5 Inégalité de Høffding

Leçons 261, 262, 266(, 253)

Ref : [Bernis & Bernis] ou [Ouvrard]

Théorème 1 (Inégalité de Høffding) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes centrées telle que pour tout $n \in \mathbb{N}$, il existe une constante $c_n > 0$ telle que

$$|X_n| < c_n \text{ p.s..}$$

Alors on a pour tout $\varepsilon > 0$

$$\mathbb{P}(|S_n| > \varepsilon) \le 2\,\mathrm{e}^{-\frac{\varepsilon^2}{2a}},$$

où S_n désigne la somme $\displaystyle\sum_{i=1}^n X_i$, et $a = \displaystyle\sum_{i=1}^n c_i^2$.

Démonstration. On commence par démontrer le lemme suivant.

Lemme 2 Soit X une variable aléatoire réelle centrée telle que $|X| < 1$ presque sûrement. Alors on a

$$\forall t \in \mathbb{R} \qquad \mathbb{E}\left(e^{tX}\right) \le \mathrm{e}^{\frac{t^2}{2}}.$$

Démonstration. On se donne un réel $x \in [-1, 1]$. Alors si $\lambda := \frac{1}{2}(1 + x)$ est dans $[0, 1]$, et donc $1 - \lambda = \frac{1}{2}(1 - x)$ aussi. De plus, on a $x = 2\lambda - 1 = \lambda\times 1 + (1 - \lambda)\times(-1)$. Donc, par convexité de l'exponentielle, on a

$$\forall t \in \mathbb{R} \qquad \mathrm{e}^{tx} = \mathrm{e}^{t(\lambda\times 1+(1-\lambda)\times(-1))} \le \lambda\,\mathrm{e}^{t} + (1 - \lambda)\,\mathrm{e}^{-t}.$$

De plus, comme X est bornée par 1, e^{tX} est intégrable pour tout $t \in \mathbb{R}$, et on a d'après ce qui précède

$$\mathrm{e}^{tX} \le \frac{1}{2}(1 + X)\,\mathrm{e}^{t} + \frac{1}{2}(1 - X)\,\mathrm{e}^{-t}.$$

Ainsi, en passant à l'espérance et en rappelant que X est centrée, on obtient

$$\mathbb{E}\left(\mathrm{e}^{tX}\right) = \mathrm{ch}(t).$$

On compare alors les développements en série entière de l'exponentielle et du cosinus hyperbolique : comme on a pour tout n l'inégalité

$$\frac{1}{(2n)!} \le \frac{1}{2^n n!},$$

on en déduit

$$\forall t \in \mathbb{R} \qquad \mathrm{ch}(t) \le \mathrm{e}^{\frac{t^2}{2}},$$

ce qui permet de déduire le résultat souhaité. $\square$

On se donne maintenant $t \in \mathbb{R}$ et $n \in \mathbb{N}$. D'après le lemme, on a

$$\mathbb{E}\left(\mathrm{e}^{t\frac{X_n}{c_n}}\right) \le \mathrm{e}^{\frac{t^2}{2}}.$$

En prenant $s = \dfrac{t}{c_n} \in \mathbb{R}$, on en déduit

$$\mathbb{E}\left(\mathrm{e}^{sX_n}\right) \le \mathrm{e}^{\frac{s^2 c_n^2}{2}},$$

et ceci est donc vrai pour tout $s \in \mathbb{R}$ et $n \in \mathbb{N}$. On en déduit par indépendance des X_i :

$$\forall t \in \mathbb{R} \qquad \mathbb{E}\left(\mathrm{e}^{tS_n}\right) = \prod_{i=1}^n \mathbb{E}\left(\mathrm{e}^{tX_i}\right) \le \mathrm{e}^{\frac{at^2}{2}},$$

où a désigne la somme $\displaystyle\sum_{i=1}^n c_i^2$.

On se donne maintenant $\varepsilon > 0$. On a alors pour $t > 0$

$$\begin{array}{rcll} \mathbb{P}(S_n > \varepsilon) & = & \mathbb{P}\left(\mathrm{e}^{tS_n} > \mathrm{e}^{t\varepsilon}\right) & \text{par criossance de l'exponentielle et positivité de } t \\ & \leq & \dfrac{\mathbb{E}\left(\mathrm{e}^{tS_n}\right)}{\mathrm{e}^{t\varepsilon}} & \text{par inégalité de Markov} \\ & \leq & \mathrm{e}^{\frac{at^2}{2} - \varepsilon t} & \end{array}$$

Or le minimum de la fonction

$$f : \left| \begin{array}{ccc} \mathbb{R}_+^* & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \dfrac{at^2}{2} - \varepsilon t \end{array} \right.$$

est atteint en $\dfrac{\varepsilon}{a}$ et vaut $-\dfrac{\varepsilon^2}{2a}$. Donc

$$\mathbb{P}(S_n > \varepsilon) \leq \mathrm{e}^{-\frac{\varepsilon^2}{2a}}.$$

Puisque les variables $-X_n$ vérifient les mêmes propriétés que les variables X_n, le résultat est aussi vrai en prenant $-S_n$ à la place de S_n. On en déduit l'inégalité de Høffding. □

Corollaire 3 On se donne des observations $X_1, .., X_n$ indépendantes identiquement réparties d'une loi inconnue de moyenne m, pour un modèle paramétrique $(\mathcal{H}^n, (\mathbb{Q}_\theta)_{\theta \in \Theta})$. On désigne alors par $\overline{X}_n$ la moyenne empirique associée aux observations $X_1, .., X_n$. On suppose qu'il existe une constante $c > 0$ telle que

$$\forall \theta \in \Theta \ \forall i \in [\![1, n]\!] \qquad |X_i - m| \leq c \ \ \mathbb{Q}_\theta - \text{p.s.}.$$

Alors, pour $\alpha \in (0, 1)$ fixé, l'intervalle

$$ICE_{1-\alpha} := \left[\overline{X}_n - c\sqrt{\frac{2}{n} \ln\left(\frac{2}{\alpha}\right)}, \overline{X}_n + c\sqrt{\frac{2}{n} \ln\left(\frac{2}{\alpha}\right)}\right]$$

est un intervalle de confiance par excès de niveau $1 - \alpha$ pour la moyenne $m = m(\theta)$.

Démonstration. Il suffit d'appliquer l'inégalité de Høffding aux X_i : on a ici $a = nc^2$, d'où

$$\forall q > 0 \quad \mathbb{P}\left(|\overline{X}_n| > q\right) = \mathbb{P}\left(|S_n| > nq\right) \leq 2\,\mathrm{e}^{-\frac{nq^2}{2c^2}}.$$

Prendre la bonne valeur de q pour avoir $1 - \alpha = 2\,\mathrm{e}^{-\frac{nq^2}{2c^2}}$ fournit alors l'intervalle de confiance $ICE_{1-\alpha}$. □