Inégalité de Hoeffding et application [1]

Leçons : 253, 260, 261, 262, *229*

[Ouv2], exercice 10.11

Théorème

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes et centrées.
De plus, on suppose que $\forall n \in \mathbb{N}^*$, X_n est ps bornée par c_n, où $c_n > 0$.
On note : $\forall n \in \mathbb{N}^*, S_n = \sum_{j=1}^{n} X_j$ et $a_n = \sum_{j=1}^{n} c_j^2$.
Alors $\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.

Démonstration :

→ Il va s'agir de démontrer le lemme qui suit.

Lemme

Soit X une variable aléatoire réelle, centrée, et ps bornée par 1.
On note L_X sa transformée de Laplace.
On a : $\forall t \in \mathbb{R}, L_X(t) \leqslant \exp\left(\frac{t^2}{2}\right)$.

Démonstration :

Soit $t \in \mathbb{R}$, on a : $\forall x \in [-1,1], tx = \frac{1-x}{2}(-t) + \frac{1+x}{2}t$.
Donc, par convexité de l'exponentielle, on en déduit : $\forall x \in [-1,1], \mathrm{e}^{tx} \leqslant \frac{1-x}{2}\mathrm{e}^{-t} + \frac{1+x}{2}\mathrm{e}^{t}$.
Mais on sait que $|X| \leqslant 1$ ps ; en particulier, e^{tX} est bornée ps donc admet un moment d'ordre 1.
Ainsi, L_X est bien définie en t et :

$$L_X(t) \leqslant \frac{\mathbb{E}[1-X]}{2}\mathrm{e}^{-t} + \frac{\mathbb{E}[1+X]}{2}\mathrm{e}^{t} = \frac{1}{2}\left(\mathrm{e}^{-t}+\mathrm{e}^{t}\right) = \operatorname{ch} t = \sum_{n=0}^{\infty}\frac{t^{2n}}{(2n)!} \leqslant \sum_{n=0}^{\infty}\frac{t^{2n}}{n!2^n} = \exp\left(\frac{t^2}{2}\right).$$ ■

→ Fixons $n \in \mathbb{N}^*$.
On applique le lemme aux variables $\frac{X_j}{c_j}$, où $j \in [\![1,n]\!]$: $\forall t \in \mathbb{R}, L_{X_j}(t) = L_{\frac{X_j}{c_j}}(tc_j) \leqslant \exp\left(\frac{t^2}{2}c_j^2\right)$.
Mais par indépendance des $\exp(tX_j)$ pour $j \in [\![1,n]\!]$, il vient :

$$\forall t \in \mathbb{R}, L_{S_n}(t) = \prod_{j=1}^{n} L_{X_j}(t) \leqslant \exp\left(\frac{t^2}{2}a_n\right).$$

→ Soient $t > 0$ et $\varepsilon > 0$; on a : $S_n > \varepsilon \Leftrightarrow \mathrm{e}^{tS_n} > \mathrm{e}^{t\varepsilon}$.
Ainsi, par l'inégalité de Markov, on obtient :

$$\mathbb{P}(S_n > \varepsilon) = \mathbb{P}\left(\mathrm{e}^{tS_n} > \mathrm{e}^{t\varepsilon}\right) \leqslant \frac{\mathbb{E}\left[\mathrm{e}^{tS_n}\right]}{\mathrm{e}^{t\varepsilon}} = \mathrm{e}^{-t\varepsilon}L_{S_n}(t) \leqslant \exp\left(-t\varepsilon + \frac{t^2}{2}a_n\right).$$

Cette inégalité est vraie pour tout $t > 0$, donc en particulier pour $t = \frac{\varepsilon}{a_n}$, où $-t\varepsilon + \frac{t^2}{2}a_n$ réalise son minimum, d'où : $\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(\frac{\varepsilon^2}{a_n^2}\frac{a_n}{2} - \frac{\varepsilon}{a_n}\varepsilon\right) = \exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.

→ Soit $\varepsilon > 0$ quelconque.
On a : $\mathbb{P}(|S_n| > \varepsilon) = \mathbb{P}(S_n > \varepsilon) + \mathbb{P}(S_n < -\varepsilon)$.
Mais on aurait pu appliquer tout ce qu'on vient de faire aux variables $-X_j$, où $j \in [\![1,n]\!]$, et donc :
$\mathbb{P}(S_n < -\varepsilon) = \mathbb{P}(-S_n > \varepsilon) \leqslant \exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.
Et finalement, $\mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2a_n}\right)$. ■

Corollaire

Soit $\alpha > 0$; on ajoute l'hypothèse supplémentaire : $\exists \beta > 0, \forall n \in \mathbb{N}^*, a_n \leqslant n^{2\alpha-\beta}$.
Alors : $\dfrac{S_n}{n^\alpha} \xrightarrow[n\to\infty]{\text{ps}} 0$.

Démonstration :

Soit $\varepsilon > 0$, on a, par Hoeffding : $\forall n \in \mathbb{N}^*, \mathbb{P}\left(|S_n| > n^\alpha \varepsilon\right) \leqslant 2\exp\left(-\dfrac{n^{2\alpha}\varepsilon^2}{2a_n}\right) \leqslant 2\exp\left(-\dfrac{n^\beta \varepsilon^2}{2}\right)$.

Mais la série $\sum_{n\geqslant 1} \exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right)$ converge (par le critère de Riemann), car $\exists N \in \mathbb{N}^*, \forall n \leqslant N, \dfrac{\varepsilon^2}{2} n^\beta \geqslant 2\ln n$ et donc $\exists N \in \mathbb{N}^*, \forall n \leqslant N, 0 \leqslant \exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right) \geqslant \dfrac{1}{n^2}$.

Ainsi, la série $\sum_{n\geqslant 1} \mathbb{P}\left(|S_n| > n^\alpha\varepsilon\right)$ converge, d'où, par Borel-Cantelli :

$$\forall \varepsilon > 0, \mathbb{P}\left(\limsup_{n\to\infty} \{|S_n| > n^\alpha\varepsilon\}\right) = 0, \text{ ie } \mathbb{P}\left(\liminf_{n\to\infty} \{|S_n| \leqslant n^\alpha\varepsilon\}\right) = 1.$$

En particulier, $\forall p \in \mathbb{N}^*, \mathbb{P}\left(\bigcup_{n\in\mathbb{N}^*} \bigcap_{k\geqslant n} \left\{\left|\dfrac{S_k}{k^\alpha}\right| \leqslant \dfrac{1}{p}\right\}\right) = 1$.

C'est-à-dire : $\forall p \in \mathbb{N}^*, \exists N_p$ négligeable, $\forall \omega \in N_p^c, \exists n \in \mathbb{N}, \forall k \geqslant n, \left|\dfrac{S_k(\omega)}{k^\alpha}\right| \leqslant \dfrac{1}{p}$.

On pose alors $N = \bigcup_{p\in\mathbb{N}^*} N_p$, alors N est négligeable et :

$$\forall \omega \in N^c, \forall p \in \mathbb{N}^*, \exists n \in \mathbb{N}, \forall k \geqslant n, \left|\frac{S_k(\omega)}{k^\alpha}\right| \leqslant \frac{1}{p}.$$

Ou, en d'autres termes : $\dfrac{S_n}{n^\alpha} \xrightarrow[n\to\infty]{\text{ps}} 0$. ■

Références

[Ouv2] J.-Y. OUVRARD – *Probabilités 2*, 3e éd., Cassini, 2009.

1. Allez, une autre application, si vous aimez les statistiques ; elle provient de la partie 3.2 du livre *Statistique mathématique*, de B. CADRE et C. VIAL, paru en 2012 aux éditions Ellipses. Plaçons-nous dans un modèle statistique $\left(\mathcal{H}^n, \{P_\theta\}_{\theta\in\Theta}\right)$ avec $\mathcal{H} \subset \mathbb{R}$ et $\Theta \subset \mathbb{R}^d$. Le paramètre d'intérêt est $g(\theta)$ avec $g : \Theta \to \mathbb{R}$ une fonction continue On suppose que $X_1, \ldots, X_n$ sont indépendantes et identiquement distribuées, de loi P_θ, bornées P_θ-ps et avec $\mathbb{E}_\theta[X_1] = g(\theta)$. Soit c une borne P_θ-presque sûre de $X_1 - g(\theta)$. On veut un intervalle de confiance pour $g(\theta)$.

Par Hoeffding, $P_\theta\left(\left|\overline{X_n} - g(\theta)\right| > \varepsilon\right) = P_\theta\left(\left|\sum_{j=1}^n \left(X_j - g(\theta)\right)\right| > n\varepsilon\right) \leqslant 2\exp\left(-\dfrac{n^2\varepsilon^2}{2nc^2}\right) = 2\exp\left(-\dfrac{n\varepsilon^2}{2c^2}\right)$. Soit $\alpha \in]0,1[$, on choisit ε de sorte que $\alpha = 2\exp\left(-\dfrac{n\varepsilon^2}{2c^2}\right)$, ie : $\varepsilon = c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}$. Dès lors, on obtient l'intervalle de confiance par excès au niveau $1-\alpha$, pour $g(\theta)$: $I_\alpha = \left[\overline{X_n} - c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}, \overline{X_n} + c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}\right]$.