39. Inégalité de Hoeffding

[Ouv09, Ch10, p132]

Énoncé

Proposition. [Inégalité de Hoeffding]
Soient $(X_i)_{1\le i\le n}$ des variables aléatoires réelles indépendantes, centrées et telles que pour tout $i \in [\![1,n]\!]$, $|X_i| \le c_i$ p.s. pour un réel c_i. Alors pour tout $t \ge 0$:

$$\mathbb{P}(|S_n| \ge t) \le 2\,\mathrm{e}^{-t^2/2\sum_{i=1}^n c_i^2} \qquad \text{où } S_n = \sum_{i=1}^n X_i$$

Corollaire. *Supposons maintenant la suite $(X_i)_{i\in\mathbb{N}}$ infinie, avec les mêmes hypothèses. S'il existe $\alpha, \beta > 0$ tels que $\sum_{i=1}^n c_i^2 \le n^{2\alpha-\beta}$, alors $\frac{S_n}{n^\alpha} \underset{n\to+\infty}{\longrightarrow} 0$ p.s..*

Développement

Commençons par l'inégalité de Hoeffding :

1. Montrons d'abord que si X est une variable aléatoire centrée et bornée par 1 presque sûrement, alors $\mathbb{E}[\mathrm{e}^{\lambda X}] \le \mathrm{e}^{\frac{\lambda^2}{2}}$ pour tout $\lambda \in \mathbb{R}$.

   Soit en effet $\lambda \in \mathbb{R}$. Pour tout $x \in [-1,1]$, on a $x = \frac{1-x}{2} \times (-1) + \frac{1+x}{2} \times 1$ et donc par convexité de $\exp(\lambda .)$, on a $\mathrm{e}^{\lambda x} \le \frac{1-x}{2}\,\mathrm{e}^{-\lambda} + \frac{1+x}{2}\,\mathrm{e}^{\lambda}$ puis comme X est centrée :

$$\mathbb{E}[\mathrm{e}^{\lambda X}] \le \frac{1}{2}(\mathrm{e}^{-\lambda} + \mathrm{e}^{\lambda}) = \cosh(\lambda) = \sum_{n\ge 0} \frac{\lambda^{2n}}{(2n)!} \le \sum_{n\ge 0} \frac{\lambda^{2n}}{n!2^n} = \mathrm{e}^{\frac{\lambda^2}{2}}$$

   puisque $(2n)! \ge n!2^n$ pour tout $n \in \mathbb{N}$ (par récurrence).

2. Soit $i \in [\![1,n]\!]$. On a :

$$\forall \lambda \in \mathbb{R}, \mathbb{E}[\mathrm{e}^{\lambda X_i}] = \mathbb{E}[\mathrm{e}^{c_i\lambda \frac{X_i}{c_i}}] \le \mathrm{e}^{\frac{\lambda^2 c_i^2}{2}}$$

   puisque $\frac{X_i}{c_i}$ est centrée et bornée par 1 presque sûrement.

   Autrement dit chaque X_i est c_i-sous-gaussienne.

3. Fixons $t \in \mathbb{R}_+$. On a alors pour tout $\lambda \ge 0$:

$$\begin{aligned}
\mathbb{P}(S_n \ge t) = \mathbb{P}(\mathrm{e}^{\lambda S_n} \ge \mathrm{e}^{\lambda t}) &\le \mathrm{e}^{-\lambda t}\,\mathbb{E}[\mathrm{e}^{\lambda \sum_{i=1}^n X_i}] && \text{par l'inégalité de Markov} \\
&= \mathrm{e}^{-\lambda t} \prod_{i=1}^n \mathbb{E}[\mathrm{e}^{\lambda X_i}] && \text{par indépendance} \\
&\le \mathrm{e}^{-\lambda t} \prod_{i=1}^n \mathrm{e}^{\frac{\lambda^2 c_i^2}{2}} = \mathrm{e}^{-\lambda t + \frac{\sum c_i^2}{2}\lambda^2} && \text{par le deuxième point}
\end{aligned}$$

   En optimisant en $\lambda = t/\sum_{i=1}^n c_i^2$ on obtient $\mathbb{P}(S_n \ge t) \le \mathrm{e}^{-t^2/2\sum_{i=1}^n c_i^2}$.

   Puisque les $(-X_i)_{1\le i\le n}$ satisfont les mêmes hypothèses que les $(X_i)_{1\le i\le n}$, on a de même $\mathbb{P}(S_n \le -t) \le \mathrm{e}^{-t^2/2\sum_{i=1}^n c_i^2}$.

   D'où finalement $\mathbb{P}(|S_n| \ge t) \le 2\,\mathrm{e}^{-t^2/2\sum_{i=1}^n c_i^2}$.

Passons au corollaire. Soit $\varepsilon > 0$. On a par l'inégalité de Hoeffding :

$$\sum_{n\in\mathbb{N}} \mathbb{P}(|S_n| > n^\alpha \varepsilon) \le 2 \sum_{n\in\mathbb{N}} \mathrm{e}^{-\frac{n^{2\alpha}\varepsilon^2}{2n^{2\alpha-\beta}}} = 2 \sum_{n\in\mathbb{N}} \mathrm{e}^{-\frac{\varepsilon^2}{2}n^\beta} < +\infty$$

la convergence de la dernière série ayant lieu puisque $\mathrm{e}^{-\frac{\varepsilon^2}{2}n^\beta} = o(n^{-2})$.

Par le lemme de Borel-Cantelli, on a donc que $\mathbb{P}(\limsup_{n\to+\infty} |S_n| > n^\alpha\varepsilon) = 0$, puis par dénombrabilité :

$$\mathbb{P}\Big(\bigcup_{k\in\mathbb{N}^*} \limsup_{n\to+\infty} |S_n| > \frac{n^\alpha}{k}\Big) = 0 \qquad \text{donc} \qquad \mathbb{P}\Big(\bigcap_{k\in\mathbb{N}^*} \liminf_{n\to+\infty} |S_n| \le \frac{n^\alpha}{k}\Big) = 1$$

Ou encore

$$\mathbb{P}\Big(\lim_{n\to+\infty} \frac{|S_n|}{n^\alpha} = 0\Big) = 1$$

Ainsi $\frac{S_n}{n^\alpha} \underset{n\to+\infty}{\longrightarrow} 0$ p.s..

Commentaires

On rappelle qu'en probabilités, $\limsup_{n\in\mathbb{N}} A_n$ désigne l'événement « A_n est réalisé pour une infinité de n », alors que $\liminf_{n\in\mathbb{N}} A_n$ désigne l'événement « A_n est réalisé pour tout n à partir d'un certain rang ».