} \quad t > 0$$

   alors $\mathbb{E}[|Z|^p]$ pour tout $0 < p < q$.

   - S'il existe $q > 0$ tel que $\mathbb{E}[|Z|^q] < +\infty$ alors

   $$\mathbb{P}(|Z| \geq t) \leq \frac{\mathbb{E}[|Z|^q]}{t^q} \quad \text{pour tout} \quad t > 0.$$

   - De manière similaire, les assertions ci-dessous sont équivalentes.
     (a) Il existe des constantes $(C, c, p) \in \mathbb{R}_+^*$ telles que $\mathbb{P}(Z \geq t) \leq Ce^{-ct^p}$ avec $t \geq 0$.

     (b) Il existe $c' > 0$ telle que $\mathbb{E}[e^{c' Z^p}] < +\infty$.

Voici quelques exemples classiques d'inégalités de concentrations obtenues à partir des conséquences de l'inégalité de Markov 7.1.1.

Exemple 7.1.1 (Bienaymé-Tchebychev)**.** Si $\phi(t) = t^2$ et $Y = \big|Z - \mathbb{E}[Z]\big|$, nous obtenons

$$\mathbb{P}\Big(\big|Z - \mathbb{E}[Z]\big| \geq t\Big) \leq \frac{1}{t^2} \mathrm{Var}(Z) \quad \text{pour tout} \quad t > 0.$$

Cette inégalité de déviation permet d'obtenir une quantification, non-asymptotique, du théorème de la limite centrale. En effet, supposons que $Z = \sum_{i=1}^n X_i$ avec $(X_i)_{i=1,\ldots,n}$ une suite de variables aléatoires i.i.d. de carré intégrable, alors nous avons

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i - \mathbb{E}[\sum_{i=1}^n X_i]\Big| \geq t\Big) \leq \frac{n}{t^2}\mathrm{Var}(X_1) \quad \text{pour tout} \quad t > 0.$$

Le choix $t = u\sqrt{n}$ illustre alors le théorème de la limite centrale.

Exemple 7.1.2 (Chernoff)**.** Si $\phi(t) = e^{\lambda t}$ (avec $\lambda > 0$) et $Z = \sum_{i=1}^n X_i$ avec $(X_i)_{i=1,\ldots,n}$ une suite de variables aléatoires i.i.d. (centrées) admettant des moments de tout ordre, nous obtenons, pour tout $t \geq 0$,

$$\begin{aligned} \mathbb{P}(Z \geq t) &\leq e^{-\lambda t}\prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] \\ &= e^{-\lambda t}\mathbb{E}[e^{\lambda X_1}]^n \\ &= e^{-\lambda t + n\Lambda_{X_1}(\lambda)} \end{aligned}$$

où, rappelons-le, $\Lambda_{X_1}(\lambda) = \log \mathbb{E}[e^{\lambda X_1}]$ avec $\lambda \in \mathbb{R}$ désigne la log-Laplace de la variable X_1. Le lecteur attentif aura remarqué que la majoration précédente est proche de celle employée dans la démonstration du théorème de Cramer 4.2.1. Nous reviendrons sur ce point dans la section suivante.

7.2 Méthode de Cramer-Chernoff

L'emploi de l'inégalité de Chernoff permet d'obtenir la proposition suivante. Implicitement nous supposons que la variable aléatoire Z admet des moments de tout ordre (de sorte que la transformée de Laplace est finie sur $\mathbb{R}$).

Proposition 7.2.1 (Cramer-Chernoff)**.** *Soit Z une variable aléatoire réelle alors, pour tout $t \geq 0$, nous avons*

$$\mathbb{P}(Z \geq t) \leq e^{-\sup_{\lambda \geq 0}\left(\lambda t - \Lambda_Z(\lambda)\right)} = e^{-\Lambda_Z^*(t)}$$

où $\Lambda_Z^* = \sup_{\lambda \geq 0} \lambda t - \Lambda_Z(\lambda)$ désigne la transformée de Fenchel-Legendre de Λ_Z.

Remarque. 1. Il est intéressant d'observer que le membre de gauche de l'inégalité précédente est un objet purement probabiliste tandis que le membre de droite est un objet analytique.

2. Cette inégalité est précisément celle que nous avons utilisée pour démontrer la majoration dans le théorème de Cramer.

Voici quelques exemples d'applications de la méthode de Cramer-Chernoff pour des lois usuelles.

Exemple 7.2.1 (Loi normale)**.** Si $Z \sim \mathcal{N}(0,1)$ alors $\mathbb{P}(Z \geq t) \leq e^{-\frac{t^2}{2}}$ pour tout $t \geq 0$.

Remarque. La majoration est ci-dessus est proche du cas optimal (bornes de Mills, cf. [33]) :

$$\frac{t}{(1+t^2)\sqrt{2\pi}}e^{-\frac{t^2}{2}} \leq \mathbb{P}(Z \geq t) \leq \frac{1}{t\sqrt{2\pi}}e^{-\frac{t^2}{2}} \quad \text{pour tout } t > 0$$

Exemple 7.2.2 (Loi de Poisson)**.** Si $Y \sim \mathcal{P}_o(\theta)$ avec $\theta > 0$ et $Z = Y - \theta$, nous avons

$$\Lambda_Z(\lambda) = \theta(e^{\lambda} - \lambda - 1) \quad \text{pour tout} \quad \lambda \in \mathbb{R}$$

et, par suite,

$$\Lambda_Z^*(t) = \theta h(\frac{t}{\theta}) \quad \text{pour tout} \quad t \geq 0$$

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

Remarque. Observons que h admet les comportements asymptotiques suivants :

- lorsque $x \to 0$, $h(x) \sim x^2$ ce qui correspond au comportement gaussien issu du théorème de la limite centrale ;
- lorsque $x \to +\infty$, $h(x) \sim x \log x$ ce qui correspond au comportement issu du théorème de la limite centrale poissonnienne. En effet, rappelons le résultat suivant (cf.[19]) :

  Théorème 7.2.1 (TCL poisonnien)*. Soit $(X_i)_{1 \leq i \leq n}$ une famille de variables aléatoires indépendantes suivant une loi de Bernoulli $\mathcal{B}_e(p_i)$ avec $p_i \in]0,1[$ pour tout $i \in \{1,\ldots,n\}$. Si $S_n = \sum_{i=1}^n X_i$ et, lorsque $n \to \infty$, les hypothèses suivantes sont vérifiées*

  $$\max_{i=1,\ldots,n} p_i \to 0 \quad et \quad \mathbb{E}[S_n] = \sum_{i=1}^n p_i \to \lambda < +\infty.$$

  Alors, lorsque $n \to +\infty$,

  $$S_n \to \mathcal{P}_o(\lambda) \quad en\ loi.$$

Exemple 7.2.3 (Loi de Bernoulli)**.** Si $Y \sim \mathcal{B}e(p)$ avec $p \in [0,1]$ et $Z = Y - p$ alors

$$\Lambda_Z(\lambda) = \log(pe^{\lambda} + 1 - p) - \lambda p \quad \text{pour tout} \quad \lambda \in \mathbb{R}$$

et

$$\Lambda_Z^*(t) = (1-p-t)\log\left(\frac{1-p-t}{1-p}\right) + (p+t)\log\left(\frac{p+t}{p}\right) \quad \text{pour tout} \quad t \in [0, 1-p]$$

Exemple 7.2.4 (Majoration principe grande déviation)**.** Si $Z = X_1 + \ldots + X_n$ avec $(X_i)_{i=1,\ldots,n}$ une suite de variables i.i.d. nous avons

$$\Lambda_Z(\lambda) = n\Lambda_{X_1}(\lambda) \quad \text{pour tout} \quad \lambda \in \mathbb{R} \quad \text{et} \quad \Lambda_Z^*(t) = n\Lambda_{X_1}^*(\frac{t}{n}) \quad \text{pour tout} \quad t \geq 0.$$

En particulier, $\mathbb{P}(Z \geq t) \leq e^{-n\Lambda_{X_1}^*(\frac{t}{n})}$ pour tout $t \geq 0$. Il s'agit précisément de la majoration employée lors de la démonstration du théorème de Cramer 4.2.1.

7.3 Variables aléatoires sous-gaussiennes

Rappelons le fait suivant : si $Z \sim \mathcal{N}(0, \sigma^2)$ alors $\Lambda_Z(\lambda) = \frac{\lambda^2\sigma^2}{2}$ pour tout $\lambda \in \mathbb{R}$. Nous allons voir qu'il est possible d'introduire une classe de variables aléatoires qui vérifie une propriété similaire et donne lieu à des conséquences intéressantes de stabilité.

Définition 7.3.1 (Loi sous-gaussienne)**.** *Une variable aléatoire Z est dite sous-gaussienne de paramètre $\sigma^2 > 0$ si*

$$\Lambda_Z(\lambda) \leq \frac{\lambda^2\sigma^2}{2} \quad \textit{pour tout} \quad \lambda \in \mathbb{R}$$

Remarque. 1. En général, $\sigma^2 \leq \mathrm{Var}(Z)$.

2. Par l'argument de Chernoff, une variable aléatoire sous-gaussienne de paramètre σ^2 vérifie une propriété de concentration au moins gaussienne. C'est-à-dire,

$$\mathbb{P}(Z \geq t) \leq e^{-\frac{t^2}{2\sigma^2}} \quad \text{pour tout} \quad t > 0$$

3. Si Z est une variable aléatoire sous-gaussienne de paramètres σ^2, $-Z$ l'est également (avec le même paramètre σ^2). Par suite,

$$\mathbb{P}(|Z| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}} \quad \text{pour tout} \quad t > 0$$

4. Il est également possible d'exhiber la classe des variables aléatoires dites « sous-gamma » de paramètres $(a, b) \in]0; +\infty[^2$. Celles-ci vérifient alors les mêmes propriétés de concentration qu'une loi gamma standard $\Gamma(a, b)$. Ces variables aléatoires vérifient alors des résultats de stabilité similaire aux lois sous-gaussiennes. Néanmoins, nous n'aborderons pas cet aspect dans ce cours et renvoyons le lecteur vers [33] pour plus de détails.

Voici quelques propriétés vérifiées par des variables aléatoires sous-gaussiennes.

Proposition 7.3.1. *1. L'ensemble des variables aléatoires sous-gaussiennes est stable par addition. Plus précisément, si $X_1, \dots, X_n$ sont mutuellement indépendantes et sous-gaussiennes de paramètres respectifs $\sigma_i^2 > 0$ alors $Z = \sum_{i=1}^n X_i$ est une variable aléatoire sous-gaussienne de paramètre $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.*

2. Il est possible de caractériser les variables alétoires sous-gaussiennes d'après la croissance de leurs moments. Plus précisément, Z est une variable aléatoire sous-gaussienne si et seulement si il existe $C > 0$ telle que

$$\|Z\|_q \leq C\sqrt{q} \quad \textit{pour tout} \quad q \geq 1$$

où $\|Z\|_q = \left(\int_\Omega |Z|^q d\mathbb{P}\right)^{1/q}$.

Remarque. Notons que la première propriété se comporte médiocrement vis-à-vis de la dimension (i.e. du nombre de variables impliquées).

Comme nous allons le voir, le résultat suivant nous assure que les variables bornées (p.s.) sont sous-gaussiennes.

Proposition 7.3.2 (Hoeffding)**.** *Si Z est une variable aléatoire réelle, centrée, telle que*

$$Z \in [a, b] \quad \textit{p.s.}$$

(avec $a, b \in \mathbb{R}$) alors Z est une variable aléatoire sous-gaussienne de paramètre $\sigma^2 = \frac{(b-a)^2}{4}$.

Démonstration. La démonstration repose sur deux arguments : un changement de mesure combiné à une développement de Taylor de la transformée de Laplace. Avant de mettre en place ceci, il va être utile d'observer le fait élémentaire suivant.

La variance de Z est bornée par σ^2. En effet, puisque $Y \in [a, b]$ p.s. nous avons

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

En particulier, $\mathrm{Var}(Z) = \mathrm{Var}(Z - \frac{b+a}{2}) \leq \sigma^2$.

Nous pouvons à présent procéder au changement de mesure. Si $\mathbb{P} = \mathcal{L}(Z)$ nous désignerons par $\mathbb{P}_\lambda$ (avec $\lambda \in \mathbb{R}$) la mesure de probabilité définie par

$$d\mathbb{P}_\lambda = \frac{e^{\lambda Z}}{\mathbb{E}[e^{\lambda Z}]} d\mathbb{P}$$

Dans ce qui suit, pour tout $\lambda \in \mathbb{R}$, Z_λ correspondra à une variable aléatoire de loi $\mathbb{P}_\lambda$. Observons au passage que $Z_\lambda \in [a, b]$ p.s. pour tout $\lambda \in \mathbb{R}$.

Comme annoncé, nous allons utilisé un développement de Taylor (à l'ordre 2) de Λ_Z. Celui-ci nous assure que, pour tout $\lambda > 0$ il existe $\theta \in]0, \lambda[$ tel que

$$\Lambda_Z(\lambda) = \Lambda_Z(0) + \lambda \Lambda'_Z(0) + \frac{\lambda^2}{2} \Lambda''(\theta)$$

Or, des calculs élémentaires montrent que $\Lambda'_Z(\lambda) = \mathbb{E}[Z_\lambda]$ et $\Lambda''_Z(\lambda) = \mathrm{Var}(Z_\lambda)$. En conséquence, puisque Z est centré et $Z_\theta \in [a, b]$ p.s., nous avons

$$\Lambda_Z(\lambda) = \frac{\lambda^2}{2} \Lambda''(\theta) \leq \frac{\lambda^2 (b-a)^2}{8}.$$

D'où, $\mathbb{E}[e^{\lambda Z}] \leq e^{\frac{\lambda^2 (b-a)^2}{8}}$ pour tout $\lambda > 0$. Il suffit ensuite d'utiliser l'inégalité de Chernoff 7.1.2pour enfin optimiser en λ afin d'obtenir l'inégalité annoncée. Le cas $\lambda < 0$ se traite de manière similaire en remplaçant Z par $-Z$ qui est une variable aléatoire à valeurs dans $[-b, -a]$. □

Ce dernier résultat permet d'obtenir un corollaire évident.

Corollaire 7.3.1 (Inégalité d'Hoeffding)**.** *Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes telles que, pour tout $i = 1, \ldots, n$,*

$$X_i \in [a_i, b_i] \quad p.s.$$

alors, pour tout $t > 0$,

$$\mathbb{P}(S \geq t) \leq e^{-\frac{2t^2}{\sigma_n^2}} \tag{7.3.1}$$

où $S = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$ et $\sigma_n^2 = \sum_{i=1}^n (b_i - a_i)^2$

Ceci permet d'obtenir, dans le cadre des variables bornées, une inégalité de concentration reflétant de manière non asymptotique le théorème de la limite centrale.

Exemple 7.3.1. Si $X_1, \dots, X_n$ sont des variables aléatoires de Rademacher (i.e. $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$) i.i.d. alors, d'après l'inégalité de Hoeffding, nous avons

$$\mathbb{P}(S \geq t) \leq e^{-\frac{t^2}{2n}} \quad \text{pour tout} \quad t > 0$$

D'où, en posant $r = \frac{t}{\sqrt{n}}$, nous avons

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \geq r\right) \leq e^{-\frac{r^2}{2}} \quad \text{pour tout} \quad r > 0.$$

Voici deux applications de l'inégalité de Hoeffding 7.3.1.

Exemple 7.3.2 (Loi forte des grands nombres)**.** Le lecteur est bien entendu familier avec le Théorème de Kolmogorov (cf. [19]) qui s'énonce comme suit. Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. telles que $\mathbb{E}[|X_1|] < +\infty$. Si $S_n = \sum_{i=1}^n X_i$ alors

$$\lim_{n \to +\infty} \frac{S_n}{n} = \mathbb{E}[X_1] \quad p.s.$$

Nous allons démontrer ce résultat à partir de l'inégalité de Hoeffding 7.3.1. Pour cela, nous allons tout d'abord supposer que $X_1 \in [a, b]$ p.s. Ainsi, d'après l'inégalité de Hoeffding nous avons

$$\mathbb{P}\left(\Big|S_n - \mathbb{E}[S_n]\Big| \geq t\right) \leq 2e^{-\frac{2t^2}{n(b-a)^2}} \quad \text{pour tout} \quad t > 0$$

Implicitement nous avons aussi utilisé l'inégalité de Hoeffding pour la variable $-X_1$ pour obtenir cette inégalité de concentration. Posons à présent $u = tn$, ceci mène à

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}[S_n]}{n}\right| \geq u\right) \leq 2e^{-\frac{2u^2 n}{(b-a)^2}} \quad \text{pour tout} \quad u > 0.$$

En particulier, le membre de droite de cette inégalité est le terme générale d'une suite sommable. En conséquence, le Lemme de Borel-Cantelli (cf. [19]) nous assure que

$$\lim_{n \to +\infty} \frac{S_n}{n} = \mathbb{E}[X_1] \quad p.s.$$

Traitons à présent le cas général. Puisque $\mathbb{E}[|X|] < +\infty$, par construction de l'intégrale de Lebesgue (les fonctions mesurables peuvent être approchées par des fonctions étagées), il est possible d'approcher X par une variable aléatoire bornée (car étagée). Plus précisément, pour tout $\epsilon > 0$ fixé, il existe Y_1 telle que

$$|X_1 - Y_1| \leq \epsilon$$

Par la suite, nous noterons par $\tilde{S}_n = \sum_{i=1}^n Y_i$ avec $(Y_i)_{i=1,\dots,n}$ une famille de copies indépendantes de Y_1. D'après la première étape de la démonstration, il existe $N_1 \in \mathbb{N}$ tel que pour tout $n \geq N_1$ nous ayons

$$\left|\frac{\tilde{S}_n}{n} - \mathbb{E}[Y]\right| \leq \epsilon.$$

En conséquence, si $n \geq N_1$, nous avons

$$\begin{aligned}\left|\frac{S_n}{n}-\mathbb{E}[X_1]\right| &\leq \left|\frac{\tilde{S}_n}{n}-\mathbb{E}[Y]\right|+\mathbb{E}[|X_1-Y_1|]+\left|\frac{S_n}{n}-\frac{\tilde{S}_n}{n}\right| \\ &\leq 2\epsilon+\frac{1}{n}\sum_{i=1}^{n}|X_i-Y_i| \\ &\leq 3\epsilon\end{aligned}$$

par construction de la suite $(Y_i)_{i=1,\dots,n}$. ϵ étant arbitraire, ceci conclu la preuve.

Exemple 7.3.3 (Version faible de la loi du logarithme itéré)**.** Considérons $X_1,\dots,X_n$ des variables aléatoires de Rademacher i.i.d.. Nous allons montrer que

$$\limsup_{n+\to+\infty}\frac{S_n}{\sqrt{2n\log n}}\leq 1 \quad p.s.$$

où, à nouveau, $S_n=\sum_{i=1}^n X_i$. Pour cela, il suffit d'utiliser l'inégalité de Hoeffding 7.3.1 : $\mathbb{P}(S_n\geq t)\leq e^{-\frac{t^2}{2n}}$ pour tout $t>0$. En effet, pour tout $\epsilon>0$, celle-ci entraine que

$$\mathbb{P}\big(S_n\geq(1+\epsilon)\sqrt{2n\log n}\big)\leq e^{-(1+\epsilon)^2\log n}=\frac{1}{n^{(1+\epsilon)^2}}.$$

Le membre de droite de l'inégalité précédente est sommable, le Lemme de Borel-Cantelli (cf. [19]) implique alors que

$$\limsup_{n+\to+\infty}\frac{S_n}{\sqrt{2n\log n}}\leq 1+\epsilon \quad p.s.$$

Il suffit ensuite de faire $\epsilon\to 0$.

Remarque. Le résultat précédent soulève quelques questions. Nous savons, d'après la loi forte des grands nombres, que

$$\frac{1}{n}S_n\to\mathbb{E}[X_1]\,p.s. \quad \text{lorsque} \quad n\to+\infty.$$

Nous savons aussi que

$$\frac{1}{\sqrt{n}}S_n\to\mathcal{N}(0,1) \quad \text{en loi lorsque} \quad n\to+\infty.$$

Si $1<p<2$ que dire de la convergence de $\frac{1}{n^{p/2}}S_n$? A quelle place se trouve la loi du logarithme itéré parmi ces résultats ?

Avant de passer à la section suivante, il est important de souligner certaines faiblesses de l'inégalité de Hoeffding :

- la taille de la variance $\mathrm{Var}(Z)$ n'apparait pas dans celle-ci.
- Le théorème limite poisonnien (cf. théorème 7.2.1) n'est également pas reflété dans cette inégalité.

Comme nous allons le voir, au prix d'hypothèses supplémentaires, il est possible de raffiner et renforcer l'inégalité d'Hoeffding.

7.4 Inégalités de Bernstein et de Bennett

Dans cette section, $X_1, \dots, X_n$ désignent à nouveau une suite de variables aléatoires mutuellement indépendantes. Nous supposons de plus que, pour tout $i = 1, \dots, n$, il existe $b > 0$ telle que $X_i \leq b$. p.s et posons $\sigma^2 = \sum_{i=1}^n \mathbb{E}[X_i^2]$.

Sous les hypothèses précédentes, nous avons l'inégalité de Bernstein suivante.

Théorème 7.4.1 (Bernstein)**.** *Pour tout $t \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{t^2}{2\sigma^2(1+\frac{bt}{3})}}$$

avec $S_n = \sum_{i=1}^n X_i$.

Remarque. 1. Si t est petit (par rapport à σ^2) le membre de droite est équivalent à $e^{-t^2/2\sigma^2}$ et reflète le comportement gaussien du théorème de la limite centrale.

2. Au contraire, lorsque t est grand (par rapport à σ^2), le membre de droite est équivalent à e^{-t}. Cette partie ne reflète pas convenablement le théorème 7.2.1.

3. Bien entendu, si $|X_i| \leq b$ pour tout $i = 1, \dots, n$, il est possible d'obtenir l'inégalité de concentration
$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2(1+\frac{bt}{3})}} \quad \text{pour tout} \quad t \geq 0$$

Démonstration. Le début de la démonstration repose sur l'inégalité de Chernoff et l'hypothèse d'indépendance. Ainsi, pour tout $t \geq 0$ et tout $\lambda > 0$, nous obtenons

$$\begin{aligned}
\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) &\leq e^{-\lambda t} \prod_{i=1}^n \mathbb{E}[e^{\lambda(X_i - \mathbb{E}[X_i])}] \\
&= e^{-\lambda t + \sum_{i=1}^n \log(\mathbb{E}[e^{\lambda X_i}]) - \lambda \mathbb{E}[X_i]} \\
&\leq e^{-\lambda t + \sum_{i=1}^n \mathbb{E}[\phi(\lambda X_i)]}
\end{aligned}$$

avec $\phi(u) = e^u - u - 1$ puisque $\log(u) \leq u - 1$ pour tout $u > 0$.

La suite de la démonstration va reposer sur différentes manière de contrôler, pour tout $i = 1, \dots, n$, $\phi(\lambda X_i)$ suivant le signe des variables indépendantes $(X_i)_{i=1,\dots,n}$.

Puisque $\phi(u) \leq \frac{u^2}{2}$ lorsque $u < 0$, nous avons, pour tout $i = 1, \dots, n$,

$$\phi(\lambda X_i) \leq \lambda^2 \frac{X_i^2}{2} \quad \text{lorsque} \quad X_i < 0$$

Lorsque $u > 0$, un développement de Taylor de la fonction ϕ entraine que, pour tout $i = 1, \dots, n$, nous avons (lorsque $X_i > 0$)

$$\begin{aligned}\phi(\lambda X_i) &\leq \frac{\lambda^2}{2}X_i^2 + \frac{\lambda^3}{3!}X_i^3 + \ldots \\ &\leq \frac{\lambda^2}{2}X_i^2\Big(1 + \frac{\lambda b}{3} + \frac{\lambda^2 b^2}{12} + \ldots\Big).\end{aligned}$$

Notons que nous avons utilisé l'hypothèse « $X_i \leq b$, pour tout $i = 1, \ldots, n$, »afin d'obtenir la dernière inégalité. En outre, lorsque $\frac{\lambda b}{3} < 1$, nous avons aussi

$$1 + \frac{\lambda b}{3} + \frac{\lambda^2 b^2}{12} + \ldots \leq \frac{1}{1 - \frac{\lambda b}{3}}.$$

D'où, pour tout $t \geq 0$ et pour tout $\frac{\lambda b}{3} < 1$, nous obtenons

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) &\leq e^{-\lambda t + \frac{\lambda^2}{2(1-\frac{\lambda b}{3})} \times \sum_{i=1}^n \mathbb{E}[X_i^2]} \\ &= e^{-\lambda t + \frac{\lambda^2 \sigma^2}{2(1-\frac{\lambda b}{3})}}.\end{aligned}$$

Il ne reste plus qu'à optimiser en $\lambda > 0$ (sous la contrainte $\frac{\lambda b}{3} < 1$) ; cela revient à choisir $\lambda = \frac{t}{\sigma^2 + \frac{bt}{3}}$. □

Sous les mêmes hypothèses et par des méthodes similaires, il est possible d'affiner l'inégalité de Bennet. Cette précision est contenue dans l'inégalité de Bennett ; cette inégalité permet alors de quantifier le théorème de la limite centrale poissonnienne 7.2.1.

Théorème 7.4.2 (Bennett). *Sous les mêmes hypothèses, nous avons, pour tout $t \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{\sigma^2}{b^2} h\left(\frac{bt}{\sigma^2}\right)}$$

avec $h(u) = (1+u)\log(1+u) - u$ pour $u > 0$.

Remarque. 1. Si $|X_i| \leq b$ pour tout $i = 1, \ldots, n$ nous obtenons

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2e^{-\frac{\sigma^2}{b^2} h\left(\frac{bt}{\sigma^2}\right)}$$

2. Si u est petit $h(u) \sim \frac{u^2}{2}$, le membre de droite de l'inégalité de Bennett reflète toujours le comportement gaussien du théorème de la limite centrale.

3. Si u est grand $h(u) \sim u \log u$, le membre de droite de l'inégalité de Bennett reflète alors le comportement poissonnien du théorème de la limite centrale.

Pour clore ce chapitre, nous évoquons la question suivante en guise de piste de réflection :

comment serait-il possible de généraliser les inégalités de concentration précédentes lorsque les variables $(X_i)_{i=1,\ldots,n}$ ne sont plus à valeurs réelles mais à valeurs dans un espace de Banach ? Quel sens donner à X_1^2 ?

7.5 Références historiques

Ces quelques notes sont tirées de [33].

Les inégalités de concentration exponentielles pour les sommes de variables aléatoires sont connues depuis quelques temps déjà. Bernstein a présenté son inégalité en 1946, Chernoff la sienne en 1952, Bennett en 1962 et Hoeffding en 1963. De nombreuses extensions de ces inégalités sont maintenant disponibles, par exemple Azuma a étendu en 1967 l'inégalité d'Hoeffding à des martingales. Dans le chapitre suivant, nous présenterons le cas de fonctions dites « aux différences bornées » introduit par Mc Diarmid en 1989.

La loi du logarithme itéré (pour une somme de variables aléatoires i.i.d.) date des travaux de Khinchin (1924) puis ceux de Kolmogorov en 1929. De nombreuses extensions ont été obtenues depuis. citons par exemple les travaux de Hartman-Wintner (1940) concernant des marches aléatoires, ceux de Strassen (1964) qui aborda ce sujet du point de vue des principes d'invariances. Mentionnons également les travaux de Ledoux et Talagrand [83] : dans cet ouvrage, ils obtiennent ce genre de résultat dans des espaces de Banach généraux par le biais d'inégalités de concentration.