Inégalité de Hoeffding

Référence : Ouvrard 2 : p. 128 + Cadre-Vial p.36 pour l'application
Lecons : : 229, 253, 260,261,262.

Notation : On note $L_X(t) = \mathbb{E}[e^{tX}]$ la transformée de Laplace de la variable aléatoire X.

Lemme

Soit X une variable aléatoire réelle $\mathbb{P}$-ps bornée par 1 ie $|X| \leqslant 1$ $\mathbb{P}$-ps. On suppose que X est centrée. On a :

$$L_X(t) \leqslant \exp\left(\frac{t^2}{2}\right)$$

Preuve :

Etape 1 : Mq pour $t \in \mathbb{R}$, $x \in [-1, 1]$, $\exp(tx) \leqslant \frac{1}{2}(1-x)\exp(-t) + \frac{1}{2}(1+x)\exp(t)$

C'est juste de la convexité car :

$$0 \leqslant \lambda_1 = \frac{1}{2}(1-x) \leqslant 1$$

$$0 \leqslant \lambda_2 = \frac{1}{2}(1+x) \leqslant 1$$

$$\lambda_1 + \lambda_2 = 1$$

Et enfin

$$tx = \lambda_1(-t) + \lambda_2 t$$

Comme $x \mapsto \exp(tx)$ est convexe (avec sa dérivée seconde), on a l'inégalité souhaitée.

Etape 2 : Mq $L_X(t) \leqslant \frac{1}{2}(\exp(-t) + \exp(t))$.

Déjà, comme X est ps bornée par 1, $\exp(tx)$ est bornée ps et admet donc une moyenne. Ainsi $L_X(t)$ existe. Avec l'étape 1 on a, pour $t \in \mathbb{R}$, et ps,

$$\exp(tX) \leqslant \frac{1}{2}(1-X)\exp(-t) + \frac{1}{2}(1+X)\exp(t)$$

et donc

$$L_X(t) \leqslant \frac{1}{2}\mathbb{E}[1-X]\exp(-t) + \frac{1}{2}\mathbb{E}[1+X]\exp(t)$$

Comme X est centrée :

$$L_X(t) \leqslant \frac{1}{2}\left(\exp(-t) + \exp(t)\right) = \operatorname{ch}(t)$$

Or $\operatorname{ch}(t) = \sum_{n \in \mathbb{N}} \frac{t^{2n}}{(2n)!}$ et $\exp\left(\frac{t^2}{2}\right) = \sum_{n \in \mathbb{N}} \frac{t^{2n}}{n! 2^n}$.

mais pour tout $n \in \mathbb{N}$, on a $n!2^n = 2 \cdot 4 \cdots 2n \leqslant (2n)!$ donc $\operatorname{ch}(t) \leqslant \exp\left(\frac{t^2}{2}\right)$.

Et finalement

$$L_X(t) \leqslant \exp\left(\frac{t^2}{2}\right)$$

■

PROPOSITION (INÉGALITÉ DE HOEFFDING - 1963)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes, 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 : $\forall n \in \mathbb{N}^*, S_n = \displaystyle\sum_{j=1}^{n} X_j$ et $a_n = \displaystyle\sum_{j=1}^{n} c_j^2$.
Alors $\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}\left(|S_n| > \varepsilon\right) \leqslant 2\exp\left(-\dfrac{\varepsilon^2}{2a_n}\right)$.

__Preuve :__

Etape 1 : Mq pour tout t, $L_{S_n}(t) \leqslant \exp\left(\dfrac{t^2}{2}a_n\right)$

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

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

Etape 2 : Mq pour tout $\varepsilon > 0$, $\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(-\dfrac{\varepsilon^2}{2a_n}\right)$

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 [1], on obtient :

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

Cette inégalité est vraie pour tout $t > 0$. On va donc minimiser en t.
$t \mapsto -t\varepsilon + \dfrac{t^2}{2}a_n$ atteint son minimum en $t = \dfrac{\varepsilon}{a_n}$ et ce minimum vaut $-\dfrac{\varepsilon^2}{2a_n}$, d'où :

$$\mathbb{P}\left(S_n > \varepsilon\right) \leqslant \exp\left(\frac{\varepsilon^2}{a_n^2}\frac{a_n}{2} - \frac{\varepsilon}{a_n}\varepsilon\right) = \exp\left(-\frac{\varepsilon^2}{2a_n}\right)$$

Etape 3 : Conclusion

Si $\varepsilon > 0$. On a les égalités :

$$\{|S_n| > \varepsilon\} = \{S_n > \varepsilon\} \sqcup \{S_n < -\varepsilon\}$$

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

$\blacksquare$

1. Soit Z une variable aléatoire réelle positive ou nulle ps. Alors $\forall a > 0, \mathbb{P}(Z \geqslant a) \leqslant \dfrac{\mathbb{E}[Z]}{a}$

Application 1 : un critère de convergence presque sûre

Proposition

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

Preuve :

Etape 1 : Mq pour tout $\varepsilon > 0$, la série de terme général $\mathbb{P}(|S_n| > n^\alpha \varepsilon)$ est convergente.

Soit $\varepsilon > 0$, on a [2], 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 $\displaystyle\sum_{n\geqslant 1} \exp\left(-\frac{n^\beta\varepsilon^2}{2}\right)$ converge (par le critère de Riemann), car à partir d'un certain rang, $\dfrac{\varepsilon^2}{2}n^\beta \geqslant 2\ln n$ et donc $0 \leqslant \exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right) \geqslant \dfrac{1}{n^2}$.

Ainsi, la série $\displaystyle\sum_{n\geqslant 1} \mathbb{P}\left(|S_n| > n^\alpha\varepsilon\right)$ converge

Etape 2 : [3] Par Borel-Cantelli [4] :

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

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

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

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

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

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

■

2. manque le /2 dans Ouvrard
3. On fait pas comme Ouvrard parce que c'est pas rigoureux.
4. Lemme de Borel-Cantelli : Soit $(A_n)_{n\geqslant 0}$ une suite dans $\mathcal{A}$. Si $\displaystyle\sum_{n\geqslant 0} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_n A_n) = 0$

Application 2 : calcul d'intervalles de confiance par excès

Rappelons les conditions de l'inégalité :

> Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes, 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 : $\forall n \in \mathbb{N}^*, S_n = \displaystyle\sum_{j=1}^{n} X_j$ et $a_n = \displaystyle\sum_{j=1}^{n} c_j^2$.
> Alors $\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}\left(|S_n| > \varepsilon\right) \leqslant 2\exp\left(-\dfrac{\varepsilon^2}{2a_n}\right)$.

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 [5], bornées P_θ-ps et avec $\mathbb{E}_\theta\left[X_1\right] = g(\theta)$ [6]. Soit c une borne P_θ-presque sûre de $X_1 - g(\theta)$ [7].
On veut un intervalle de confiance pour $g(\theta)$ donc on cherche par exemple $\mathbb{P}\left(\left|\overline{X_n} - g(\theta)\right| > \varepsilon\right)$
Par Hoeffding,

$$P_\theta\left(\left|\overline{X_n} - g(\theta)\right| > \varepsilon\right) = P_\theta\left(\left|\sum_{j=1}^{n}\left(X_j - g(\theta)\right)\right| > n\varepsilon\right) \leqslant 2\exp\left(-\frac{n^2\varepsilon^2}{2nc^2}\right) = 2\exp\left(-\frac{n\varepsilon^2}{2c^2}\right)$$

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 [8] au niveau $1-\alpha$, pour $g(\theta)$:

$$I_\alpha = \left[\overline{X_n} - c\sqrt{\frac{2}{n}\ln\frac{2}{\alpha}}, \overline{X_n} + c\sqrt{\frac{2}{n}\ln\frac{2}{\alpha}}\right]$$

Notes :

✓ **A l'oral**, selon la leçon, on fera l'une des deux applications (de préférence la deuxième). En allant un peu vite lemme 4'46, lemme + inega = 9', appli 1= 4'41, appli 2=3'37 ou 4'27 si on détaille pour des bernoulli. L'idée c'est si appli 1 aller vite sur Hoeffding quitte à prendre le temps pour l'appli qui manipule des choses délicates. Si appli 2, prendre plus de temps sur hoeffding et voir à la fin comment j'ai le temps de détailler.
♣ Wassily Hoeffding (1914 - 1991) est un statisticien et probabiliste finnois. Il est l'un des fondateurs des statistiques non-paramétriques. Il est surtout connu pour cette inégalité.

5. Par exemple, $\mathcal{B}(\theta)^{\otimes n}$
6. Toujours dans notre exemple, $g(\theta) = \theta$.
7. Ex : on va avoir $1 \leqslant -\theta \leqslant X_1 - \theta \leqslant 1 - \theta \leqslant 1$ donc bornées par 1.
8. Un intervalle de confiance par excès pour un paramètre θ de niveau de confiance $1-\alpha$ est une statistique I telle que $\mathbb{P}_\theta(\theta \in I) \geqslant \alpha$.