INÉGALITÉS EXPONENTIELLES POUR LES MARTINGALES

BERNARD BERCU

Université Bordeaux 1, Institut de Mathématiques de Bordeaux,
UMR 5251, 351 cours de la libération, 33405 Talence cedex, France.
Bernard.Bercu@math.u-bordeaux1.fr

Résumé. Le but de ce cours est de proposer un panorama des inégalités exponentielles pour les martingales avec des applications en statistique. La première partie porte sur un bref aperçu des théorèmes limites classiques en probabilités avec une motivation pour l'obtention d'inégalités exponentielles. La seconde partie est consacrée aux inégalités exponentielles classiques pour les sommes de variables aléatoires indépendantes. On va revenir sur les inégalités de Hoeffding, Bennett et Bernstein. La troisième partie porte sur les inégalités exponentielles pour les martingales. On va tout d'abord rappeler les inégalités de Azuma-Hoeffding et de Freedman puis l'on étudiera les inégalités exponentielles obtenues par De la Peña pour les martingales autonormalisées. La quatrième partie est consacrée à une nouvelle notion en probabilités de variables aléatoires lourdes à gauche ou à droite. Via ce nouveau concept, on va montrer comment étendre les inégalités de De la Peña.

Table des matières

1. Introduction

On débute ce cours par un petit panorama autour des théorèmes limites classiques en probabilités en passant par une motivation statistique.

Soit $(X_1, X_2, \ldots, X_n)$ une suite de variables aléatoires indépendantes et de même loi de carré intégrable sur $\mathbb{R}$, de moyenne m et de variance $\sigma^2 > 0$. On suppose que m et σ^2 sont inconnues et l'on cherche à obtenir un intervalle de confiance exact pour m en utilisant la moyenne empirique et la variance empirique données par

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{et} \qquad S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2.$$

On dispose tout d'abord de la loi des grands nombres (LGN) qui nous apprend que $\overline{X}_n$ est un bon estimateur de la moyenne inconnue m.

Théorème 1.1 (Loi des grands nombres)**.** *Soit (X_n) une suite de variables aléatoires indépendantes et de même loi intégrable sur $\mathbb{R}$ et de moyenne m. Alors, on a*

$$\lim_{n\to\infty} \overline{X}_n = m \qquad p.s. \tag{1.1}$$

La LGN ne sert pas à l'obtention d'un intervalle de confiance pour m. On a besoin des fluctuations dans la LGN données par le théorème limite central (TLC) afin de proposer un intervalle de confiance asymptotique pour m.

Théorème 1.2 (Théorème limite central)**.** *Soit (X_n) une suite de variables aléatoires indépendantes et de même loi de carré intégrable sur $\mathbb{R}$, de moyenne m et de variance σ^2. Alors, on a*

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \tag{1.2}$$

On ne peut utiliser ce TLC car la variance σ^2 est inconnue. Cependant, par la LGN et l'égalité

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} X_k^2 - \frac{n}{n-1}(\overline{X}_n)^2,$$

S_n^2 converge presque sûrement vers σ^2. On déduit alors du TLC que

$$\sqrt{n}\left(\frac{\overline{X}_n - m}{S_n}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On a donc un intervalle de confiance asymptotique symétrique pour m

$$\left[\overline{X}_n - a\frac{S_n}{\sqrt{n}}, \overline{X}_n + a\frac{S_n}{\sqrt{n}}\right]$$

avec, pour un niveau de confiance $1-\alpha$, $\mathbb{P}(|Z| \leq a) = 1-\alpha$ où Z suit la loi $\mathcal{N}(0,1)$. On pourrait penser à utiliser les propriétés de grandes déviations de la suite $(\overline{X}_n)$. Cependant, cette démarche est illusoire et vouée à l'échec car la fonction de taux du principe de grandes déviations dépend très souvent des paramétres inconnus m et σ^2. Pour s'en convaincre, on va énoncer le théorème de Bahadur-Rao et son application immédiate dans le cas gaussien.

Théorème 1.3 (Théorème de Bahadur-Rao)**.** *Soit (X_n) une suite de variables aléatoires réelles indépendantes et de même loi intégrable sur $\mathbb{R}$ et de moyenne m. On suppose que cette loi est absolument continue et que sa log-Laplace L est finie sur $\mathbb{R}$ tout entier. On note I sa transformée de Legendre donnée, pour tout $x \in \mathbb{R}$, par*

$$I(x) = \sup_{t\in\mathbb{R}}\{xt - L(t)\}.$$

Alors, la suite $(\overline{X}_n)$ satisfait un principe de grandes déviations précis. En particulier, pour tout $x \in \mathbb{R}$, il existe une suite $(d_k(x))$ telle que, pour tout $p \geq 0$ et n assez grand, si $x > m$

$$\mathbb{P}(\overline{X}_n \geq x) = \frac{\exp(-nI(x))}{\sigma_x t_x\sqrt{2\pi n}}\left[1 + \sum_{k=1}^{p}\frac{d_k(x)}{n^k} + \mathcal{O}\Big(\frac{1}{n^{p+1}}\Big)\right] \tag{1.3}$$

tandis que si $x < m$

$$\mathbb{P}(\overline{X}_n \leq x) = -\frac{\exp(-nI(x))}{\sigma_x t_x\sqrt{2\pi n}}\left[1 + \sum_{k=1}^{p}\frac{d_k(x)}{n^k} + \mathcal{O}\Big(\frac{1}{n^{p+1}}\Big)\right] \tag{1.4}$$

où t_x est donné par $L'(t_x) = x$ et $\sigma_x^2 = L''(t_x)$. Les coefficients $d_k(x)$ peuvent être calculés explicitement en fonction des dérivées successives de L au point t_x.

Remarque 1.4. *Le théorème de Bahadur-Rao peut facilement être établi dans le cas gaussien avec une succession d'intégrations par parties. Soit (X_n) une suite de variables aléatoires indépendantes et de même loi normale $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$. Alors, pour tout $x > 0$, $p \geq 0$ et n assez grand, on a avec $I(x) = x^2/2\sigma^2$,*

$$\mathbb{P}\Big(\Big|\overline{X}_n - m\Big| \geq x\Big) = \frac{2\sigma\exp(-nI(x))}{x\sqrt{2\pi n}}\left[1 + \sum_{k=1}^{p}\frac{d_k(x)}{n^k} + \mathcal{O}\Big(\frac{1}{n^{p+1}}\Big)\right].$$

On peut donc réaliser qu'obtenir un intervalle de confiance exact pour m si les variables aléatoires (X_n) ne sont pas bornées est loin d'être aisé. On va maintenant se placer dans le cadre simplifié des variables aléatoires bornées en supposant que (X_n) est une suite de variables aléatoires réelles indépendantes et de même loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$. Il est immédiat que $m = p$ et $\sigma^2 = p(1-p)$. Afin d'obtenir un intervalle de confiance pour p, une première approche naïve consiste à utiliser l'inégalité de Markov qui entraîne que pour tout $a > 0$

$$\mathbb{P}(|\overline{X}_n - p| \geq a) \leq \frac{p(1-p)}{na^2}.$$

Par l'inégalité élémentaire $4p(1-p) \leq 1$, il en découle que

$$\mathbb{P}(|\overline{X}_n - p| \leq a) \geq 1 - \frac{1}{4na^2}.$$

Si l'on pose $\alpha = 1/4na^2$, on a dès que $n > 1/4a^2$, $0 < \alpha < 1$. Un premier intervalle de confiance exact pour p, avec un niveau de confiance $1 - \alpha$, est donné par

$$I(p) = \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}, \overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right].$$

Une alternative est d'utiliser le TLC rencontré précédemment. On a

$$\overline{X}_n - p = \frac{\sqrt{p(1-p)}}{\sqrt{n}} Y_n \quad \text{avec} \quad Y_n = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}}.$$

Par le TLC, (Y_n) converge en loi vers une variable aléatoire Y de loi $\mathcal{N}(0,1)$. De plus, pour tout $a > 0$, on a $\mathbb{P}(|\overline{X}_n - p| \leq a) \geq \mathbb{P}(|Y_n| \leq 2a\sqrt{n})$. Par suite, si l'on approche la loi de Y_n par la loi $\mathcal{N}(0,1)$, un second intervalle de confiance asymptotique pour p est donné par

$$J(p) = \left[\overline{X}_n - \frac{a}{2\sqrt{n}}, \overline{X}_n + \frac{a}{2\sqrt{n}}\right]$$

avec, pour un niveau de confiance $1-\alpha$, $\mathbb{P}(|Y| \leq a) = 1-\alpha$. On peut aussi utiliser le TLC associé à la convergence presque sûre de $\overline{X}_n(1-\overline{X}_n)$ vers la variance $p(1-p)$. Il en découle que

$$\sqrt{n}\,\frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On a donc un troisième intervalle de confiance asymptotique pour p donné par

$$K(p) = \left[\overline{X}_n - a\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + a\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

avec, pour un niveau de confiance $1-\alpha$, $\mathbb{P}(|Z| \leq a) = 1-\alpha$ où Z suit la loi $\mathcal{N}(0,1)$. Finalement, on va voir par l'inégalité exponentielle de Hoeffding que pour tout $a > 0$

$$\mathbb{P}(|\overline{X}_n - p| \geq a) \leq 2\exp(-2na^2).$$

On peut en déduire que

$$\mathbb{P}(|\overline{X}_n - p| \leq a) \geq 1 - 2\exp(-2na^2).$$

Si l'on pose $\alpha = 2\exp(-2na^2)$, on a dès que $n > \log 2/2a^2$, $0 < \alpha < 1$. On trouve donc un quatrième intervalle de confiance exact pour p, avec un niveau de confiance $1-\alpha$, donné par

$$L(p) = \left[\overline{X}_n - \sqrt{\frac{\log(2/\alpha)}{2n}}, \overline{X}_n + \sqrt{\frac{\log(2/\alpha)}{2n}}\right].$$

La figure suivante compare les bornes de ces quatre intervalles de confiance avec le choix standard $\alpha = 5\%$ et pour n variant de 10 à 50. Seule les bornes provenant de $K(p)$ sont aléatoires. On peut noter que les intervalles de confiance $J(p)$ et $K(p)$ sont toujours plus précis que $I(p)$ et $L(p)$, à condition bien sûr que l'approximation gaussienne soit justifiée. A contrario, les bornes dans $I(p)$ et $L(p)$, obtenues grâce aux inégalités de Markov et Hoeffding, sont toujours valables pour toute valeur assez grande de n. On comprend donc sur cet exemple l'intérêt de l'inégalité de Hoeffding, surtout pour les petites valeurs de α.

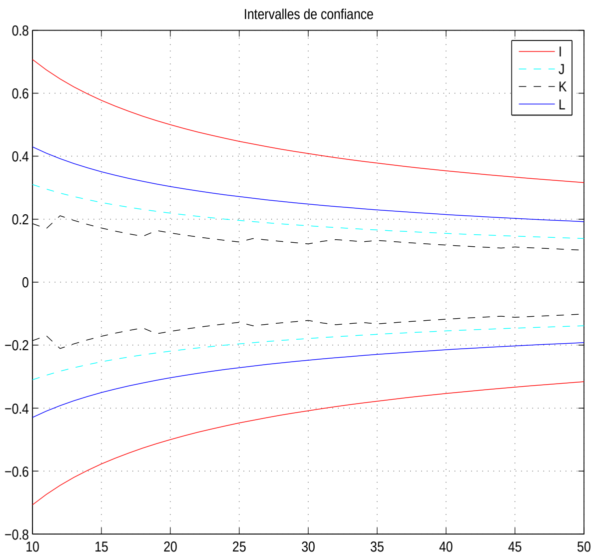

FIG. 1. Comparaison des intervalles de confiance

2. Sur les sommes de variables aléatoires indépendantes

On va tout d'abord revenir sur la célèbre inégalité de Hoeffding pour les sommes de variables aléatoires indépendantes et bornées. On parlera ensuite des inégalités de Bennett et Bernstein. Le lecteur curieux pourra consulter [3], [15] ainsi que l'excellent cours de McDiarmid [18].

2.1. **Inégalité de Hoeffding.**

Théorème 2.1 (Inégalité de Hoeffding)**.** *Soit* (X_n) *une suite de variables aléatoires indépendantes. On suppose que, pour tout* $1 \leq k \leq n$*, on peut trouver des constantes* $a_k < b_k$ *telles que* $a_k \leq X_k \leq b_k$ *p.s. Si*

$$S_n = \sum_{k=1}^{n} X_k$$

alors, pour tout $x \geq 0$*, on a*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq 2 \exp\left(-\frac{2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right). \tag{2.1}$$

La preuve de l'inégalité de Hoeffding repose sur le lemme suivant.

Lemme 2.2. *Soit* X *une variable aléatoire réelle centrée telle que* $a \leq X \leq b$ *p.s. avec* $a < b$*. Alors, pour tout* $t > 0$*, on a*

$$\mathbb{E}[\exp(tX)] \leq \exp\left(\frac{t^2}{8}(b-a)^2\right). \tag{2.2}$$

Preuve. Par la convexité de l'exponentielle, on a pour tout $a \leq x \leq b$

$$\exp(tx) \leq \frac{b-x}{b-a}\exp(ta) + \frac{x-a}{b-a}\exp(tb).$$

Par passage à l'espérance, comme $\mathbb{E}[X] = 0$, on a

$$\begin{aligned}\mathbb{E}[\exp(tX)] &\leq \frac{b}{b-a}\exp(ta) - \frac{a}{b-a}\exp(tb), \\ &\leq (1-p)\exp(-py) + p\exp((1-p)y)\end{aligned}$$

avec $p = -a/(b-a)$ et $y = (b-a)t$ ce qui entraîne $at = -py$ et $bt = (1-p)y$. On en déduit que

$$\mathbb{E}[\exp(tX)] \leq \exp(h(y))$$

avec $\exp(h(y)) = \exp(-py)((1-p)+p\exp(y))$ donc $h(y) = -py+\log(1-p+p\exp(y))$. Cependant, il est clair que

$$\begin{aligned}h'(y) &= -p + \frac{p}{p+(1-p)\exp(-y)}, \\ h''(y) &= \frac{p(1-p)\exp(-y)}{(p+(1-p)\exp(-y))^2} \leq \frac{1}{4}.\end{aligned}$$

Comme $h(0) = 0$ et $h'(0) = 0$, on a par la formule de Taylor qu'il existe $\theta \in \mathbb{R}$ avec $0 \leq |\theta| \leq |y|$ tel que

$$h(y) = h(0) + yh'(0) + \frac{y^2}{2}h''(\theta) \leq \frac{y^2}{8} = \frac{t^2}{8}(b-a)^2$$

ce qui achève la preuve du lemme 2.2. $\square$

Preuve du théorème 2.1. Pour tous $x \geq 0$ et $t > 0$, on a par l'inégalité de Markov

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) &= \mathbb{P}(\exp(t(S_n - \mathbb{E}[S_n])) \geq \exp(tx)), \\ &\leq \exp(-tx)\mathbb{E}[\exp(t(S_n - \mathbb{E}[S_n]))], \\ &\leq \exp(-tx)\mathbb{E}\Big[\exp\Big(t\sum_{k=1}^{n} Y_k\Big)\Big] \qquad (2.3)\end{aligned}$$

avec, pour tout $1 \leq k \leq n$, $Y_k = X_k - \mathbb{E}[X_k]$. La suite (Y_n) est constituée de variables aléatoires indépendantes et centrées avec, pour tout $1 \leq k \leq n$, $c_k \leq Y_k \leq d_k$ p.s. où $c_k = a_k - \mathbb{E}[X_k]$ et $d_k = b_k - \mathbb{E}[X_k]$. On peut noter que $d_k - c_k = b_k - a_k$. Il découle alors de l'inégalité (2.2) que pour tout $t > 0$

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(t\sum_{k=1}^{n} Y_k\Big)\Big] &= \mathbb{E}\Big[\prod_{k=1}^{n}\exp(tY_k)\Big] = \prod_{k=1}^{n}\mathbb{E}\left[\exp(tY_k)\right], \\ &\leq \prod_{k=1}^{n}\exp\left(\frac{t^2}{8}(b_k-a_k)^2\right) = \exp\left(\frac{t^2 v_n}{8}\right) \qquad (2.4)\end{aligned}$$

avec

$$v_n = \sum_{k=1}^{n} (b_k - a_k)^2.$$

Les inégalités (2.3) et (2.4) entraînent alors que pour tous $x \geq 0$ et $t > 0$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-tx + \frac{t^2 v_n}{8}\right).$$

En prenant $t = 4x/v_n$, on en déduit que

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{2x^2}{v_n}\right).$$

En remplaçant X_k par $-X_k$, on montre de la même manière que pour tout $x \geq 0$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \leq -x) \leq \exp\left(-\frac{2x^2}{v_n}\right).$$

Finalement, pour tout $x \geq 0$

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) = \mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) + \mathbb{P}(S_n - \mathbb{E}[S_n] \leq -x) \leq 2\exp\left(-\frac{2x^2}{v_n}\right)$$

ce qui achève la preuve de l'inégalité de Hoeffding. □

Exercice 1. Majoration de la variance dans Hoeffding Soit X une variable aléatoire réelle centrée telle que $a \leq X \leq b$ p.s. avec $a < b$. Montrer que la borne supérieure dans (2.2) est associée à la majoration de la variance

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

Exercice 2. Variable aléatoire sous-gaussienne Soit X une variable aléatoire réelle de carré intégrable, centrée et de variance σ^2. On dit que X est sous-gaussienne s'il existe une constante $a > 0$ telle que, pour tout $t \in \mathbb{R}$

$$\mathbb{E}[\exp(tX)] \leq \exp\Big(\frac{a^2 t^2}{2}\Big).$$

Le plus petit réel $a > 0$ satisfaisant cette inégalité, noté $a(X)$, s'appelle le moment sous-gaussien de X et l'on a $\sigma^2 \leq a(X)$.

1) Si X suit la loi $\mathcal{N}(0, \sigma^2)$, vérifier que X est sous-gaussienne avec $a(X) = \sigma^2$.
2) Montrer que, si X est une variable aléatoire réelle centrée telle que $|X| \leq c$ p.s. où c est une constante > 0, alors X est sous-gaussienne avec $a(X) \leq c^2$.

Indication : On utilisera que, pour tout $t \in \mathbb{R}$ et pour tout $x \in [-1, 1]$

$$\exp(tx) \leq \exp(t^2/2) + x\sinh(t).$$

Exercice 3. LGN sous-gaussienne Soit (X_n) une suite de variables aléatoires réelles indépendantes, centrées et de carré intégrable et soit

$$S_n = \sum_{k=1}^{n} X_k.$$

Pour tout $n \geq 1$, on suppose que X_n est sous-gaussienne avec $a(X_n) \leq 1$.

1) Montrer que pour tout $x \geq 0$

$$\mathbb{P}(|S_n| \geq x) \leq 2\exp\Big(-\frac{x^2}{2n}\Big).$$

2) En déduire par le lemme de Borel-Cantelli que

$$\lim_{n\to\infty} \frac{S_n}{n} = 0 \qquad \text{p.s.}$$

Exercice 4. LGN et TLC Rademacher Soit (ε_n) une suite de variables aléatoires indépendantes et de même loi de Rademacher symétrique c'est-à-dire que, pour tout $n \geq 1$, $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$. Soit (X_n) la suite donnée par $X_n = n^a \varepsilon_n$ avec $a > 0$ et soit

$$S_n = \sum_{k=1}^{n} X_k.$$

1) Montrer que pour tout $t \in \mathbb{R}$

$$L_n(t) = \mathbb{E}[\exp(tS_n)] = \prod_{k=1}^{n} \cosh(tk^a).$$

2) En déduire que pour tout $t \in \mathbb{R}$

$$\mathbb{E}[\exp(tS_n)] \leq \exp\Big(\frac{t^2 v_n}{2}\Big) \qquad \text{avec} \qquad v_n = \sum_{k=1}^{n} k^{2a}.$$

3) Montrer que pour tout $x \geq 0$

$$\mathbb{P}(|S_n| \geq x) \leq 2\exp\Big(-\frac{x^2}{2v_n}\Big).$$

4) En déduire par le lemme de Borel-Cantelli que

$$\lim_{n\to\infty} \frac{S_n}{n^{2a+1}} = 0 \qquad \text{p.s.}$$

5) Montrer finalement le TLC

$$\frac{S_n}{n^a\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\Big(0, \frac{1}{2a+1}\Big).$$

2.2. **Inégalité de Bennett.** On aborde maintenant l'inégalité de Bennett dans laquelle entre en jeu la variance de S_n au lieu d'une majoration de cette variance.

Théorème 2.3 (Inégalité de Bennett)**.** *Soit (X_n) une suite de variables aléatoires indépendantes et de carré intégrable. On suppose que, pour tout $1 \leq k \leq n$, il existe une constante $c > 0$ telle que $X_k - \mathbb{E}[X_k] \leq c$ p.s. Si*

$$S_n = \sum_{k=1}^{n} X_k \qquad \textit{et} \qquad v_n = \sum_{k=1}^{n} \mathit{Var}(X_k)$$

alors, pour tout $x \geq 0$, on a

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{v_n}{c^2} h\left(\frac{xc}{v_n}\right)\right) \tag{2.5}$$

avec $h(x) = (1+x)\log(1+x) - x$. En particulier, pour tout $x \geq 0$, on a

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{x^2}{2(v_n + xc/3)}\right). \tag{2.6}$$

Remarque 2.4. *Il est facile de vérifier que pour tout $x \geq 0$*

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

Sinon, on peut se reporter à la figure 2. Par suite, l'inégalité (2.6), due à Bernstein, découle immédiatement de l'inégalité de Bennett. De plus, si $v_n = \sigma^2 n$ avec $\sigma^2 > 0$ et $x = o(n)$, alors la majoration dans (2.6) est semblable à $\exp(-x^2/2\sigma^2 n)$. Elle correspond à un comportement sous-gaussien car par le TLC

$$\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

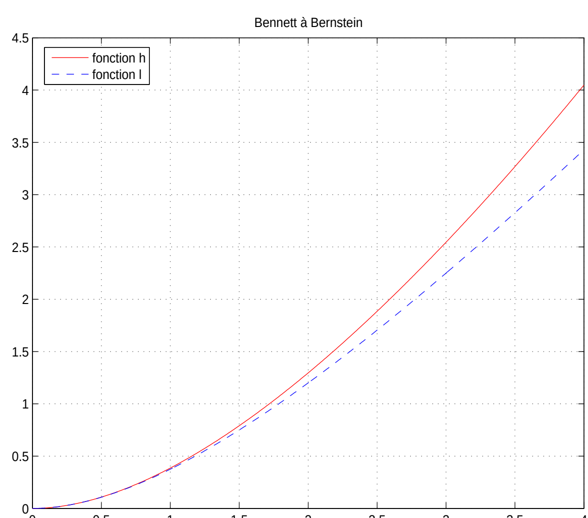

FIG. 2. Comparaison des fonctions h et ℓ

La preuve de l'inégalité de Bennett repose sur le lemme suivant.

Lemme 2.5. *Soit f la fonction définie par $f(0) = 1/2$ et pour tout $x \in \mathbb{R}^*$*

$$f(x) = \frac{\exp(x) - 1 - x}{x^2}.$$

Alors, f est une fonction croissante. De plus, soit X une variable aléatoire réelle centrée et de carré intégrable, de variance $\sigma^2 > 0$. On suppose que $X \leq c$ p.s. avec $c > 0$. Alors, pour tout $t \geq 0$, on a

$$\mathbb{E}[\exp(tX)] \leq 1 + t^2 f(tc)\sigma^2 \leq \exp(t^2 f(tc)\sigma^2). \tag{2.7}$$

Preuve. On peut voir aisément que f est croissante car pour tout $x \in \mathbb{R}^*$

$$f'(x) = \frac{g(x)}{x^3}$$

avec $g(x) = (x-2)\exp(x) + 2 + x$. Il est clair que $g'(x) = (x-1)\exp(x) + 1$ et $g''(x) = x\exp(x)$. Par suite, comme $g'(0) = 0$, $g'(x) \geq 0$ pour tout $x \in \mathbb{R}$. La fonction g est donc croissante et comme $g(0) = 0$, alors pour tout $x \geq 0$, $g(x) \geq 0$ tandis que pour tout $x \leq 0$, $g(x) \leq 0$ ce qui entraîne que pour tout $x \in \mathbb{R}^*$, $f'(x) > 0$ donc f est croissante. Si $X \leq c$, alors pour tout $t > 0$, $tX \leq tc$ p.s. donc $f(tX) \leq f(tc)$ p.s. Comme $\mathbb{E}[X] = 0$, il en découle que pour tout $t \geq 0$

$$\mathbb{E}[\exp(tX)] = 1 + \mathbb{E}[t^2X^2 f(tX)] \leq 1 + \mathbb{E}[t^2X^2 f(tc)] = 1 + t^2 f(tc)\sigma^2.$$

La dernière inégalité de (2.7) est immédiate car, pour tout $x \in \mathbb{R}$, $1 + x \leq \exp(x)$. $\square$

Preuve du théorème 2.3. Pour tous $x \geq 0$ et $t > 0$, on a déjà vu par l'inégalité de Markov et l'hypothèse d'indépendance sur la suite (X_n) que

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp(-tx) \prod_{k=1}^{n} \mathbb{E}[\exp(tY_k)] \tag{2.8}$$

avec $Y_k = X_k - \mathbb{E}[X_k]$. La suite (Y_n) est constituée de variables aléatoires indépendantes et centrées avec $\mathrm{Var}(Y_k) = \mathrm{Var}(X_k)$. De plus, pour tout $1 \leq k \leq n$, on a $Y_k \leq c$ p.s. On tire alors des inégalités (2.7) et (2.8) que pour tous $x \geq 0$ et $t > 0$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\big(-tx + t^2 f(tc) v_n\big)$$

avec $v_n = \displaystyle\sum_{k=1}^{n} \mathrm{Var}(X_k)$. On minimise cette majoration en prenant

$$t = \frac{1}{c} \log\Big(1 + \frac{xc}{v_n}\Big).$$

On trouve alors que

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{v_n}{c^2} h\Big(\frac{xc}{v_n}\Big)\right)$$

ce qui termine la preuve de l'inégalité de Bennett. $\square$

2.3. **Inégalité de Bernstein.** On va maintenant remplacer une hypothèse de bornitude par une hypothèse de moment.

Théorème 2.6 (Inégalité de Bernstein)**.** *Soit* (X_n) *une suite de variables aléatoires indépendantes et soit*

$$S_n = \sum_{k=1}^{n} X_k \qquad et \qquad v_n = \sum_{k=1}^{n} Var(X_k).$$

On suppose que pour tout entier $p \geq 2$*, il existe une constante* $c > 0$ *telle que*

$$\sum_{k=1}^{n} \mathbb{E}[|X_k - \mathbb{E}[X_k]|^p] \leq \frac{p! c^{p-2}}{2} v_n. \tag{2.9}$$

Alors, pour tout $x \geq 0$*, on a*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{v_n}{c^2}\,\ell\left(\frac{xc}{v_n}\right)\right) \tag{2.10}$$

avec $\ell(x) = (1+x) - \sqrt{1+2x}$*. En particulier, pour tout* $x \geq 0$*, on a*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{x^2}{2(v_n + xc)}\right). \tag{2.11}$$

Remarque 2.7. *Il est facile de vérifier que pour tout* $x \geq 0$

$$\ell(x) \geq k(x) = \frac{x^2}{2(1+x)}.$$

On peut aussi se reporter à la figure 3. L'inégalité (2.10) entraîne directement (2.11).

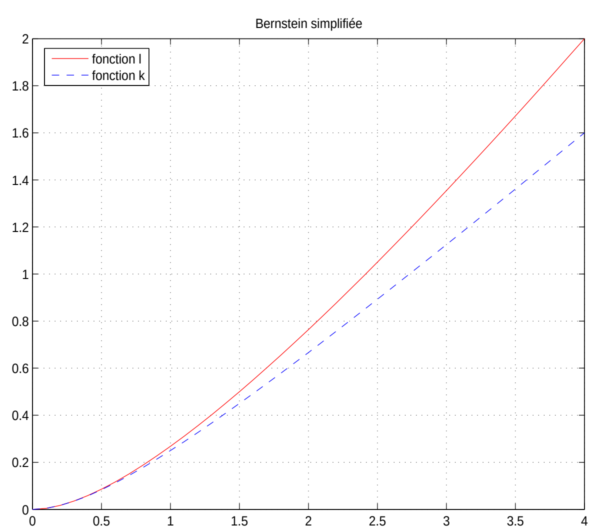

FIG. 3. Comparaison des fonctions ℓ et h

Remarque 2.8. *Il est clair que l'inégalité de Bernstein (2.11) entraîne (2.6) pour les variables aléatoires bornées. En effet, soit (X_n) une suite de variables aléatoires indépendantes avec, pour tout $1 \leq k \leq n$, $|X_k - \mathbb{E}[X_k]| \leq a$ p.s. où $a > 0$. Alors, pour tout entier $p \geq 2$, on a*

$$\sum_{k=1}^{n} \mathbb{E}[|X_k - \mathbb{E}[X_k]|^p] \leq a^{p-2}\, v_n \leq \frac{p!c^{p-2}}{2}\, v_n$$

avec $c = a/3$ ce qui implique (2.6).

Remarque 2.9. *Sous l'hypothèse de moment (2.9), on a également pour tout $x \geq 0$*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq 2\exp\left(-\frac{x^2}{2(v_n + xc)}\right). \tag{2.12}$$

Preuve du théorème 2.6. Pour tous $x \geq 0$ et $t > 0$, on a déjà vu que

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp(-tx)\prod_{k=1}^{n}\mathbb{E}[\exp(tY_k)] \tag{2.13}$$

avec $Y_k = X_k - \mathbb{E}[X_k]$. Cependant, on a par théorème de convergence monotone

$$\mathbb{E}[\exp(tY_k)] \leq 1 + \mathbb{E}\left[\sum_{p=2}^{\infty}\frac{|tY_k|^p}{p!}\right] = 1 + \sum_{p=2}^{\infty}\frac{t^p}{p!}\mathbb{E}[|Y_k|^p].$$

Via l'inégalité élémentaire $1 + x \leq \exp(x)$, on en déduit que

$$\mathbb{E}[\exp(tY_k)] \leq \exp\left(\sum_{p=2}^{\infty}\frac{t^p}{p!}\mathbb{E}[|Y_k|^p]\right). \tag{2.14}$$

On tire alors des inégalités (2.13) et (2.14) que pour tous $x \geq 0$ et $t > 0$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-tx + \sum_{k=1}^{n}\sum_{p=2}^{\infty}\frac{t^p}{p!}\mathbb{E}[|Y_k|^p]\right).$$

La condition de moment (2.9) entraîne alors que

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) &\leq \exp\left(-tx + \frac{t^2 v_n}{2}\sum_{p=2}^{\infty}(tc)^{p-2}\right),\\ &\leq \exp\left(-tx + \frac{t^2 v_n}{2(1-tc)}\right) \qquad (2.15)\end{aligned}$$

pourvu que $0 < tc < 1$. On peut minimiser cette majoration en prenant

$$t = \frac{1}{c}\left(1 - \sqrt{\frac{v_n}{2xc + v_n}}\right).$$

On trouve alors que

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{v_n}{c^2}\ell\Big(\frac{xc}{v_n}\Big)\right)$$

ce qui achève la preuve de l'inégalité de Berstein. □

3. Inégalités exponentielles pour les martingales

Les martingales possèdent une place privilégiée en probabilités car elles sont la généralisation naturelle de la notion de sommes de variables aléatoires indépendantes. Il y a une pléiade de beaux résultats à raconter sur les martingales. On ne parlera ici que du comportement asymptotique des martingales à temps discret avant d'aborder les inégalités exponentielles de Azuma-Hoeffding, Freedman et De la Peña.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité muni d'une filtration $\mathbb{F} = (\mathcal{F}_n)$ où $\mathbb{F}$ est une suite croissante de sous-tribus de $\mathcal{A}$ et $\mathcal{F}_n$ est la tribu des évènements se produisant avant l'instant n. Une suite (M_n) de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, est adaptée à $\mathbb{F}$ si, pour tout $n \geq 0$, M_n est $\mathcal{F}_n$-mesurable.

Définition 3.1. *Soit (M_n) une suite de variables aléatoires réelles, intégrable et adaptée à $\mathbb{F}$. On dit que (M_n) est une martingale (MG), sous-martingale (sMG) ou surmartingale (SMG) si, pour tout $n \geq 0$, on a respectivement*

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n, \qquad \mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n, \qquad \mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n.$$

Remarque 3.2. *Une martingale reste constante en espérance conditionnelle tandis qu'une sMG croît et qu'une SMG décroît en espérance conditionnelle.*

Exercice 5. Somme Soit (X_n) une suite de variables aléatoires indépendantes et intégrables avec, pour tout $n \geq 0$, $\mathbb{E}[X_n] = m$ et soit

$$S_n = \sum_{k=1}^{n} X_k.$$

Montrer que (S_n) est une MG, sMG ou SMG suivant que $m = 0$, $m \geq 0$ ou $m \leq 0$, respectivement. On peut penser en particulier à la somme de variables aléatoires indépendantes de loi de Rademacher $\mathcal{R}(p)$ avec $0 < p < 1$.

Exercice 6. Produit Soit (X_n) une suite de variables aléatoires indépendantes, positives et intégrables avec, pour tout $n \geq 0$, $\mathbb{E}[X_n] = m$ et soit

$$P_n = \prod_{k=1}^{n} X_k.$$

Montrer que (P_n) est une MG, sMG ou SMG suivant que $m = 1$, $m \geq 1$ ou $m \leq 1$, respectivement. On peut penser en particulier au produit de variables aléatoires indépendantes de loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$.

En analyse, l'étude de la convergence des suites de nombres réels repose sur des critères faciles à vérifier. Il en va de même en probabilités comme on peut le voir sur le théorème de convergence de Doob dont on trouvera une preuve dans [19]

Théorème 3.3 (Théorème de Doob)**.** *Soit (M_n) une martingale, sous-martingale ou surmartingale, bornée dans $\mathbb{L}^1$, c'est-à-dire satisfaisant*

$$\sup_{n \geq 0} \mathbb{E}[|M_n|] < +\infty.$$

Alors, (M_n) converge presque sûrement vers une variable aléatoire intégrable M_∞.

Théorème 3.4. *Soit (M_n) une martingale bornée dans $\mathbb{L}^p$ avec $p \geq 1$, c'est-à-dire satisfaisant*

$$\sup_{n \geq 0} \mathbb{E}[|M_n|^p] < +\infty.$$

Si $p > 1$, alors (M_n) converge presque sûrement et dans $\mathbb{L}^p$ vers une variable aléatoire M_∞. Par contre, si $p = 1$, alors (M_n) converge presque sûrement et cette convergence n'a lieu dans $\mathbb{L}^1$ que si (M_n) est équi-intégrable donc, si

$$\lim_{a \to \infty} \sup_{n \geq 0} \mathbb{E}[|M_n| \mathrm{I}_{(|M_n| \geq a)}] = 0.$$

Exercice 7. Martingale autorégressive Soit (X_n) la suite de variables aléatoires définie, pour tout $n \geq 0$, par

$$X_{n+1} = \theta X_n + (1 - \theta)\varepsilon_{n+1}$$

où $0 < \theta < 1$ et $X_0 = p$ avec $0 < p < 1$. On note $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ et l'on suppose que, pour tout $n \geq 0$, la loi de ε_{n+1} sachant $\mathcal{F}_n$ est la loi de Bernoulli $\mathcal{B}(X_n)$.

1) Vérifier que, pour tout $n \geq 0$, $0 < X_n < 1$.
2) Montrer que (X_n) est une martingale qui converge presque sûrement et dans $\mathbb{L}^2$ vers une variable aléatoire L.
3) Montrer que, pour tout $n \geq 0$,
$$\mathbb{E}[(X_{n+1} - X_n)^2] = (1 - \theta)^2 \mathbb{E}[X_n(1 - X_n)].$$
4) Calculer $\mathbb{E}[L(1 - L)]$ puis conclure que L suit une loi de Bernoulli $\mathcal{B}(p)$.

Exercice 8. Urne de Polya. Une urne contient une boule blanche et une boule rouge. On tire une boule dans l'urne, on regarde sa couleur, puis on la remplace par deux boules de la même couleur, ce qui donne la nouvelle composition de l'urne après l'instant 1. On réitère ensuite la même procédure. Après l'instant n, il y a donc $n + 2$ boules dans l'urne. Soit X_n le nombre de boules blanches dans l'urne après l'instant n et soit M_n la proportion associée

$$M_n = \frac{X_n}{n + 2}.$$

1) Montrer que X_n est uniformément distribuée sur $\{1, 2, \ldots, n + 1\}$.
2) Montrer que (M_n) est une martingale qui converge presque sûrement et dans tous les $\mathbb{L}^p$ avec $p \geq 1$, vers une variable aléatoire L de loi uniforme sur $[0, 1]$.

3.1. **Martingales de carré intégrable.** On va maintenant énoncer la LGN et le TLC pour les martingales de carré intégrable qui généralisent les résultats pour les sommes de variables aléatoires indépendantes de la section 1. Le lecteur curieux trouvera des preuves abordables dans [5], [10], [14], [17], [19]. Tout d'abord, une martingale (M_n) est de carré intégrable si, pour tout $n \geq 0$, $\mathbb{E}[M_n^2] < +\infty$. Dans ce cas, (M_n^2) est une sMG positive et intégrable.

Définition 3.5. *Soit (M_n) une martingale de carré intégrable. On appelle processus croissant associé à (M_n), la suite $(<M>_n)$ définie par $<M>_0= 0$ et, pour tout $n \geq 1$,*

$$<M>_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}]$$

avec $\Delta M_n = M_n - M_{n-1}$

Théorème 3.6 (Loi des grands nombres)**.** *Soit (M_n) une martingale de carré intégrable et soit $(<M>_n)$ son processus croissant. On pose*

$$<M>_\infty = \lim_{n\to\infty} <M>_n .$$

1) Sur $\{<M>_\infty < \infty\}$, (M_n) converge p.s. vers M_∞ de carré intégrable.
2) Sur $\{<M>_\infty = \infty\}$, on a

$$\lim_{n\to\infty} \frac{M_n}{<M>_n} = 0 \qquad p.s.$$

Remarque 3.7. *Une conséquence utile est que si $<M>_n = \mathcal{O}(a_n)$ où (a_n) est une suite déterministe positive croissante vers l'infini, alors $M_n = o(a_n)$ p.s.*

Théorème 3.8 (Théorème limite central)**.** *Soit (M_n) une martingale de carré intégrable et soit $(<M>_n)$ son processus croissant. Soit (a_n) une suite déterministe, positive, croissante vers l'infini. On suppose que*
1) Il existe une limite déterministe $\ell \geq 0$ telle que

$$\frac{<M>_n}{a_n} \xrightarrow[n\to+\infty]{\mathbb{P}} \ell.$$

2) La condition de Lindeberg est satisfaite c'est-à-dire pour tout $\varepsilon > 0$

$$\frac{1}{a_n} \sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^2 \mathrm{I}_{(|\Delta M_k| \geq \varepsilon\sqrt{a_n})} | \mathcal{F}_{k-1}] \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Alors, on a

$$\frac{1}{\sqrt{a_n}} M_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \ell). \tag{3.1}$$

De plus, si $\ell > 0$, on a

$$\sqrt{a_n} \frac{M_n}{<M>_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \ell^{-1}). \tag{3.2}$$

3.2. **Inégalité de Azuma-Hoeffding.** On revient à présent sur l'inégalité de Hoeffding pour les martingales à accroissements bornés, encore appelée inégalité de Azuma-Hoeffding [1],[15].

Théorème 3.9 (Inégalité de Azuma-Hoeffding)**.** *Soit (M_n) une martingale de carré intégrable avec $M_0 = 0$. On suppose que, pour tout $1 \leq k \leq n$, on peut trouver des constantes $a_k < b_k$ telles que $a_k \leq \Delta M_k \leq b_k$ p.s. Alors, pour tout $x \geq 0$, on a*

$$\mathbb{P}(|M_n| \geq x) \leq 2 \exp\Big(-\frac{2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\Big). \tag{3.3}$$

Preuve du théorème 3.9. On adopte la même démarche que pour l'inégalité de Hoeffding. Tout d'abord, il est clair que pour tout $n \geq 1$, $M_n = M_{n-1} + \Delta M_n$. Pour tout $t > 0$, on a donc par passage à l'espérance conditionnelle

$$\mathbb{E}[\exp(tM_n)] = \mathbb{E}[\mathbb{E}[\exp(tM_n)|\mathcal{F}_{n-1}]] = \mathbb{E}[\exp(tM_{n-1})\mathbb{E}[\exp(t\Delta M_n)|\mathcal{F}_{n-1}]]. \tag{3.4}$$

Cependant, (M_n) est une martingale ce qui implique $\mathbb{E}[\Delta M_n|\mathcal{F}_{n-1}] = 0$. De plus, on a $a_n \leq \Delta M_n \leq b_n$ p.s. Il découle alors du lemme 2.2 que pour tout $t > 0$

$$\mathbb{E}[\exp(t\Delta M_n)|\mathcal{F}_{n-1}] \leq \exp\left(\frac{t^2}{8}(b_n - a_n)^2\right). \tag{3.5}$$

Par suite, on déduit de (3.4) et (3.5) que

$$\mathbb{E}[\exp(tM_n)] \leq \exp\left(\frac{t^2}{8}(b_n - a_n)^2\right)\mathbb{E}[\exp(tM_{n-1})]. \tag{3.6}$$

On tire alors de (3.6) que

$$\mathbb{E}[\exp(tM_n)] \leq \exp\left(\frac{t^2 v_n}{8}\right) \quad \text{avec} \quad v_n = \sum_{k=1}^{n}(b_k - a_k)^2 \tag{3.7}$$

car $M_0 = 0$. L'inégalité de Markov et (3.7) entraînent alors que pour tout $x \geq 0$, $t > 0$

$$\begin{aligned} \mathbb{P}(M_n \geq x) &\leq \exp(-tx)\mathbb{E}[\exp(tM_n)] \leq \exp\left(-tx + \frac{t^2 v_n}{8}\right), \\ &\leq \exp\left(-\frac{2x^2}{v_n}\right) \end{aligned}$$

avec $t = 4x/v_n$. En remplaçant M_n par $-M_n$, on montre de la même manière que pour tout $x \geq 0$

$$\mathbb{P}(M_n \leq -x) \leq \exp\left(-\frac{2x^2}{v_n}\right).$$

On peut donc conclure que pour tout $x \geq 0$

$$\mathbb{P}(|M_n| \geq x) = \mathbb{P}(M_n \geq x) + \mathbb{P}(M_n \leq -x) \leq 2\exp\left(-\frac{2x^2}{v_n}\right)$$

ce qui achève la preuve de l'inégalité de Azuma-Hoeffding. □

3.3. **Inégalité de Freedman.** On aborde l'inégalité de Freedman [13], à comparer avec l'inégalité de Bennett, dans laquelle entre en jeu le processus croissant $<M>_n$.

Théorème 3.10 (Inégalité de Freedman)**.** *Soit* (M_n) *une martingale de carré intégrable avec* $M_0 = 0$*. On suppose que, pour tout* $1 \leq k \leq n$*, il existe une constante* $c > 0$ *telle que* $\Delta M_k \leq c$ *p.s. Alors, pour tous* $x \geq 0$ *et* $y > 0$*, on a*

$$\mathbb{P}(M_n \geq x, <M>_n \leq y) \leq \exp\left(-\frac{x^2}{2(y + cx)}\right). \tag{3.8}$$

Remarque 3.11. *Soit* (M_n) *une martingale de carré intégrable avec* $M_0 = 0$. *On suppose que* (M_n) *vérifie la condition de Bernstein, c'est-à-dire que pour tout entier* $p \geq 2$, *il existe une constante* $c > 0$ *telle que*

$$\sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^p | \mathcal{F}_{k-1}] \leq \frac{p! c^{p-2}}{2} <M>_n .$$

Alors, pour tous $x \geq 0$ *et* $y > 0$, *on a*

$$\mathbb{P}(|M_n| \geq x, <M>_n \leq y) \leq 2 \exp\Big(-\frac{x^2}{2(y+cx)}\Big).$$

Des raffinements sur l'inégalité de Freedman se trouvent dans [7], [12] *et* [20].

La preuve de l'inégalité de Freedman repose sur le lemme suivant dans lequel on utilise à nouveau la fonction f définie par $f(0) = 1/2$ et pour tout $x \in \mathbb{R}^*$

$$f(x) = \frac{\exp(x) - 1 - x}{x^2}.$$

Lemme 3.12. *Soit* (M_n) *une martingale de carré intégrable avec* $M_0 = 0$ *telle que pour tout* $1 \leq k \leq n$, $\Delta M_k \leq c$ *p.s. avec* $c > 0$. *Pour tout* $t > 0$, *on pose*

$$V_n(t) = \exp\Big(tM_n - t^2 f(tc) <M>_n\Big).$$

Alors, $(V_n(t))$ *est une surmartingale positive avec* $\mathbb{E}[V_n(t)] \leq 1$.

Preuve. On reprend la même stratégie que dans la preuve du lemme 2.5. Tout d'abord, pour tout $t > 0$, on a

$$V_n(t) = V_{n-1}(t) \exp\Big(t\Delta M_n - t^2 f(tc) \Delta <M>_n\Big) \tag{3.9}$$

avec $\Delta <M>_n = \mathbb{E}[(\Delta M_n)^2 | \mathcal{F}_{n-1}]$. Cependant, on a $\mathbb{E}[\Delta M_n | \mathcal{F}_{n-1}] = 0$ et par hypothèse $\Delta M_n \leq c$ p.s. On tire alors du lemme 2.5 que pour tout $t > 0$

$$\begin{aligned} \mathbb{E}[\exp(t\Delta M_n)|\mathcal{F}_{n-1}] &\leq 1 + t^2 f(tc) \mathbb{E}[(\Delta M_n)^2 | \mathcal{F}_{n-1}], \\ &\leq \exp\Big(t^2 f(tc) \mathbb{E}[(\Delta M_n)^2 | \mathcal{F}_{n-1}]\Big) \end{aligned}$$

ce qui entraîne que

$$\mathbb{E}\Big[\exp\Big(t\Delta M_n - t^2 f(tc) \Delta <M>_n\Big) \Big| \mathcal{F}_{n-1}\Big] \leq 1. \tag{3.10}$$

Il découle alors de (3.9) et (3.10) que pour tout $t > 0$, $\mathbb{E}[V_n(t)|\mathcal{F}_{n-1}] \leq V_{n-1}(t)$. Par passage à l'espérance, on en déduit que $\mathbb{E}[V_n(t)] \leq \mathbb{E}[V_{n-1}(t)] \leq 1$ car $M_0 = 0$ et $<M>_0 = 0$. On peut conclure que, pour tout $t > 0$, $(V_n(t))$ est une surmartingale positive avec $\mathbb{E}[V_n(t)] \leq 1$. $\square$

Preuve du théorème 3.10. Pour tous $x \geq 0$ et $y > 0$, soit

$$A_n = \{M_n \geq x, <M>_n \leq y\}.$$

Par l'inégalité de Markov, on a pour tout $t > 0$

$$
\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2 f(tc)}{2} <M>_n\Big) \exp\Big(\frac{t^2 f(tc)}{2} <M>_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \exp\Big(\frac{t^2 f(tc) y}{2} - \frac{tx}{2}\Big)\mathbb{E}\Big[\sqrt{V_n(t)}\mathrm{I}_{A_n}\Big], \\
&\leq \exp\Big(\frac{t^2 f(tc) y}{2} - \frac{tx}{2}\Big)\sqrt{\mathbb{E}[V_n(t)]\mathbb{P}(A_n)}, \\
&\leq \exp\Big(\frac{t^2 f(tc) y}{2} - \frac{tx}{2}\Big)\sqrt{\mathbb{P}(A_n)} \qquad (3.11)
\end{aligned}
$$

via l'inégalité de Cauchy-Schwarz et le lemme 3.12. Par suite, en divisant chaque côté de (3.11) par $\sqrt{\mathbb{P}(A_n)}$, on en déduit que

$$\mathbb{P}(A_n) \leq \exp\Big(t^2 f(tc) y - tx\Big).$$

On minimise cette majoration en prenant

$$t = \frac{1}{c}\log\Big(1 + \frac{xc}{y}\Big)$$

ce qui conduit à

$$\mathbb{P}(A_n) \leq \exp\left(-\frac{y}{c^2}h\Big(\frac{xc}{y}\Big)\right) \tag{3.12}$$

avec $h(x) = (1+x)\log(1+x) - x$. Finalement, on a déjà vu que pour tout $x \geq 0$

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

L'inégalité (3.12) entraîne alors que

$$\mathbb{P}(A_n) \leq \exp\Big(-\frac{x^2}{2(y+cx)}\Big)$$

ce qui achève la preuve de l'inégalité de Freedman. □

3.4. **Inégalité de De la Peña.** Afin de s'affranchir de toute hypothèse de bornitude ou de moment, De la Peña [7] propose une inégalité exponentielle pour (M_n) conditionnellement symétrique, qui fait intervenir sa variation quadratique totale

$$[M]_n = \sum_{k=1}^{n} \Delta M_k^2.$$

Définition 3.13. *Soit (M_n) une martingale adaptée à $\mathbb{F} = (\mathcal{F}_n)$. On dit que (M_n) est conditionnellement symétrique si pour tout $n \geq 1$, la loi de ΔM_n sachant $\mathcal{F}_{n-1}$ est symétrique.*

Théorème 3.14 (Inégalité de De la Peña)**.** *Soit* (M_n) *une martingale de carré intégrable et conditionnellement symétrique avec* $M_0 = 0$*. Alors, pour tous* $x \geq 0$ *et* $y > 0$*, on a*

$$\mathbb{P}(M_n \geq x, [M]_n \leq y) \leq \exp\Big(-\frac{x^2}{2y}\Big). \tag{3.13}$$

Pour les martingales autonormalisées, on a également le résultat suivant.

Théorème 3.15. *Soit* (M_n) *une martingale de carré intégrable et conditionnellement symétrique avec* $M_0 = 0$*. Alors, pour tous* $x \geq 0$*,* $y > 0$ *et* $a \geq 0$*,* $b > 0$

$$\mathbb{P}\Big(\frac{M_n}{a + b[M]_n} \geq x\Big) \leq \sqrt{\mathbb{E}\Big[\exp\Big(-x^2\Big(ab + \frac{b^2}{2}[M]_n\Big)\Big)\Big]},$$

$$\mathbb{P}\Big(\frac{M_n}{a + b[M]_n} \geq x, [M]_n \geq y\Big) \leq \exp\Big(-x^2\Big(ab + \frac{b^2 y}{2}\Big)\Big).$$

Remarque 3.16. *On trouvera dans* [8]*,* [9]*,* [11] *une revue et des extensions de ces résultats. Par une lecture attentive de* [7]*, on peut réaliser que si* (M_n) *est conditionnellement symétrique, alors pour tous* $x \geq 0$ *et* $y > 0$

$$\mathbb{P}(|M_n| \geq x, [M]_n \leq y) \leq 2\exp\Big(-\frac{x^2}{2y}\Big). \tag{3.14}$$

Preuve. On trouvera dans la section suivante une preuve des théorèmes 3.14 et 3.15 sans l'hypothèse de symétrie associée à (M_n). □

3.5. **Martingales gaussiennes.** Afin d'obtenir une inégalité exponentielle pour (M_n) semblable à celle de De la Peña en remplaçant la variation quadratique $[M]_n$ par le processus croissant $<M>_n$, on doit supposer que (M_n) est conditionnellement gaussienne.

Définition 3.17. *Soit* (M_n) *une martingale adaptée à* $\mathbb{F} = (\mathcal{F}_n)$*. On dit que* (M_n) *est conditionnellement gaussienne si pour tout* $n \geq 1$*, la loi de* ΔM_n *sachant* $\mathcal{F}_{n-1}$ *est la loi* $\mathcal{N}(0, \Delta <M>_n)$*.*

Théorème 3.18. *Soit* (M_n) *une martingale de carré intégrable et conditionnellement gaussienne avec* $M_0 = 0$*. Alors, les résultats des théorèmes 3.14 et 3.15 sont vrais en remplaçant partout* $[M]_n$ *par* $<M>_n$*. De plus, pour tous* $x \geq 0$*,* $a \geq 0$ *et* $b > 0$*, on a*

$$\mathbb{P}\Big(\frac{M_n}{a + b <M>_n} \geq x\Big) \leq \inf_{p>1} \Big(\mathbb{E}\Big[\exp\Big(-(p-1)x^2\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\Big)^{1/p}. \tag{3.15}$$

Preuve du théorème 3.18. La preuve s'inspire du papier de De la Peña [7]. Tout d'abord, pour tous $n \geq 0$ et $t \in \mathbb{R}$, soit

$$W_n(t) = \exp\Big(tM_n - \frac{t^2}{2} <M>_n\Big).$$

Comme la martingale (M_n) est conditionnellement gaussienne, il est clair que, pour tout $t \in \mathbb{R}$, $(W_n(t))$ est une martingale positive avec $\mathbb{E}[W_n(t)] = 1$. Pour tous $x \geq 0$ et $y > 0$, soit

$$A_n = \{M_n \geq x, <M>_n \leq y\}.$$

Par l'inégalité de Markov, on a

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \mathbb{E}\Big[\sqrt{W_n(t)}\exp\Big(\frac{t^2}{4} <M>_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \exp\Big(\frac{t^2 y}{4} - \frac{tx}{2}\Big)\sqrt{\mathbb{P}(A_n)} \qquad (3.16)
\end{aligned}$$

via l'inégalité de Cauchy-Schwarz et l'égalité $\mathbb{E}[W_n(t)] = 1$. En divisant chaque côté de (3.16) par $\sqrt{\mathbb{P}(A_n)}$ et en choisissant $t = x/y$, on obtient

$$\mathbb{P}(A_n) \leq \exp\Big(-\frac{x^2}{2y}\Big).$$

On va continuer la preuve dans le cas particulier $a = 0$ et $b = 1$ car la démarche est exactement la même dans le cas général. Pour tout $x > 0$, soit

$$B_n = \{M_n \geq x <M>_n\}.$$

Par l'inégalité de Markov et l'inégalité de Holder, on a pour tout $t > 0$ et $q > 1$

$$\begin{aligned}
\mathbb{P}(B_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q} <M>_n\Big)\mathrm{I}_{B_n}\Big], \\
&\leq \mathbb{E}\Big[(W_n(t))^{1/q}\exp\Big(\frac{t}{2q}(t-2x) <M>_n\Big)\mathrm{I}_{B_n}\Big], \\
&\leq \Big(\mathbb{E}\Big[\exp\Big(\frac{tp}{2q}(t-2x) <M>_n\Big)\Big]\Big)^{1/p} \qquad (3.17)
\end{aligned}$$

car $\mathbb{E}[W_n(t)] = 1$ et $1/p + 1/q = 1$. Comme $p/q = p - 1$, on tire de (3.17) avec $t = x$

$$\mathbb{P}(B_n) \leq \inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2} <M>_n\Big)\Big]\Big)^{1/p}$$

ce qui entraîne immédiatement (3.15). Dans le cas particulier $p = 2$, on trouve que

$$\mathbb{P}(B_n) \leq \sqrt{\mathbb{E}\Big[\exp\Big(-\frac{x^2}{2} <M>_n\Big)\Big]}.$$

Finalement, pour tous $x \geq 0$ et $y > 0$, soit

$$C_n = \{M_n \geq x <M>_n, <M>_n \geq y\}.$$

En reprenant la preuve de (3.16), on a pour tout $0 < t < 2x$

$$\begin{aligned}
\mathbb{P}(C_n) &\leq \mathbb{E}\Big[\sqrt{W_n(t)}\exp\Big(\frac{t}{4}(t-2x) <M>_n\Big)\mathrm{I}_{C_n}\Big], \\
&\leq \exp\Big(\frac{ty}{4}(t-2x)\Big)\mathbb{E}\Big[\sqrt{W_n(t)}\mathrm{I}_{C_n}\Big], \\
&\leq \exp\Big(\frac{ty}{y}(t-2x)\Big)\sqrt{\mathbb{P}(C_n)}. \qquad (3.18)
\end{aligned}$$

En divisant chaque côté de (3.18) par $\sqrt{\mathbb{P}(C_n)}$, il advient avec $t = x$ que

$$\mathbb{P}(C_n) \leq \exp\Big(-\frac{x^2}{2y}\Big)$$

ce achève la preuve du théorème 3.18. □

Exercice 9. Processus autorégressif On considère le processus autorégressif

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}$$

où X_n et ε_n sont l'observation et le bruit associés au modèle. On suppose que (ε_n) est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ avec $\sigma^2 > 0$. On suppose également que l'état initial X_0 est indépendant de (ε_n) et de loi $\mathcal{N}(0, \tau^2)$ avec $\tau^2 \geq \sigma^2$. Le processus est dit stable si $|\theta| < 1$, instable si $|\theta| = 1$ et explosif si $|\theta| > 1$. Dans les trois cas, on estime θ par l'estimateur des moindres carrés

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=1}^n X_{k-1}^2}$$

L'objectif de cet exercice est de proposer une inégalité exponentielle simple pour $\widehat{\theta}_n$. On se place tout d'abord dans le cas stable avec $|\theta| < 1$.

1) Montrer que l'on a l'égalité

$$\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{<M>_n}$$

avec

$$M_n = \sum_{k=1}^n X_{k-1}\varepsilon_k \qquad \text{et} \qquad <M>_n = \sigma^2 \sum_{k=1}^n X_{k-1}^2.$$

2) Montrer par l'inégalité de Cauchy-Schwarz que

$$X_n^2 \leq \frac{1}{1-|\theta|}\Big(X_0^2 + \sum_{k=1}^n |\theta|^{n-k}\varepsilon_k^2\Big) \quad \text{et} \quad \sum_{k=0}^n X_k^2 \leq \Big(\frac{1}{1-|\theta|}\Big)^2\Big(X_0^2 + \sum_{k=1}^n \varepsilon_k^2\Big).$$

3) En déduire que presque sûrement

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = 0 \qquad \text{et} \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k^2 = \frac{\sigma^2}{1-\theta^2}.$$

4) Montrer par la LGN pour les martingales que

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta \qquad \text{p.s.}$$

5) Montrer également via le TLC pour les martingales que

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 1-\theta^2).$$

6) On n'impose maintenant aucune contrainte sur θ. Vérifier tout d'abord que (M_n) est une martingale conditionnellement gaussienne et en déduire que pour tous $n \geq 1$ et $x \geq 0$

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \inf_{p>1} \left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2\sigma^4} {<}M{>}_n\Big)\Big]\right)^{1/p}.$$

7) Montrer que pour tous $n \geq 1$ et $t < 0$, $\mathbb{E}[\exp(t {<}M{>}_n)] \leq (1-2\sigma^4 t)^{-n/2}$.

8) Pour tout $y > 0$, on pose

$$\ell(y) = \frac{\log(1+y)}{x^2+y}.$$

Déduire des questions 6) et 7) que pour tous $n \geq 1$ et $x \geq 0$

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \inf_{y>0} \exp\Big(-\frac{nx^2}{2}\ell(y)\Big).$$

9) Si $h(y) = (1+y)\log(1+y) - y$, vérifier que

$$\ell'(y) = \frac{x^2 - h(y)}{(1+y)(x^2+y)^2}.$$

10) Conclure que pour tous $n \geq 1$ et $x > 0$

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \exp\Big(-\frac{nx^2}{2(1+y_x)}\Big)$$

où y_x est l'unique solution positive de l'équation $h(y_x) = x^2$. En particulier, vérifier que si $0 < x < 1/2$

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \exp\Big(-\frac{nx^2}{2(1+2x)}\Big).$$

4. Toujours un peu plus loin

On propose tout d'abord une inégalité exponentielle sans aucune hypothèse sur (M_n). On parlera ensuite d'une nouvelle notion en probabilités de variables aléatoires lourdes à gauche ou à droite [6], ce qui va nous permettre de généraliser les inégalités de De la Peña.

Théorème 4.1. *Soit (M_n) une martingale de carré intégrable avec $M_0 = 0$. Alors, pour tous $x \geq 0$ et $y > 0$, on a*

$$\mathbb{P}\Big(|M_n| \geq x, [M]_n + <M>_n \leq y\Big) \leq 2\exp\Big(-\frac{x^2}{2y}\Big). \tag{4.1}$$

Pour les martingales autonormalisées, on a également le résultat suivant.

Théorème 4.2. *Soit (M_n) une martingale de carré intégrable avec $M_0 = 0$. Alors, pour tous $x \geq 0$, $y > 0$ et $a \geq 0$, $b > 0$*

$$\mathbb{P}\left(\frac{|M_n|}{a + b <M>_n} \geq x, <M>_n \geq [M]_n + y\right) \leq 2\exp\left(-x^2\Big(ab + \frac{b^2 y}{2}\Big)\right),$$

$$\mathbb{P}\left(\frac{|M_n|}{a + b <M>_n} \geq x, [M]_n \leq y <M>_n\right)$$

$$\leq 2 \inf_{p>1} \left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{(1+y)}\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\right)^{1/p}.$$

La preuve des théorèmes 4.1 et 4.2 repose sur le lemme suivant.

Lemme 4.3. *Pour tout $x \in \mathbb{R}$, on a*

$$f(x) = \exp\Big(x - \frac{x^2}{2}\Big) \leq g(x) = 1 + x + \frac{x^2}{2}. \tag{4.2}$$

De plus, soit X une variable aléatoire réelle centrée et de carré intégrable, de variance $\sigma^2 > 0$. Pour tout $t \in \mathbb{R}$, on pose

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2}{2}X^2\Big)\Big]. \tag{4.3}$$

Alors, pour tout $t \in \mathbb{R}$, on a

$$L(t) \leq 1 + \frac{t^2}{2}\sigma^2. \tag{4.4}$$

Remarque 4.4. *L'inégalité (4.2), illustrée par la figure 4, est également utilisée par Barlow, Jacka et Yor* [2] *ainsi que par Dzhaparidze et Van Zanten* [11] *pour l'obtention d'inégalités exponentielles pour les martingales à temps continu.*

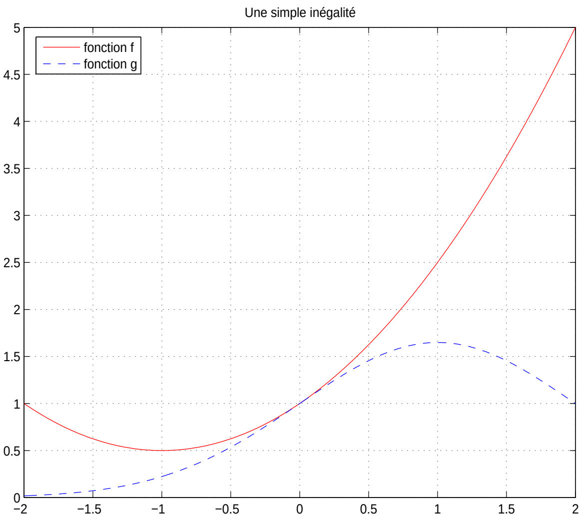

FIG. 4. Comparaison des fonctions f et g

Preuve. On va montrer que pour tout $x \geq 0$

$$\exp\Big(x - \frac{x^2}{2}\Big) \leq 1 + x$$

tandis que pour tout $x \leq 0$

$$\exp(x) \leq 1 + x + \frac{x^2}{2}$$

ce qui entraînera bien sûr (4.2). Pour tout $x \geq 0$, soit

$$h(x) = 1 + x - \exp\Big(x - \frac{x^2}{2}\Big).$$

On a

$$h'(x) = 1 + (1-x)\exp\Big(x - \frac{x^2}{2}\Big) \quad \text{et} \quad h''(x) = x(2-x)\exp\Big(x - \frac{x^2}{2}\Big).$$

Il est clair que la fonction h' est minimum en 0 et comme $h'(0) = 0$, $h'(x) \geq 0$ pour tout $x \geq 0$. La fonction h est donc croissante et comme $h(0) = 0$, on a pour tout $x \geq 0$, $h(x) \geq 0$. De plus, pour tout $x \leq 0$, soit

$$\ell(x) = 1 + x + \frac{x^2}{2} - \exp(x).$$

On a $\ell'(x) = 1 + x - \exp(x)$ et $\ell''(x) = 1 - \exp(x)$. Pour tout $x \leq 0$, $\ell''(x) \geq 0$. La fonction ℓ est convexe et minimum en 0 et comme $\ell(0) = 0$, $\ell(x) \geq 0$ pour tout $x \leq 0$, ce qui achève la preuve de l'inégalité (4.2). Il découle immédiatement de (4.2) que pour tout $t \in \mathbb{R}$

$$L(t) = \mathbb{E}[f(tX)] \leq \mathbb{E}[g(tX)] = 1 + \frac{t^2}{2}\sigma^2.$$

Preuve des théorèmes 4.1 et 4.2. Tout d'abord, pour tous $n \geq 0$ et $t \in \mathbb{R}$, soit

$$V_n(t) = \exp\Big(tM_n - \frac{t^2}{2}\Big([M]_n + < M >_n\Big)\Big).$$

Pour tout $t \in \mathbb{R}$, $(V_n(t))$ est une surmartingale positive avec $\mathbb{E}[V_n(t)] \leq 1$ car

$$V_n(t) = V_{n-1}(t)\exp\Big(t\Delta M_n - \frac{t^2}{2}\Big(\Delta[M]_n + \Delta < M >_n\Big)\Big)$$

donc

$$\mathbb{E}[V_n(t)|\mathcal{F}_{n-1}] \leq V_{n-1}(t)\Big(1 + \frac{t^2}{2}\Delta < M >_n\Big)\exp\Big(-\frac{t^2}{2}\Delta < M >_n\Big) \leq V_{n-1}(t)$$

via l'inégalité élémentaire $(1+x) \leq \exp(x)$. Il en découle que, pour tout $t \in \mathbb{R}$, $(V_n(t))$ est une surmartingale positive avec $\mathbb{E}[V_n(t)] \leq 1$. On pose

$$Z_n = [M]_n + < M >_n .$$

Pour tous $x \geq 0$ et $y > 0$, soit

$$A_n = \{|M_n| \geq x, Z_n \leq y\}.$$

On a $A_n = A_n^+ \cup A_n^-$ avec $A_n^+ = \{M_n \geq x, Z_n \leq y\}$ et $A_n^- = \{M_n \leq -x, Z_n \leq y\}$. Par l'inégalité de Markov, on a pour tout $t > 0$

$$\begin{aligned}
\mathbb{P}(A_n^+) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n^+}\Big], \\
&\leq \mathbb{E}\Big[\sqrt{V_n(t)}\exp\Big(\frac{t^2}{4}Z_n - \frac{tx}{2}\Big)\mathrm{I}_{A_n^+}\Big], \\
&\leq \exp\Big(\frac{t^2 y}{4} - \frac{tx}{2}\Big)\sqrt{\mathbb{P}(A_n^+)}
\end{aligned}$$

car $\mathbb{E}[V_n(t)] \leq 1$. En divisant de chaque côté par $\sqrt{\mathbb{P}(A_n^+)}$ et en choisissant $t = x/y$, on obtient

$$\mathbb{P}(A_n^+) \leq \exp\Big(-\frac{x^2}{2y}\Big).$$

On trouve également la même majoration pour $\mathbb{P}(A_n^-)$ ce qui termine la preuve du théorème 4.1. La preuve du théorème 4.2 est laissé au lecteur car elle est semblable à celle du théorème 3.18. □

4.1. **Variables aléatoires lourdes à gauche ou à droite.** On va maintenant introduire la notion de variables aléatoires lourdes à gauche ou à droite [6] qui va nous permettre d'affiner le lemme 4.3.

Définition 4.5. *On dit qu'une variable aléatoire X est lourde à gauche si $\mathbb{E}[X] = 0$ et, pour tout $a > 0$, $\mathbb{E}[T_a(X)] \leq 0$ où $T_a(X) = \min(|X|, a)sign(X)$ est la version tronquée de X*

$$T_a(X) = \begin{cases} a & si \quad X \geq a, \\ X & si \quad -a \leq X \leq a, \\ -a & si \quad X \leq -a. \end{cases}$$

De plus, X est lourde à droite si $-X$ est lourde à gauche.

Exercice 10. Lourde à gauche, droite ou symétrique Soit X une variable aléatoire de fonction de répartition F.

1) Montrer en intégrant par parties que $\mathbb{E}[T_a(X)] = -H(a)$ où H est la fonction definie pour tout $a > 0$ par
$$H(a) = \int_0^a F(-x) - (1 - F(x))\, dx.$$
2) En déduire que X est lourde à gauche si $\mathbb{E}[X] = 0$ et pour tout $a > 0$
$$H(a) \geq 0.$$
3) Vérifier que X est symétrique si et seulement si X est lourde à gauche et à droite.

Exercice 11. Variables discrètes. Soit Y une variable aléatoire positive discrète d'espérance m et soit
$$X = Y - m.$$

1) Si Y est de loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$, montrer que X est lourde à gauche, lourde à droite, ou symétrique suivant que $p < 1/2$, $p > 1/2$, ou $p = 1/2$, respectivement.
2) Si Y est de loi géometrique $\mathcal{G}(p)$ avec $0 < p < 1$, montrer que X est toujours lourde à gauche.
3) Si Y est de loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$, montrer que X est lourde à gauche dès que
$$2 \exp(-\lambda) \sum_{k=0}^{[\lambda]} \frac{\lambda^k}{k!} \geq 1.$$

Exercice 12. Variables continues. Soit Y une variable aléatoire positive absolument continue d'espérance m et soit
$$X = Y - m.$$

1) Si Y est de loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$, montrer que X est toujours lourde à gauche.
2) Si Y est de loi Gamma $\mathcal{G}(a, \lambda)$ avec $a, \lambda > 0$, montrer que X est toujours lourde à gauche.
3) Si Y est de loi de Paréto $\mathcal{P}(a, \lambda)$ avec $a, \lambda > 0$, c'est-à-dire $Y = a \exp(Z)$ où Z est de loi exponentielle $\mathcal{E}(\lambda)$, montrer que X est toujours lourde à gauche.
4) Si Y est de loi log-normale $\mathcal{LN}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma^2 > 0$, c'est-à-dire $Y = \exp(Z)$ où Z est de loi $\mathcal{N}(m, \sigma^2)$, montrer que X est toujours lourde à gauche.

Le lemme suivant est la clef de voûte de tout ce qui va suivre.

Lemme 4.6. *Soit X une variable aléatoire réelle et pour tout $t \in \mathbb{R}$, soit*

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2}{2}X^2\Big)\Big].$$

1) *Si X est lourde à gauche alors pour tout $t \geq 0$, $L(t) \leq 1$.*
2) *Si X est lourde à droite alors pour tout $t \leq 0$, $L(t) \leq 1$.*
3) *Si X est symétrique alors pour tout $t \in \mathbb{R}$, $L(t) \leq 1$.*

Preuve. Soit f la fonction définie pour tout $x \in \mathbb{R}$ par

$$f(x) = \exp\Big(x - \frac{x^2}{2}\Big).$$

On a $f'(x) = (1-x)f(x)$ et $f'(-x) = (1+x)f(-x)$. On pose

$$a(x) = f'(-x) \qquad \text{et} \qquad b(x) = f'(-x) - f'(x).$$

Il est facile de vérifier que pour tout $x > 0$, $0 < a(x) < 1$, $0 < b(x) < 2$ et $a'(x) < 0$ car $a'(x) = -(2x + x^2)f(-x)$. Pour tout $t \in \mathbb{R}$, on a

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2X^2}{2}\Big)\Big] = \int_{\mathbb{R}} f(tx)\, dF(x)$$

où F est la loi associée à X. En intégrant par parties, on a pour tout $t \in \mathbb{R}$

$$\begin{aligned}
L(t) &= \Big[f(tx)F(x)\Big]_{-\infty}^{+\infty} - t\int_{\mathbb{R}} f'(tx)F(x)\, dx = -t\int_{\mathbb{R}} f'(tx)F(x)\, dx, \\
&= -t\int_{-\infty}^{0} f'(tx)F(x)\, dx - t\int_{0}^{+\infty} f'(tx)F(x)\, dx, \\
&= -t\int_{0}^{+\infty} f'(-tx)F(-x)\, dx - t\int_{0}^{+\infty} f'(tx)F(x)\, dx. \qquad (4.5)
\end{aligned}$$

Cependant, on a

$$t\int_0^{\infty} f'(tx)\, dx = \Big[\exp\Big(tx - \frac{t^2x^2}{2}\Big)\Big]_0^{+\infty} = -1.$$

On tire alors de (4.5) que pour tout $t \in \mathbb{R}$

$$1 - L(t) = tI(t)$$

où

$$\begin{aligned}
I(t) &= \int_0^{+\infty} f'(-tx)F(-x)dx - \int_0^{+\infty} f'(tx)(1-F(x))dx, \\
&= \int_0^{+\infty} f'(-tx)F(-x)dx - \int_0^{+\infty} (f'(tx) - f'(-tx) + f'(-tx))(1-F(x))dx, \\
&= \int_0^{+\infty} f'(-tx)(F(-x) - (1-F(x)))dx + \int_0^{+\infty} (f'(-tx) - f'(tx))(1-F(x))dx, \\
&= A(t) + B(t)
\end{aligned}$$

avec

$$\begin{aligned} A(t) &= \int_0^{+\infty} a(tx)(F(-x) - (1 - F(x)))\,dx, \\ B(t) &= \int_0^{+\infty} b(tx)(1 - F(x))\,dx. \end{aligned}$$

On suppose dans un premier temps que X est lourde à gauche. On doit montrer que, pour tout $t > 0$, $I(t) \geq 0$. On a clairement $B(t) \geq 0$ car $b(tx) > 0$. De plus, pour tout $a > 0$, soit

$$H(a) = \int_0^a F(-x) - (1 - F(x))\,dx.$$

Comme $H'(a) = F(-a) - (1 - F(x))$ presque partout, on a en intégrant par parties

$$\begin{aligned} A(t) &= \int_0^{+\infty} a(tx)H'(x)\,dx = \Big[a(tx)H(x)\Big]_0^{+\infty} - \int_0^{+\infty} ta'(tx)H(x)\,dx, \\ &= -t\int_0^{+\infty} a'(tx)H(x)\,dx \qquad (4.6) \end{aligned}$$

car $H(0) = 0$ et

$$\lim_{a\to\infty} H(a) = -\mathbb{E}[X] = 0.$$

D'autre part, on a déjà vu que pour tout $a \geq 0$, $H(a) \geq 0$ et pour tout $x > 0$, $a'(x) < 0$. Par conséquent, on déduit de (4.6) que pour tout $t > 0$, $A(t) \geq 0$. La relation $I(t) = A(t)+B(t)$ entraîne alors comme attendu que $I(t) \geq 0$ donc $L(t) \leq 1$ pour tout $t > 0$, ce qui achève la preuve du lemme 4.6 partie 1). Ensuite, si X est lourde à droite, $-X$ est lourde à gauche. En remplaçant X par $-X$, on trouve par le lemme 4.6 partie 1) que $L(t) \leq 1$ for tout $t < 0$. Finalement, le lemme 4.6 partie 3) découle immédiatement des parties 1) et 2). On peut aussi le voir directement car si X est symétrique, alors pour tout $t \in \mathbb{R}$

$$\begin{aligned} L(t) &= \int_{\mathbb{R}} f(tx)\,dF(x) = \int_0^{+\infty} (f(tx) + f(-tx))\,dF(x), \\ &= 2\int_0^{+\infty} \exp(-t^2x^2/2)\cosh(tx)\,dF(x) \leq 1 \end{aligned}$$

car, pour tout $x \in \mathbb{R}$, $\cosh(x) \leq \exp(x^2/2)$. □

4.2. **Inégalités exponentielles pour les martingales lourdes à gauche.** On va maintenant appliquer aux martingales la notion de variables aléatoires lourdes à gauche ou à droite.

Définition 4.7. *Soit (M_n) une martingale de carré intégrable adaptée à la filtration $\mathbb{F} = (\mathcal{F}_n)$. On dit que (M_n) est lourde à gauche si tous ses accroissements sont conditionnellement lourds à gauche. En d'autres termes, pour tous $n \geq 1$ et $a > 0$, $\mathbb{E}[T_a(\Delta M_n)|\mathcal{F}_{n-1}] \leq 0$. De plus, (M_n) est lourde à droite si $(-M_n)$ est lourde à gauche.*

Notre démarche est très simple : si (M_n) est lourde à gauche, on peut espérer obtenir une inégalité exponentielle à droite tandis que si (M_n) est lourde à droite, on doit pouvoir montrer une inégalité exponentielle à gauche. On va ainsi retrouver les inégalités de De la Peña [7] sous la seule contrainte que (M_n) soit lourde à gauche.

Théorème 4.8. *Soit (M_n) une martingale de carré intégrable, lourde à gauche avec $M_0 = 0$. Alors, pour tous $x \geq 0$ et $y > 0$, on a*

$$\mathbb{P}\Big(M_n \geq x, [M]_n \leq y\Big) \leq \exp\Big(-\frac{x^2}{2y}\Big). \tag{4.7}$$

Pour les martingales autonormalisées, on a aussi le résultat suivant.

Théorème 4.9. *Soit (M_n) une martingale de carré intégrable, lourde à gauche avec $M_0 = 0$. Alors, pour tous $x \geq 0$, $y > 0$ et $a \geq 0$, $b > 0$*

$$\mathbb{P}\left(\frac{M_n}{a + b[M]_n} \geq x\right) \leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)x^2\Big(ab + \frac{b^2}{2}[M]_n\Big)\Big)\Big]\right)^{1/p},$$

$$\mathbb{P}\left(\frac{M_n}{a + b[M]_n} \geq x, [M]_n \geq y\right) \leq \exp\left(-x^2\Big(ab + \frac{b^2 y}{2}\Big)\right).$$

De plus, on a

$$\mathbb{P}\left(\frac{M_n}{a + b <M>_n} \geq x, [M]_n \leq y <M>_n\right)$$

$$\leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{y}\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\right)^{1/p}.$$

Preuve des théorèmes 4.8 et 4.9. Pour tous $n \geq 0$ et $t \in \mathbb{R}$, soit

$$W_n(t) = \exp\Big(tM_n - \frac{t^2}{2}[M]_n\Big).$$

Il découle clairement du lemme 4.6 que si (M_n) est lourde à gauche alors pour tout $t \geq 0$, $(W_n(t))$ est une surmartingale positive avec $\mathbb{E}[W_n(t)] \leq 1$. De plus, si (M_n) est lourde à droite alors pour tout $t \leq 0$, $(W_n(t))$ est une surmartingale positive avec $\mathbb{E}[W_n(t)] \leq 1$. Enfin, si (M_n) est conditionnellement symétrique alors pour tout $t \in \mathbb{R}$, $(W_n(t))$ est une surmartingale positive avec $\mathbb{E}[W_n(t)] \leq 1$. La preuve du théorème 4.8 est laissé au lecteur car elle est exactement la même que celle du théorème 4.1 en remplaçant $(V_n(t))$ par $(W_n(t))$. On va ensuite montrer le théorème 4.9 dans le cas particulier $a = 0$ et $b = 1$ car la démarche est la même dans le cas général. Pour tout $x \geq 0$, soit

$$A_n = \{M_n \geq x[M]_n\}.$$

Par l'inégalité de Holder, on a pour tous $t > 0$ et $q > 1$

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q}[M]_n\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{t^2}{2q}[M]_n\Big)\exp\Big(\frac{t}{2q}(t-2x)[M]_n\Big)\mathrm{I}_{A_n}\Big], \\
&\leq \Big(\mathbb{E}\Big[\exp\Big(\frac{tp}{2q}(t-2x)[M]_n\Big)\Big]\Big)^{1/p} \qquad (4.8)
\end{aligned}$$

car (M_n) est lourde à gauche donc, pour tout $t \geq 0$, $\mathbb{E}[W_n(t)] \leq 1$. L'inégalité (4.8) entraîne avec $p/q = p-1$ et $t = x$ que

$$\mathbb{P}(A_n) \leq \inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2}[M]_n\Big)\Big]\Big)^{1/p}.$$

Ensuite, pour tous $x \geq 0$ et $y > 0$, soit

$$B_n = \{M_n \geq x[M]_n, [M]_n \geq y\}.$$

Comme précédemment, on trouve que pour tout $0 < t < 2x$

$$\begin{aligned}
\mathbb{P}(B_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}[M]_n\Big)\exp\Big(\frac{t}{4}(t-2x)[M]_n\Big)\mathrm{I}_{B_n}\Big], \\
&\leq \exp\Big(\frac{ty}{4}(t-2x)\Big)\mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}[M]_n\Big)\mathrm{I}_{B_n}\Big], \\
&\leq \exp\Big(\frac{ty}{4}(t-2x)\Big)\sqrt{\mathbb{P}(B_n)}, \\
&\leq \exp\Big(-\frac{x^2y}{2}\Big)
\end{aligned}$$

en choisissant la valeur $t = x$. Finalement, la dernière inégalité du théorème 4.9 est laissée au lecteur. □

Exercice 13. Processus de Galton-Watson. Le processus de Galton-Watson sert à modéliser la dynamique d'une population. On part d'un unique ancêtre $X_0 = 1$. Il peut donner naissance à $1, 2, \ldots$ enfants qui constituent la première génération. Chaque individu de la première génération peut donner naissance à $1, 2, \ldots$ enfants et l'ensemble de tous les descendants forme la seconde génération et ainsi de suite. Le nombre d'individus X_n de la n^e génération est donc donné, pour tout $n \geq 1$, par

$$X_n = \sum_{k=1}^{X_{n-1}} Y_{n,k}$$

où $Y_{n,k}$ correspond au nombre d'enfants du k^e individu de la $(n-1)^e$ génération. On suppose que pour tout $n \geq 1$, $Y_{n,k}$ et X_{n-1} sont indépendantes. On suppose également que $(Y_{n,k})$ est une suite de variables aléatoires indépendantes et de même loi, à valeurs dans $\mathbb{N}^*$, d'espérance m et de variance $\sigma^2 > 0$. La loi de $(Y_{n,k})$ est appelée loi de reproduction. On se place dans le cas surcritique avec $m > 1$ que l'on estime par l'estimateur de Lotka-Nagaev

$$\widehat{m}_n = \frac{X_n}{X_{n-1}}.$$

1) Montrer que le processus de Galton-Watson peut s'ecrire sous la forme autorégressive
$$X_{n+1} = mX_n + \varepsilon_{n+1}$$
avec $\mathbb{E}[\varepsilon_{n+1}|\mathcal{F}_n] = 0$ et $\mathbb{E}[\varepsilon_{n+1}^2|\mathcal{F}_n] = \sigma^2 X_n$. On pose alors
$$M_n = \frac{X_n}{m^n}.$$
2) Montrer que (M_n) est une martingale bornée dans $\mathbb{L}^2$ qui converge presque sûrement et dans $\mathbb{L}^2$ vers une variable aléatoire $M > 0$ dont on précisera l'espérance et la variance.
3) En déduire que
$$\lim_{n\to\infty} \widehat{m}_n = m \qquad \text{p.s.}$$
4) Soit L la log-Laplace de la loi de reproduction recentrée, donnée pour tout $t \in \mathbb{R}$ par $L(t) = \log \mathbb{E}[\exp(t(Y_{n,k} - m))]$. Vérifier que pour tout $t \in \mathbb{R}$
$$\mathbb{E}[\exp(t\varepsilon_{n+1})|\mathcal{F}_n] = \exp(L(t)X_n).$$
5) On suppose que L est finie sur un intervalle $[-c, c]$ avec $c > 0$. On note I sa transformée de Legendre
$$I(x) = \sup_{|t|\leq c}\{xt - L(t)\}$$
et $J(x) = \min(I(x), I(-x))$. Montrer que pour tous $n \geq 1$ et $x \geq 0$
$$\mathbb{P}(|\widehat{m}_n - m| \geq x) \leq 2\mathbb{E}[\exp(-J(x)X_{n-1})].$$
6) Dans le cas particulier où la loi de reproduction est la loi géométrique $\mathcal{G}(p)$ avec $0 < p < 1$, conclure que pour tous $n \geq 1$ et $x \geq 0$
$$\mathbb{P}(|\widehat{m}_n - m| \geq x) \leq \frac{2p^n \exp(-J(x))}{p(1 - \exp(-J(x)))}.$$

Exercice 14. Estimateurs à noyau. Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, de densité de probabilité $f \in \mathcal{C}^1(\mathbb{R})$ à dérivée bornée. Soit K une fonction positive bornée, appelée noyau, telle que

$$\int_{\mathbb{R}} K(x)\, dx = 1, \qquad \int_{\mathbb{R}} xK(x)\, dx = 0, \qquad \int_{\mathbb{R}} K^2(x)\, dx = \sigma^2.$$

Il est également classique de supposer la convergence des intégrales

$$\int_{\mathbb{R}} |x|K(x)\, dx, \qquad \int_{\mathbb{R}} x^2K(x)\, dx, \qquad \int_{\mathbb{R}} |x|K^2(x)\, dx.$$

On peut par exemple choisir le noyau uniforme $K(x) = 1/(2a)\mathrm{I}_{(|x|\leq a)}$ avec $a > 0$, ou bien le noyau d'Epanechnikov $K(x) = 3/(4b)(1 - x^2/b^2)\mathrm{I}_{(|x|\leq b)}$ avec $b > 0$, ou encore le noyau gaussien $K(x) = \exp(-x^2/2)/\sqrt{2\pi}$. On estime f par l'estimateur à noyau $\widehat{f}_n$ défini pour tout $x \in \mathbb{R}$ par

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\Big(\frac{X_i - x}{h_i}\Big)$$

où $h_i = 1/i^\alpha$ avec $0 < \alpha < 1$. Pour tout $1 \leq i \leq n$, on pose

$$\varepsilon_i(x) = \frac{1}{h_i}K\Big(\frac{X_i - x}{h_i}\Big).$$

1) Montrer que pour tout $x \in \mathbb{R}$, on a l'égalité

$$\widehat{f}_n(x) - f(x) = \frac{M_n(x)}{n} + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\varepsilon_i(x)] - f(x)$$

où $(M_n(x))$ est la martingale de carré intégrable donnée par

$$M_n(x) = \sum_{i=1}^{n}\varepsilon_i(x) - \mathbb{E}[\varepsilon_i(x)] \quad \text{et} \quad < M(x) >_n = \sum_{i=1}^{n}\mathrm{Var}(\varepsilon_i(x)).$$

2) Vérifier par le théorème des accroissements finis que

$$\lim_{n\to\infty}\frac{< M(x) >_n}{n^{1+\alpha}} = \frac{\sigma^2 f(x)}{1+\alpha} \qquad \text{p.s.}$$

3) En déduire par la LGN pour les martingales que

$$\lim_{n\to\infty}\widehat{f}_n(x) = f(x) \qquad \text{p.s.}$$

4) Si $1/5 < \alpha < 1$, montrer également via le TLC pour les martingales que

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2 f(x)}{1+\alpha}\Big).$$

5) Trouver une inégalité exponentielle associée à cet estimateur à noyau.

Références

[1] Azuma, K. Weighted sums of certain dependent random variables, Tôkuku Mathematical Journal, vol. 19, p. 357-367, 1967.

[2] Barlow, M. T., Jacka, S. D. and Yor, M. Inequalities for a pair of processes stopped at a random time, Proc. London Math. Soc., vol. 52, p. 142-172, 1986.

[3] Bennett, G. Probability inequalities for the sum of independent random variables, Journal of the Amer. Statist. Assoc., vol. 57, p. 33-45, 1962.

[4] Bercu, B. On large deviations in the Gaussian autoregressive process : stable, unstable and explosive cases, Bernoulli, vol. 7, p. 299-316, 2001.

[5] Bercu, B et Chafai, D. Modélisation stochastique et simulation, Mathématiques appliquées pour le Master, SMAI, Dunod, 2007.

[6] Bercu, B. and Touati, A. Exponential inequalities for self-normalized martingales with applications, to appear in Annals of Applied Probability, 18, 2008.

[7] De la Peña, V. H. A general class of exponential inequalities for martingales and ratios. Annals of Probability, vol. 27, p. 537-564, 1999.

[8] De la Peña, V. H., Klass, M. J. and Lai, T. L. Self-normalized processes : exponential inequalities, moments bounds and iterated logarithm law, Annals of Probability, vol. 32, p. 1902-1933, 2004.

[9] De la Peña, V. H., Klass, M. J. and Lai, T. L. Pseudo-maximization and self-normalized processes, Probability Surveys, vol. 4, p. 172-192, 2007.

[10] Duflo, M. Random Iterative Models, Springer Verlag, Berlin, 1997.

[11] Dzhaparidze, K. and Van Zanten, J. H. On Bernstein-type inequalities for martingales. Stochastic Processes and their Applications, vol. 93, p. 109-117, 2001.

[12] Van de Geer, S. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Annals of Statistics, vol. 23, p. 1779-1801, 1995.

[13] Freedman, D. A. On tail probabilities for martingales. Annals of Probability, vol. 3, p. 100-118, 1975.

[14] Hall, P. and Heyde, C. C. Martingale limit theory and its application, Academic Press, New York, 1980.

[15] Hoeffding, W. J. Probability inequalities sums of bounded random variables, Journal of the Amer. Statist. Assoc., vol. 58, p. 713-721, 1963.

[16] Liptser, R. and Spokoiny, V. Deviation probability bound for martingales with applications to statistical estimation. Statistic and Probability Letters, vol. 46, p. 347-357, 2000.

[17] Mazliak, L., Priouret, P. et Baldi, P. Martingales et chaînes de Markov, Hermann, Paris, 1998.

[18] McDiarmid, C. Concentration, in Probabilistic methods for algorithmic discrete mathematics, Springer-Verlag, Berlin, p. 195-248, 1998.

[19] Neveu, J. Martingales à temps discret, Masson, Paris, 1972.

[20] Pinelis, I. Optimum bounds for the distributions of martingales in Banach spaces, Annals of Probability, vol. 22, p. 1679-1706, 1994.