

Rappels

On rappelle les éléments suivants du cours de Probabilités, considérés comme acquis par la suite:

- **Loi (forte) des grands nombres:** Si $X_1, \dots, X_n$ sont des v.a. i.i.d. sur $\mathbb{R}$ satisfaisant $\mathbb{E}|X_1| < +\infty$, alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(X_1).$$

- **Théorème central limite:** Si $X_1, \dots, X_n$ sont des v.a. i.i.d. sur $\mathbb{R}$, satisfaisant $\mathbb{E}|X_1|^2 < +\infty$, alors

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \xrightarrow[n \rightarrow +\infty]{\sim} \mathcal{N}(0, \text{Var}(X_1)).$$

On rappelle aussi les définitions équivalentes de la convergence en loi (notée $\rightsquigarrow$): $X_n \rightsquigarrow X$ ssi, de manière équivalente

1. $\forall f \in C_b(\mathbb{R}) \quad \mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$.
2. Si F_X est continue en t , $F_{X_n}(t) \rightarrow F_X(t)$ (fonctions de répartition).
3. Pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ (fonctions caractéristiques).

- **Inégalité de Markov:** Si X est positive, alors, pour tout $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Exercices: outils de base en concentration

EXERCICE 1 (Autour de Byenaymé Tchebychev). — Soit X une variable réelle satisfaisant $\mathbb{E}(X^2) < +\infty$.

1. Montrer l'inégalité de Byenaymé Tchebychev: $\forall t > 0 \quad \mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}$.
2. Montrer l'inégalité de Cantelli: $\forall t > 0 \quad \mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \frac{\text{Var}(X)}{\text{Var}(X) + t^2}$.
3. Comparer ces deux inégalités.
4. Montrer l'inégalité de Paley-Zygmund: en supposant de plus $X \geq 0$,

$$\forall t \in [0, 1] \quad \mathbb{P}(X \geq t\mathbb{E}(X)) \geq (1-t)^2 \frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}.$$

EXERCICE 2 (Inégalité de Hoeffding). — On va commencer par prouver le Lemme de Hoeffding: si Z est une v.a. telle que $Z \in [a, b]$ p.s. et $\mathbb{E}(Z) = 0$, alors

$$\forall t \in \mathbb{R} \quad \psi_Z(t) \leq \frac{(b-a)^2 t^2}{8},$$

où $\psi_Z(t) := \log(\mathbb{E}(\exp(tZ)))$. Dans la suite on se donne une telle v.a. Z .

1. Montrer que Ψ_Z est bien définie.

2. Montrer que, si Y est à valeurs dans $[a, b]$, alors

$$\text{Var}(Y) \leq \frac{(b-a)^2}{4}.$$

3. En déduire que, pour tout $t \in \mathbb{R}$, $\psi''(t) \leq (b-a)^2/4$.

4. Conclure.

Dans une deuxième partie on va prouver l'inégalité de Hoeffding. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que $X_i \in [a_i, b_i]$ p.s. et $\mathbb{E}(X_i) = 0$. On note $S = \sum_{i=1}^n X_i$.

5. Montrer que, pour tout $\varepsilon \geq 0$ et $t \in \mathbb{R}$,

$$\mathbb{P}(S \geq \varepsilon) \leq \exp(-t\varepsilon + \psi_S(t)).$$

6. En choisissant soigneusement t , montrer que, pour tout $\varepsilon \geq 0$,

$$\mathbb{P}(S \geq \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

ce qui constitue l'inégalité de Hoeffding.

7. Application: si $X_1, \dots, X_n$ sont i.i.d. de loi $\mathcal{B}(p)$, montrer que, pour tout $\varepsilon \geq 0$,

$$\mathbb{P}(|\bar{X}_n - p| \geq \varepsilon) \leq 2 \exp(-2n\varepsilon^2).$$

Exercice: un résultat utile de convergence

EXERCICE 3 (Convergence des quantiles). — Soit X v.a. sur $\mathbb{R}$ de fonction de répartition F . Pour $q \in [0, 1]$, le **quantile** d'ordre q de X est défini par

$$q(p) = \inf \{t \in \mathbb{R} \mid F(t) \geq p\}.$$

Soient $X_1, \dots, X_n$ i.i.d. de même loi que X . On note $q_n(p)$ le quantile empirique (associé à la fonction de répartition empirique $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t}$). On se donne $p \in [0, 1]$, et on suppose que F est strictement croissante en q .

1. En notant $X_{(1)} \leq \dots \leq X_{(n)}$ l'échantillon trié (attention plus i.i.d.), montrer que $q_n = X_{[np]}$.

2. Soit $\varepsilon > 0$. Montrer que

$$\mathbb{P}(q_n \leq q - \varepsilon) \leq \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq q - \varepsilon} \geq [np]\right).$$

3. En déduire l'existence de $\delta_-(\varepsilon) > 0$ tel que

$$\mathbb{P}(q_n \leq q - \varepsilon) \leq \exp(-2n\delta_-(\varepsilon)^2).$$

4. Soit $\varepsilon > 0$. Montrer que

$$\mathbb{P}(q_n \leq q + \varepsilon) \leq \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq q + \varepsilon} \leq [np]\right).$$

5. En déduire l'existence de $\delta_+(\varepsilon) > 0$ tel que, pour n assez grand,

$$\mathbb{P}(q_n \geq q + \varepsilon) \leq \exp(-2n\delta_+(\varepsilon)^2).$$

6. En déduire que $q_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} q$.

Correction TDO STMA.

Exercice 1:

$$1) P(|X - E(X)| \geq t) = P(|X - E(X)|^2 \geq t^2)$$

$|X - E(X)|^2$ est une variable aléatoire positive ; d'après l'inégalité de Markov :

$$P(|X - E(X)| \geq t) \leq \frac{E(|X - E(X)|^2)}{t^2} = \frac{\text{Var}(X)}{t^2}.$$

2) Deux solutions : - Cauchy-Schwarz
- Markov

Cauchy - Schwarz :

$$\text{On pose } Y = E(X) - X + t.$$

$$\text{Alors } 0 < t = E[Y].$$

$$\text{Donc } 0 < t = E[Y] \leq E[Y \mathbb{1}_{Y \geq 0}].$$

On peut passer au cas où car tout est positif.

$$t^2 \leq \mathbb{E}[Y \mathbb{1}_{Y \geq 0}] \leq \mathbb{E}[Y^2] P(Y \geq 0)$$

Cauchy-Schwarz

ok car $Y \in L^2$

Rappel: $\mathbb{E}(\mathbb{1}_A) = P(A)$

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)} \sqrt{\mathbb{E}(Y^2)}$$

$$\begin{aligned} \text{Or } \mathbb{E}(Y^2) &= \text{Var}(Y) + \mathbb{E}(Y)^2 \\ &= \text{Var}(X) + t^2 \end{aligned}$$

car $Y = X + ct$

$$\text{Donc } P(Y \geq 0) \geq \frac{t^2}{t^2 + \text{Var}(X)}$$

On passant au complémentaire :

$$P(Y \leq 0) = P(X \geq \mathbb{E}(X) + t) \leq \frac{\text{Var}(X)}{t^2 + \text{Var}(X)}$$

Markov:

qu'il faut à poser $\tilde{X} = X - \mathbb{E}(X)$, on montre le résultat pour X centrée.

Soit $u > 0$:

$$P(X \geq t) = P(X+u \geq t+u)$$

Comme $t+u \geq 0$,

$$P(X \geq t) = P((X+u)^2 \geq (t+u)^2)$$

$$\stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}((X+u)^2)}{(t+u)^2} = \frac{\text{Var}(X) + u^2}{(t+u)^2}$$

Le membre de gauche ne dépend pas de u , on peut optimiser en u , i.e trouver la valeur de u qui rend le borne optimale. Après étude de la fonction, on trouve $u = \sqrt{\text{Var}(X)}$ qui est bien positif et cela donne le résultat voulu.

3) Pour $t > 0$, Cantelli est mieux et la borne est ≤ 1 (ceci pour une borne de proba).

En appliquant Cantelli $0 - X$ et en sommant les bornes, on a

$$P(|X - E(X)| > t) \leq 2 \frac{\text{Var}(X)}{t^2 + \text{Var}(X)},$$

mieux pour $t < \text{Var}(X)$ mais en fait dans ce cas la borne est triviale (≥ 1).

4) On utilise Cauchy-Schwarz :

$$E(X) = E(X \mathbb{1}_{X \leq t E(X)}) + E(X \mathbb{1}_{X > t E(X)})$$

$$\leq t E(X) + E(X^2)^{1/2} P(X > t E(X))^{1/2}.$$

Et puisque $E(X) \geq 0$ et $t \in (0, 1)$ on peut isoler $P(X > t E(X))^{1/2}$, passer au carré et on a le résultat.

Exercice 2:

1) Z est bornée par $M = \max(|a|, |b|)$ donc $e^t Z$ est bornée pour tout t donc admet une espérance, ce qui assure que ψ_2 est définie pour tout $t \in \mathbb{R}$.

2) Cela part du résultat suivant

lemme: $E(Z) = \argmin_{a \in \mathbb{R}} E((X-a)^2)$

preuve: Soit $a \in \mathbb{R}$:

$$E((X-a)^2) = \text{Var}(X-a) + E(X-a)^2$$

$$= \text{Var}(X) + (E(X) - a)^2$$



Donc, $\text{Var}(Y) \leq E\left(\left(Y - \frac{a+b}{2}\right)^2\right)$

et $\left|Y - \frac{a+b}{2}\right| \leq \frac{b-a}{2}$ ps donc

$$\text{Var}(Y) \leq \left(\frac{b-a}{2}\right)^2 = \frac{(b-a)^2}{4}.$$

3) Calculons les deux premières dérivées de Ψ_2 . Pour cela, il faut appliquer le théorème de dérivation sous l'intégrale.

La fonction $(t, z) \mapsto e^t z$ est

P_2 -intégrable en z et dérivable en t ,
de dérivée
 $ze^t z$.

P_2 -ps, on a $|ze^t z| \leq \eta e^{t\eta}$.

Ainsi, si $t \in [-T, T]$, $|ze^t z| \leq \eta e^{T\eta}$,

qui est P_2 -intégrable indépendamment de t
car P_2 est une mesure de proba.

Par le théorème, Ψ_2 est dérivable sur $[-T, T]$
pour tout T donc sur $\mathbb{R}$ avec :

$$\forall t \in \mathbb{R}, \psi_2'(t) = \frac{\mathbb{E}[Ze^{tZ}]}{\mathbb{E}[e^{tZ}]}.$$

Avec des arguments similaires, on montre que $\psi \in C^2$, avec:

$$\forall t \in \mathbb{R}, \psi_2''(t) = \frac{\mathbb{E}[Z^2 e^{tZ}]}{\mathbb{E}[e^{tZ}]} - \left(\frac{\mathbb{E}[Ze^{tZ}]}{\mathbb{E}[e^{tZ}]} \right)^2.$$

$$= \mathbb{E}[Z^2 f_t(Z)] - \mathbb{E}[Z f_t(Z)]^2,$$

où $f_t(z) = \frac{e^{tz}}{\mathbb{E}[e^{tZ}]}$ est une densité de proba.

$$\text{Donc } \psi_2''(t) = \text{Var}(Y),$$

où Y a pour loi $dP_Y(y) = f_t(y) dP_Z(y)$.

(Autrement dit, Y a pour densité f_t par rapport à la loi de Z) CF RAPPEL À LA FIN

En particulier, $\forall \epsilon \in [a, b]$ ps et donc 2) donne le résultat.

4) On peut intégrer deux fois le résultat de 3, en remarquant que $\psi(0) = 0$ et que $\psi'_2(0) = E(Z) = 0$ par ² hypothèse.

↳ de pour $t \geq 0$, pour $t < 0$, utiliser $\psi_2(t) = \psi_2(-t)$

5) D'après l'inégalité de Markov:

$$\begin{aligned} P(S \geq \epsilon) &= P(e^{tS} \geq e^{t\epsilon}) \\ &\leq \frac{E(e^{tS})}{e^{t\epsilon}} = e^{\psi_S(t) - t\epsilon} \end{aligned}$$

6) D'après la question 4, on a

$$\psi_S(t) = \sum_{i=1}^n \psi_{X_i}(t) \leq \frac{t^2}{8} \sum_{i=1}^n (b_i - a_i)^2.$$

↑
indépendance

$$\text{On en déduit: } P(S \geq \epsilon) \leq e^{\frac{t^2}{8} \sum (b_i - a_i)^2 - t\epsilon}.$$

À gauche, cela ne dépend plus de t donc on optimise $-b + \text{optimal}$ et $\frac{4\varepsilon}{\sum (b_i - a_i)^2}$ et on obtient le résultat.

7) Il s'agit d'une application de 6, avec

$$Y_i = X_i - E(X_i), \quad |Y_i| \leq 1 \text{ ps.}$$

On applique l'inégalité à $S = \sum Y_i$ puis à $-S$, puis on écrit :

$$\begin{aligned} P(|\bar{X}_n - p| \geq \varepsilon) &\leq P(\bar{X}_n - p \geq \varepsilon) + P(\bar{X}_n - p \leq -\varepsilon) \\ &\leq e^{-2n\varepsilon^2} + e^{-2n\varepsilon^2} = 2e^{-2n\varepsilon^2}. \end{aligned}$$

Remarque : En comparaison avec Bernoulli. Recherche, cette borne est bien meilleure quand $n \rightarrow \infty$; la décroissance est beaucoup plus rapide.

Rappel sur les variables à densité:

Soient μ et ν deux mesures réelles.

On dit que ν admet une densité par rapport à μ s'il existe une fonction mesurable positive f telle que pour tout A mesurable:

$$\nu(A) = \int \mathbb{1}_A d\nu = \int \mathbb{1}_A f d\mu.$$

Dans ce cas, on note $f = \frac{d\nu}{d\mu}$, de sorte à

$$\text{pouvoir écrire } \nu(A) = \int \mathbb{1}_A \frac{d\nu}{d\mu} d\mu = \int \mathbb{1}_A d\nu.$$

Dans la plupart des cas, on a $\mu = \text{Lebesgue}$, et quand on dit "variable à densité", on sous-entend "à densité par rapport à la mesure de Lebesgue".

Mais dans notre cas à nous, l'astuce était d'écrire $\psi'_2(t)$ comme

$$E[Z^2 f_t(Z)] - E[Z f_t(Z)]^2$$

et de reconnaître que f_t est une densité.

Si on l'écrit avec des intégrales, cela donne :

$$\Psi''_2(t) = \int Z^2 f_t(z) dP_Z(z) - \left(\int Z f_t(z) dP_Z(z) \right)^2.$$

En posant $dP_Y(z) = f_t(z) dP_Z(z)$, ce qui correspond à introduire une variable Y dont la densité par rapport à P_Z est f_t , on peut réécrire cela :

$$\Psi''_2(t) = \int Z^2 dP_Y(z) - \left[\int Z dP_Y(z) \right]^2$$

$$= E(Y^2) - E(Y)^2$$

$$= \text{Var}(Y).$$

Autre rappel: théorème de Radon-Nikodym.

Si μ est σ -fini, alors on a l'équivalence:

$$[\mu(A)=0 \Rightarrow \nu(A)=0] \text{ssi}$$

$$[\nu \text{ a une densité par rapport à } \mu].$$

En conséquence, puisque

$$P(Z \notin (a,b)) = 0$$

et que ν a une densité par rapport à Z ,
on a

$$P(Y \notin (a,b)) = 0.$$

Cela se voit simplement:

$$P(Y \in (a,b)) = \int_a^b dP_Y = \int_a^b f_Y(z) dP_Z(z)$$

$$\text{Donc } Y \in (a,b) \text{ et } \gamma_Z''(t) = \text{Var}(Y) \leq \frac{(b-a)^2}{4}.$$

