

Inégalité de Hoeffding: Bernis p. 217 ~~260-263~~ - 261 - 262 - 266

Soit $(X_n)_n$ une suite de V.A. $\mathbb{R}$ indépendantes et centrées. On suppose qu'il existe $(c_n)_n \in (\mathbb{R}^+)^{\mathbb{N}}$ / $\forall n \in \mathbb{N}^+, |X_n| \leq c_n$ p.s. $\hookrightarrow \mathbb{E}(X_n) = 0$

Alors, $\forall \varepsilon > 0, \forall n \in \mathbb{N}^+$, on a: $\mathbb{P}(|S_n| > \varepsilon) \leq 2 \exp \left(\frac{-\varepsilon^2}{2 \sum_{k=1}^n c_k^2} \right)$ où $S_n = X_1 + \dots + X_n$

Démonstration: 1) cas d'une seule V.A. bornée.

•) Soit $x \in [-1, 1]$, alors
$$\begin{cases} \frac{1-x}{2} + \frac{1+x}{2} = 1 \\ \frac{1-x}{2}, \frac{1+x}{2} \in [0, 1] \\ \frac{1-x}{2} \times (-t) + \frac{1+x}{2} \times t = tx \end{cases}$$

Par convexité de $\exp$, on a alors $\forall t \in \mathbb{R}, e^{tx} \leq \frac{1-x}{2} e^{-t} + \frac{1+x}{2} e^t$

•) Soit X une V.A. $\mathbb{R}$, centrée et bornée par 1. $\hookrightarrow |tx| \leq |t|$ $\hookrightarrow |\exp(tx)| \leq \exp |tx| \leq \exp |t|$
 $\forall t \in \mathbb{R}, tX$ est bornée p.s. et donc $\exp(tX)$ aussi.

Donc $t \mapsto \exp(tX)$ est intégrable et donc la transformée de Laplace de X existe sur $\mathbb{R}$ tout entier.

$$\begin{aligned} \text{Soit } t \in \mathbb{R}, L_X(t) = \mathbb{E}(e^{tX}) &\leq \frac{1}{2} \mathbb{E}(1-x) e^{-t} + \frac{1}{2} \mathbb{E}(1+x) e^t \\ &\leq \frac{1}{2} e^{-t} + \frac{1}{2} e^t = \cosh(t) \end{aligned}$$

$$\text{Or } \cosh(t) = \sum_{h=0}^{\infty} \frac{t^{2h}}{(2h)!} \quad \text{et} \quad \forall h \in \mathbb{N}^+, 2^h h! = (2 \times 1) \times (2 \times 2) \times \dots \times (2 \times h) \leq (2h)! \\ \text{ex: } 2 \cdot 3 \cdot 4 \cdot 5 \cdot \dots \cdot (2h) \cdot 2h$$

$$\text{Donc } \cosh(t) \leq \sum_{h=0}^{\infty} \frac{t^{2h}}{2^h h!} = e^{t^2/2}$$

$$\text{Donc } L_X(t) \leq e^{t^2/2} \quad \forall t \in \mathbb{R}$$

2) cas général: fin de la preuve.

•) Soient $n \in \mathbb{N}^+, t \in \mathbb{R}$

Par continuité donc monotonie de $\exp$, $(\exp(tX_k))_k$ est une famille de V.A. $\mathbb{R}$ indépendantes.

De plus, $\forall h \in \mathbb{N}$, $\frac{X_h}{ch}$ est centré et $\frac{X_h}{ch} \leq 1$ p.s.

Donc par 1) et par indépendance de $(\exp(tX_h))_h$; on a:

$$L_{S_n}(t) = \mathbb{E} \left(\exp \left(t \sum_{h=1}^n X_h \right) \right) \stackrel{=}{=} \mathbb{E} \left(\prod_{h=1}^n \exp(tX_h) \right) = \prod_{h=1}^n \mathbb{E}(\exp(tX_h)) = \prod_{h=1}^n \left(\mathbb{E} \left(\exp \left(ct \frac{X_h}{ch} \right) \right) \right)$$

$$\stackrel{\mathbb{E}(e^{tX}) \leq e^{t^2/2}}{\leq} \prod_{h=1}^n e^{c^2 t^2 / 2} = \exp \left(\frac{t^2}{2} \sum_{h=1}^n c^2 \right)$$

•) Soient $\varepsilon > 0$, $n \in \mathbb{N}^*$.

$$\forall t > 0, \mathbb{P}(S_n > \varepsilon) = \mathbb{P}(\exp(tS_n) > \exp(t\varepsilon)) \text{ par stricte croissance de } \exp$$

$$\leq \frac{1}{\exp(t\varepsilon)} \underbrace{\mathbb{E}(\exp(tS_n))}_{L_{S_n}(t)} \text{ par inégalité de Markov car } \exp(tS_n) \text{ est une VA pos}$$

$$\leq \exp \left(\frac{t^2}{2} \sum_{h=1}^n c^2 - t\varepsilon \right)$$

•) Cette inégalité est optimale pour $t_{\min} = -\frac{b}{2a} = \frac{\varepsilon}{\sum_{h=1}^n c^2}$

$$\text{donc } \forall t > 0, \mathbb{P}(S_n > \varepsilon) \leq \exp \left(\frac{1}{2} \frac{\varepsilon^2}{\left(\sum_{h=1}^n c^2 \right)^2} \sum_{h=1}^n c^2 - \frac{\varepsilon}{\sum_{h=1}^n c^2} \times \varepsilon \right)$$

$$\leq \exp \left(- \frac{\varepsilon^2}{2 \sum_{h=1}^n c^2} \right)$$

•) $(-X_n)_n$ vérifie les mêmes hypothèses que $(X_n)_n$ donc $\mathbb{P}(-S_n > \varepsilon) \leq \exp \left(\frac{-\varepsilon^2}{2 \sum_{h=1}^n c^2} \right)$

$$\text{donc } \mathbb{P}(|S_n| > \varepsilon) = \mathbb{P}(\varepsilon < S_n \text{ ou } \varepsilon < -S_n)$$

$$= \mathbb{P}(S_n > \varepsilon) + \mathbb{P}(-S_n > \varepsilon) \text{ par incompatibilité des événements}$$

$$\leq 2 \exp \left(\frac{-\varepsilon^2}{2 \sum_{h=1}^n c^2} \right) \quad \square$$

Application 2: Soit $n \in \mathbb{N}^*$. On étudie le modèle Mark $(\mathcal{I}_0, \mathcal{I}_n, (\mathcal{B}(U, p))_{p \in \mathcal{D}(U)})$

Soit $X_1, \dots, X_n$ un n -échantillon de loi $\mathcal{B}(U, p)$, $p \in \mathcal{D}(U)$

Pour $\alpha \in \mathcal{D}(0, 1)$, déterminons un intervalle de confiance par exi au niveau $1-\alpha$ du paramètre p .

Soit $\alpha \in]0,1[$.

On applique l'inégalité de Hoeffding à $(X_k - p)_k$, borne par 1 :

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} S_n - p\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n (X_k - p)\right| > \varepsilon\right) = \mathbb{P}\left(\left|\sum_{k=1}^n (X_k - p)\right| > n\varepsilon\right) \\ &\leq 2 \exp\left(-\frac{n^2 \varepsilon^2}{2 \sum_{k=1}^n 1}\right) = 2 \exp\left(-\frac{n \varepsilon^2}{2}\right) \end{aligned}$$

Posons $\varepsilon_2 = \sqrt{\frac{2}{n} \ln\left(\frac{2}{\alpha}\right)}$ de sorte que $\alpha = 2 \exp\left(-\frac{n \varepsilon_2^2}{2}\right)$

$$\text{Alors } \mathbb{P}\left(\left|\frac{1}{n} S_n - p\right| \leq \varepsilon_2\right) = 1 - \mathbb{P}\left(\left|\frac{1}{n} S_n - p\right| > \varepsilon_2\right) \geq 1 - \alpha$$

Donc $\left[\frac{1}{n} S_n - \varepsilon_2, \frac{1}{n} S_n + \varepsilon_2\right]$ est un intervalle de confiance par excès de p au niveau de confiance $1 - \alpha$.

↳ Soit $(x_2, \dots, x_n) \in \{0,1\}^n$ une réalisation de $(X_2, \dots, X_n)$. Notons $S_n = \sum_{i=1}^n x_i$
Si $\alpha = 0,05$, alors avec une proba sup à 95%, le param p appartient à l'intervalle $\left[\frac{1}{n} S_n - \varepsilon_2, \frac{1}{n} S_n + \varepsilon_2\right]$

↳ Comparaison avec Tchebychev : pour $\alpha \in]0,1[$ fixé, et un échantillon fixé, le meilleur intervalle sera le + étroit.

$$\varepsilon_2 = \frac{1}{2\sqrt{n\alpha}} \quad ; \quad \left[\frac{1}{n} S_n - \varepsilon_2, \frac{1}{n} S_n + \varepsilon_2\right] \dots$$

Une étude nous donne $\psi: \alpha \mapsto \varepsilon_2^2(\alpha) = \varepsilon_2^2(n)$ msg $\psi(\alpha) \leq 0$ sur $]0, \alpha_0[$ avec $\alpha_0 \approx 0,0297$. Donc Hoeff > Tcheb pour des petites vals de α

Application 2: CV ps. Soit $\alpha > 0$. On suppose qu'il existe $\beta > 0 \mid \forall n \in \mathbb{N}^+$,
 $\sum_{k=1}^n d_k^2 \leq n^{2\alpha - \beta}$. Alors $\frac{S_n}{n^\alpha} \rightarrow 0$ ps

$$\text{Soient } \varepsilon > 0, n \in \mathbb{N}^+, \mathbb{P}(|S_n| > n^\alpha \varepsilon) \leq 2 \exp\left(-\frac{n^{2\alpha} \varepsilon^2}{2 \sum_{k=1}^n d_k^2}\right) \leq 2 \exp\left(-\frac{n^\beta \varepsilon^2}{2}\right)$$

$$\text{Or } \exp\left(-\frac{n^2 \varepsilon^2}{2}\right) = o\left(\frac{1}{n^2}\right) \text{ donc par comparaison: } \sum_{n \geq 2} P(|S_n| > n^\alpha \varepsilon) < \infty$$

Par le lemme de Borel-Cantelli: $\forall \varepsilon > 0$, $P\left(\limsup_{n \rightarrow \infty} \left\{ \left| \frac{S_n}{n^\alpha} \right| > \varepsilon \right\}\right) = 0$

$$\text{Donc } \forall \varepsilon \in \mathbb{Q}_+^*, 1 = P\left(\limsup_{n \rightarrow \infty} \left\{ \left| \frac{S_n}{n^\alpha} \right| \leq \varepsilon \right\}\right) = P\left(\bigcup_{n \in \mathbb{N}^*} \bigcap_{m \geq n} \left\{ \left| \frac{S_m}{m^\alpha} \right| \leq \varepsilon \right\}\right)$$

Donc $\forall \varepsilon \in \mathbb{Q}_+^*$, il existe $F_\varepsilon \subset \mathbb{R}$ de mesure pleine / $\forall w \in F_\varepsilon$, $\exists n \in \mathbb{N}^* / \forall m \geq n$, $\left| \frac{S_m(w)}{m^\alpha} \right| \leq \varepsilon$

Donc pour $F = \bigcap_{\varepsilon \in \mathbb{Q}_+^*} F_\varepsilon$ de mesure pleine, $\forall w \in F$, $\forall \varepsilon \in \mathbb{Q}_+^*$, $\exists n \in \mathbb{N}^* / \forall m \geq n$, $\left| \frac{S_m(w)}{m^\alpha} \right| \leq \varepsilon$

; Ici, $\forall \varepsilon \in \mathbb{Q}_+^*$ donc $\frac{S_n}{n^\alpha} \rightarrow 0$ ps.
par densité 2 p. 50

Lemme de Borel-Cantelli: Soit (f_n) une suite d'événements tq $\sum_{n=0}^{+\infty} P(f_n) < \infty$.

$$\text{Alors } P\left(\limsup_{n \rightarrow +\infty} f_n\right) = 0$$

Démonstration: $\limsup_{n \rightarrow +\infty} f_n = \bigcap_{n \in \mathbb{N}^*} \bigcup_{m \geq n} f_m$

Donc $\forall m \in \mathbb{N}$, $\limsup_{n \rightarrow +\infty} f_n \subset \bigcup_{k \geq m} f_k$

$$\text{Par } \bar{P} \text{ de } P: 0 \leq P\left(\limsup_{n \rightarrow +\infty} f_n\right) \leq P\left(\bigcup_{k \geq m} f_k\right) \leq \sum_{k=m}^{+\infty} P(f_k) \xrightarrow{m \rightarrow \infty} 0$$

comme rest d'une série CV

i.e: ps, seul un nb fini de f_n est réalisé

si $P(\limsup f_n) = 1 \Leftrightarrow$ ps, une infé de f_n est réalisée.