Inégalité de Hoeffding

Références : Ouvrard, *Probabilités 2*, 10.11

Théorème.

Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes et centrées. On suppose de plus $|X_n| \leqslant c_n$ presque partout, où $c_n > 0$. Alors, en notant $S_n = \sum_{j=1}^{n} X_j$, on a $\forall \varepsilon > 0$,

$$\mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2\sum_{j=1}^n c_j^2}\right).$$

Lemme.

Soit X une variable aléatoire réelle centrée et bornée par 1 ps, alors $\forall t \in \mathbb{R}$, $\mathbb{E}[e^{tX}] \leqslant e^{\frac{t^2}{2}}$.

Démonstration. Soit $x \in [-1, 1]$, alors par convexité de l'exponentielle, on a

$$e^{tx} = \exp\left(\frac{1-x}{2} \times (-t) + \frac{1+x}{2} \times t\right) \leqslant \frac{1-x}{2}e^{-t} + \frac{1+x}{2}e^{t}.$$

On rappelle que l'exponentielle est convexe car de dérivée seconde positive, et on peut utiliser l'inégalité car $\frac{1-x}{2} + \frac{1+x}{2} = 1$ et $\frac{1-x}{2}$, $\frac{1+x}{2}$ sont dans $[0, 1]$.

X étant bornée presque surement par 1, on peut utiliser l'inégalité précédente : $e^{tX} \leqslant \frac{1-X}{2}e^{-t} + \frac{1+X}{2}e^{t}$. En intégrant, on obtient $\mathbb{E}[e^{tX}] \leqslant \frac{1}{2}(e^{-t} + e^{t}) = \cosh(t)$ car X est centrée.

Puis $\cosh(t) = \sum \frac{t^{2n}}{(2n)!} \leqslant \sum \frac{t^{2n}}{2^n n!} = e^{\frac{t^2}{2}}$. En effet, $\frac{(2n)!}{n!} = (2n)(2n-1)...(n+1) \geqslant 2 \times 2 \times ... \times 2 = 2^n$.
D'où il vient $\mathbb{E}[e^{tX}] \leqslant e^{\frac{t^2}{2}}$. □

On va maintenant prouver l'inégalité de Hoeffding.

Démonstration. Commençons par étudier $\mathbb{P}(S_n > \varepsilon)$.
On a $\forall t > 0$, $\mathbb{P}(S_n > \varepsilon) = \mathbb{P}(e^{tS_n} > e^{t\varepsilon})$.
L'inégalité de Markov donne alors $\mathbb{P}(e^{tS_n} > e^{t\varepsilon}) \leqslant \frac{\mathbb{E}[e^{tS_n}]}{e^{t\varepsilon}}$ car e^{tS_n} est positive.
De plus, $\mathbb{E}[e^{tS_n}] = \prod_{j=1}^{n} \mathbb{E}[e^{tX_j}]$ par indépendance et $\mathbb{E}[e^{tX_j}] = \mathbb{E}[e^{(c_j t)\frac{X_j}{c_j}}] \leqslant e^{\frac{c_j^2 t^2}{2}}$ par le lemme.
On en déduit $\mathbb{E}[e^{tS_n}] \leqslant \prod_{j=1}^{n} e^{\frac{c_j^2 t^2}{2}} = \exp\left(\frac{t^2 \sum_{j=1}^n c_j^2}{2}\right)$.
Finalement $\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(\frac{t^2 \sum_{j=1}^n c_j^2}{2} - t\varepsilon\right)$.

On définit $a := \sum_{j=1}^{n} c_j^2$, alors on va tenter de minimiser sur $\mathbb{R}^{+*}$ la fonction de t, $\frac{at^2}{2} - t\varepsilon$. Elle atteint son minimum en $t_0 = \frac{\varepsilon}{a}$ et elle vaut alors $-\frac{\varepsilon^2}{2a}$. On a donc $\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(-\frac{\varepsilon^2}{2a}\right)$.

Pour finir, si on refait le même raisonnement avec la suite $(-X_i)_i$, on trouve $\mathbb{P}(S_n < -\varepsilon) = \mathbb{P}(-S_n > \varepsilon) \leqslant$

$\exp\left(-\frac{\varepsilon^2}{2a}\right)$.

D'où $\mathbb{P}(|S_n| > \varepsilon) = \mathbb{P}(S_n > \varepsilon) + \mathbb{P}(S_n < -\varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2a}\right) = 2\exp\left(-\frac{\varepsilon^2}{2\sum_{j=1}^n c_j^2}\right)$. □

Corollaire.

Si il existe α, $\beta > 0$ tels que $\sum_{j=1}^n c_j^2 \leqslant n^{2\alpha-\beta}$, alors $\frac{S_n}{n^\alpha}$ converge presque surement vers 0.

Démonstration. Soit $\varepsilon > 0$, on utilise l'inégalité de Hoeffding : $\mathbb{P}(|S_n| > n^\alpha\varepsilon) \leqslant 2\exp\left(-\frac{n^{2\alpha}\varepsilon^2}{2\sum_{j=1}^n c_j^2}\right)$.

On a alors sous nos hypothèses : $\mathbb{P}\left(\frac{|S_n|}{n^\alpha} > \varepsilon\right) \leqslant 2\exp\left(-\frac{n^\beta\varepsilon^2}{2}\right)$.

La série de terme général $\exp\left(-\frac{n^\beta\varepsilon^2}{2}\right)$ converge car à termes positifs et $\exp\left(-\frac{n^\beta\varepsilon^2}{2}\right) = o\left(\frac{1}{n^2}\right)$.

On en déduit par Borel-Cantelli que $\mathbb{P}\left(\liminf_n\left\{\frac{|S_n|}{n^\alpha} < \varepsilon\right\}\right) = 1$, ce qui est une manière de dire que $\frac{|S_n|}{n^\alpha} \to 0$ presque surement.[1] □

Remarques : • Si on compare les inégalités de Bienaymé-Tchebychev et Hoeffding sur des lois de Bernoulli ($c_n = 1$), on a

$\mathbb{P}(|Sn - np| > \sqrt{n}\varepsilon) \leqslant \frac{p(1-p)}{\varepsilon^2}$ pour Bienaymé-Tchebychev,

et $\mathbb{P}(|Sn - np| > \sqrt{n}\varepsilon) \leqslant 2\exp\left(-\frac{\epsilon^2}{2}\right)$ pour Hoeffding.

C'est quand même carrément mieux avec Hoeffding !

• Cette inégalité permet d'obtenir des intervalles de confiance... mais on préfère souvent utiliser le TCL.

• Une généralisation de cette inégalité est l'inégalité d'Azuma :

Théorème.

Soit $(X_n)_n$ une martingale issue de 0 dont les accroissements sont contrôlés par une suite déterministe (c_n), c'est à dire $|X_n - X_{n-1}| \leqslant c_n$ presque surement pour tout $n \geqslant 1$. Alors pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X_n| \geqslant \varepsilon) \leqslant 2e^{-\frac{\lambda^2}{2\sigma_n^2}}, \text{ où } \sigma_n^2 = \sum_{i=1}^n c_i^2 .$$

On peut appliquer ce résultat au calcul du nombre de couleurs nécessaires pour colorier un graphe à n sommets avec la condition que deux sommets reliés par une arête ne peuvent être de la même couleur.

Il y a au maximum $\binom{n}{k}$ arêtes possibles. On note Y_k la variable aléatoire de loi de Bernoulli $b(p)$ décidant de si on rajoute l'arête k ou non. On note enfin χ le nombre minimal de couleurs nécessaires selon les arêtes présentes.

On montre alors $\mathbb{P}(|\chi - \mathbb{E}[\chi]| \geqslant \varepsilon\sqrt{n-1}) \leqslant 2e^{-\frac{\lambda^2}{2}}$.

1.

$$\begin{aligned} X_n \to X ps &\Leftrightarrow \mathbb{P}\left(\bigcap_{\varepsilon>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m\geqslant n}\{|X_m - X| < \varepsilon\}\right) = 1 \\ &\Leftrightarrow \mathbb{P}\left(\bigcap_{k>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m\geqslant n}\left\{|X_m - X| < \frac{1}{k}\right\}\right) = 1 \\ &\Leftrightarrow \lim_{k\to\infty}\mathbb{P}\left(\liminf_n\left\{|X_n - X| < \frac{1}{k}\right\}\right) = 1 \end{aligned}$$