Université Paris-Dauphine **2022-2023**
Cours de “Intégrale de Lebesgue et probabilités" Décembre 2022

Chapitres 8 - Conditionnement

Table des matières

Dans ce chapitre on désigne par $(\Omega, \mathcal{A}, P)$ un espace probabilisé, c'est-à-dire, un espace mesurable $(\Omega, \mathcal{A})$ muni d'une probabilité P. Sont écrites en rouge les parties hors programme, en violet les parties traitées en TD (résultats à connaitre pour sa culture) et en orange les parties modifiées (précisées) par rapport au cours en amphi.

1 Conditionnement $\mathcal{L}^1$

Dans cette section, nous définissons la notion d'espérance conditionnelle dans un cadre $\mathcal{L}^1$. Une extension possible est de définir cette notion pour des va positives (sans condition d'intégrabilité). Etant donnée une sous-tribu $\mathcal{C} \subset \mathcal{A}$, on note $L^p(\mathcal{C}) := L^p(\Omega, \mathcal{C}, P)$.

Théorème 1.1. *Soient $\mathcal{B} \subset \mathcal{A}$ une sous-tribu et X une var $\mathcal{L}^1(\mathcal{A})$. Il existe une var $Y \in L^1(\mathcal{B})$ unique, notée $Y = \mathbf{E}(X|\mathcal{B})$ et appellée espérance conditionnelle de X par rapport à $\mathcal{B}$, telle que*

$$\mathbf{E}(Y\, Z) = \mathbf{E}(X\, Z) \qquad \forall\, Z = \mathbf{1}_B,\, B \in \mathcal{B}, \text{ (et donc } \forall\, Z \in L^\infty(\mathcal{B})). \tag{1}$$

Si $X \in \mathcal{L}^2(\mathcal{A})$ alors $Y \in L^2(\mathcal{B})$ est la solution du problème de minimisation

$$\mathbf{E}((Y - X)^2) = \min_{Z \in L^2(\mathcal{B})} \mathbf{E}((Z - X)^2). \tag{2}$$

Preuve du Théorème 1.1. Commeçons par l'unicité. Supposons qu'il existe deux var Y_1 et Y_2 de $L^1(\Omega, \mathcal{B}, P)$ telles que

$$\forall\, B \in \mathcal{B}, \quad \mathbf{E}(Y_1\, \mathbf{1}_B) = \mathbf{E}(Y_2\, \mathbf{1}_B) = \mathbf{E}(X\, \mathbf{1}_B).$$

En prenant $B := \{Y_1 - Y_2 > 0\} \in \mathcal{B}$, on trouve

$$\forall\, B \in \mathcal{B}, \quad \mathbf{E}((Y_1 - Y_2)\, \mathbf{1}_{\{Y_1 - Y_2 > 0\}}) = 0,$$

ce qui prouve $Y_1 - Y_2 \le 0$ p.s.. Et on montre de la même manière $Y_1 - Y_2 \ge 0$ p.s..

Pour l'existence, nous commeçons par supposer $X \ge 0$. Soit alors Q la mesure définie sur $(\Omega, \mathcal{B})$ par

$$\forall\, B \in \mathcal{B}, \quad Q(B) := \int_B X dP.$$

On voit clairement que $Q << P$ sur $(\Omega, \mathcal{B})$. Le Théorème de Radon-Nikodym assure l'existence d'une variable aléatoire positive Y sur $(\Omega, \mathcal{B})$ telle que

$$\forall\, B \in \mathcal{B}, \quad Q(B) = \int_B Y dP,$$

ce qui est bien le résultat annoncé. Si de plus, $X \in L^1(\mathcal{A})$ alors $Y \in L^1(\mathcal{B})$, ce que l'on voit en prenant $B = \Omega$.

Si maintenant, on suppose juste $X \in \mathcal{L}^1(\mathcal{A})$, on écrit $X = X_+ - X_-$ avec $0 \le X_\pm \in \mathcal{L}^1(\mathcal{A})$ et l'étape précédente nous dit qu'il existe $0 \le Y_\pm \in L^1(\mathcal{B})$ telles que $\mathbf{E}(Y_\pm \mathbf{1}_B) = \mathbf{E}(X_\pm \mathbf{1}_B)$. On en déduit que $Y := Y_+ - Y_-$ convient.

On suppose enfin $X \in \mathcal{L}^2(\mathcal{A})$. En prenant $Z := Y\,\mathbf{1}_{|Y|\le n} \in L^\infty(\mathcal{B})$ dans (1), on obtient

$$\mathbf{E}(Y^2 \mathbf{1}_{|Y|\le n}) = \mathbf{E}(Y\,(Y\mathbf{1}_{|Y|\le n})) = \mathbf{E}(X\,Y\mathbf{1}_{|Y|\le n}) \le \mathbf{E}(X^2)^{1/2}\,\mathbf{E}(Y^2\mathbf{1}_{|Y|\le n})^{1/2},$$

et donc

$$\mathbf{E}(Y^2) = \lim_{n\to\infty} \mathbf{E}(Y^2 \mathbf{1}_{|Y|\le n}) \le \mathbf{E}(X^2).$$

De plus, pour tout $Z \in L^2(\mathcal{B})$, on a

$$\begin{aligned} \mathbf{E}((X-Z)^2) &= \mathbf{E}((X-Y)^2) + 2\mathbf{E}((X-Y)(Y-Z)) + \mathbf{E}((Y-Z)^2) \\ &= \mathbf{E}((X-Y)^2) + \mathbf{E}((Y-Z)^2), \end{aligned}$$

puisque $Y - Z$ est $\mathcal{B}$ mesurable. On en déduit immédiatement (2). $\square$

Les principales propriétés de l'espérance conditionnelle sont résumées ci-dessous

Théorème 1.2. *Soient $\mathcal{B}, \mathcal{C} \subset \mathcal{A}$ deux sous-tribus et X, Y deux var $\mathcal{L}^1(\mathcal{A})$. Alors*

- $\mathbf{E}(\lambda\, X + Y|\mathcal{B}) = \lambda\,\mathbf{E}(X|\mathcal{B}) + \mathbf{E}(Y|\mathcal{B});$ (3)
- $X \le Y \quad \Rightarrow \quad \mathbf{E}(X|\mathcal{B}) \le \mathbf{E}(Y|\mathcal{B});$ (4)

- $\mathbf{E}(\mathbf{E}(X|\mathcal{B})) = \mathbf{E}(X), \quad \mathbf{E}(1|\mathcal{B}) = 1;$ (5)
- $\mathbf{E}(X|\mathcal{B}) = \mathbf{E}(X) \quad$ si $\quad X \perp\!\!\!\perp \mathcal{B};$ (6)
- $\mathbf{E}(X\,Y|\mathcal{B}) = Y\,\mathbf{E}(X|\mathcal{B}),\ \mathbf{E}(Y|\mathcal{B}) = Y \quad$ si $\quad Y$ est $\mathcal{B}$-*mesurable*; (7)
- $\mathbf{E}(\mathbf{E}(X|\mathcal{C})|\mathcal{B}) = \mathbf{E}(\mathbf{E}(X|\mathcal{B})|\mathcal{C}) = \mathbf{E}(X|\mathcal{C}) \quad$ si $\quad \mathcal{C} \subset \mathcal{B};$ (8)

- $\varphi(\mathbf{E}(X|\mathcal{B})) \le \mathbf{E}(\varphi(X)|\mathcal{B}) \quad$ si φ est une fonction convexe, (9)

ce qui implique $|\mathbf{E}(X|\mathcal{B})|^p \le \mathbf{E}(|X|^p|\mathcal{B})$ *et* $\mathbf{E}(X|\mathcal{B}) \in L^p$ *si* $X \in L^p$; (10)
- $\mathbf{E}(X_n|\mathcal{B}) \to \mathbf{E}(X|\mathcal{B})$ en norme $L^p \quad$ si $\quad X_n \to X$ en norme L^p; (11)

- $\mathbf{E}(X_n|\mathcal{B}) \nearrow \mathbf{E}(X|\mathcal{B}) \quad$ si $\quad X_n \nearrow X$ p.s.; (12)
- $\mathbf{E}(\liminf X_n|\mathcal{B}) \le \liminf \mathbf{E}(X_n|\mathcal{B}) \quad$ si $\quad X_n \ge 0;$ (13)
- $\mathbf{E}(X_n|\mathcal{B}) \to \mathbf{E}(X|\mathcal{B})$ p.s. et $L^1 \quad$ si $\quad X_n \to X$ p.s. et $|X_n| \le Z \in L^1$. (14)

Preuve du Théorème 1.2. (3) En notant $X' = \mathbf{E}(X|\mathcal{B})$ et $Y' = \mathbf{E}(Y|\mathcal{B})$, pour tout $Z \in L^\infty(\mathcal{B})$, on a

$$\mathbf{E}((\lambda X' + Y')Z) = \lambda \mathbf{E}(X'Z) + \mathbf{E}(Y'Z) = \lambda \mathbf{E}(XZ) + \mathbf{E}(YZ) = \mathbf{E}((\lambda X + Y)Z).$$

(4) Si $X \ge 0$, en notant $Y := \mathbf{E}(X|\mathcal{B})$ et $B := \{Y \le 0\}$, on a

$$0 \le \mathbf{E}(X\mathbf{1}_B) = \mathbf{E}(Y\mathbf{1}_B) \le 0 \text{ et } Y\mathbf{1}_B \le 0,$$

de sorte que $Y\mathbf{1}_B = 0$ et donc $Y \geq 0$.

(5) On a $\mathbf{E}(\mathbf{E}(X|\mathcal{B})) = \mathbf{E}(X)$ en prenant $Z = 1$ dans la définition (1) de l'espérance conditionnelle. L'identité $\mathbf{E}(1|\mathcal{B}) = 1$ est immédiate.

(6) Si $X \perp\!\!\!\perp \mathcal{B}$, on a $\mathbf{E}(XZ) = \mathbf{E}(X)\mathbf{E}(Z) = \mathbf{E}(\mathbf{E}(X)Z)$ pour tout $Z \in L^\infty(\mathcal{B})$, ce qui prouve bien $\mathbf{E}(X|\mathcal{B}) = \mathbf{E}(X)$.

(7) Si $Y \in L^\infty(\mathcal{B})$, pour $Z \in L^\infty(\mathcal{B})$, on a

$$\mathbf{E}(\mathbf{E}(X|\mathcal{B})YZ) = \mathbf{E}(XYZ) = \mathbf{E}(\mathbf{E}(XY|\mathcal{B})Z),$$

ce qui prouve $\mathbf{E}(X\,Y|\mathcal{B}) = Y\,\mathbf{E}(X|\mathcal{B})$. En particulier, en prenant $X = 1$, on a $\mathbf{E}(Y|\mathcal{B}) = Y$.

(8) Si $\mathcal{C} \subset \mathcal{B}$, alors $\mathbf{E}(X|\mathcal{C})$ est $\mathcal{B}$-mesurable, et donc $\mathbf{E}(\mathbf{E}(X|\mathcal{C})|\mathcal{B}) = \mathbf{E}(X|\mathcal{C})$ d'après (7). Pour $W \in L^\infty(\mathcal{C}) \subset L^\infty(\mathcal{B})$, on calcule également

$$\mathbf{E}(\mathbf{E}(\mathbf{E}(X|\mathcal{B})|\mathcal{C})W) = \mathbf{E}(\mathbf{E}(X|\mathcal{B})W) = \mathbf{E}(XW) = \mathbf{E}(\mathbf{E}(X|\mathcal{C})W),$$

où on a successivement utilisé la définition de $\mathbf{E}(\mathbf{E}(X|\mathcal{B})|\mathcal{C})$, de $\mathbf{E}(X|\mathcal{B})$ et de $\mathbf{E}(X|\mathcal{C})$.

(9) Notons

$$A_\varphi := \{(a,b) \in \mathbb{R}^2;\, \forall x \in \mathbb{R},\ \varphi(x) \geq ax + b\}.$$

Il est clair que

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \sup_{(a,b)\in A_\varphi} (ax+b) = \sup_{(a,b)\in A_\varphi \cap \mathbb{Q}^2} (ax+b).$$

Du fait que $A_\varphi \cap \mathbb{Q}^2$ est dénombrable, on en déduit que p.s.

$$\mathbf{E}[\varphi(X)|\, \mathcal{B}] = \mathbf{E}\big[\sup_{(a,b)\in A_\varphi\cap\mathbb{Q}^2} (aX+b)|\, \mathcal{B}\big] \geq \sup_{(a,b)\in A_\varphi\cap\mathbb{Q}^2} \mathbf{E}[aX+b|\, \mathcal{B}] = \varphi(\mathbf{E}[X|\mathcal{B}]).$$

L'assertion (10) s'en déduit immédiatement.

(11) De (10), on déduit

$$\mathbf{E}(|\mathbf{E}(X_n|\mathcal{B}) - \mathbf{E}(X|\mathcal{B})|^p) = \mathbf{E}(|\mathbf{E}(X_n - X|\mathcal{B})|^p) \leq \mathbf{E}(|X_n - X|^p) \to 0,$$

si $X_n \to X$ dans L^p.

(12) Si $X_n \nearrow X$ p.s. on a $Y_n := \mathbf{E}(X_n|\mathcal{B}) \nearrow Y$ p.s. avec Y $\mathcal{B}$-mesurable. D'après le théorème de convergence monotone, pour tout $B \in \mathcal{B}$, on a

$$\mathbf{E}(Y\mathbf{1}_B) = \lim \mathbf{E}(Y_n\mathbf{1}_B) = \lim \mathbf{E}(X_n\mathbf{1}_B) = \mathbf{E}(X\mathbf{1}_B),$$

ce qui dit bien que $Y = \mathbf{E}(X|\mathcal{B})$. Les assertions (13) et (14) s'en déduisent comme dans le chapitre sur les théorèmes de passage à la limite. $\square$

Exercice 1.3. *Faire la démonstration détaillée de (13) et (14).*

On peut préciser le point (6).

Théorème 1.4. *Deux sous tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes si, et seulement si,*

$$\mathbf{E}(X|\mathcal{B}_1) = \mathbf{E}(X) \text{ pour toute var } X\ \mathcal{B}_2\text{-mesurable.}$$

Preuve du Théorème 1.4. Le sens direct n'est rien d'autre que (6). Montrons la réciproque et supposons que

$$\mathbf{E}(X|\mathcal{B}_1) = \mathbf{E}(X) \text{ pour toute var } X\ \mathcal{B}_2\text{-mesurable.}$$

Pour $Z \in L^\infty(\mathcal{B}_1)$ on a donc

$$\mathbf{E}(XZ) = \mathbf{E}(\mathbf{E}(X|\mathcal{B}_1)Z) = \mathbf{E}(\mathbf{E}(X)Z) = \mathbf{E}(X)\mathbf{E}(Z),$$

ce qui est bien dire que $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes. $\square$

Définition 1.5. *Soit X, Y deux var. On appelle espérance conditionnelle de Y sachant X, la var* $\mathbf{E}(Y|X) := \mathbf{E}(Y|\sigma(X))$.

D'après le Théorème I.4.10, il existe une fonction φ telle que $\mathbf{E}(Y|X) = \varphi(X)$. On peut donc voir l'espérance conditionnelle comme une façon de definir une "meilleur fonction" φ de sorte que $\varphi(X)$ "soit proche" de Y.

Corollaire 1.6. *Deux variables aléatoires X et Y sont indépendantes si, et seulement si,*

$$\mathbf{E}(h(X)|Y) = \mathbf{E}(h(X)), \quad \forall\, h \,\text{mesurables.}$$

Théorème 1.7. *Soit $\mathcal{B}$ une sous-tribu et soient X et Y deux variables aléatoires telles que X est indépendante de $\mathcal{B}$ et Y est $\mathcal{B}$-mesurable. Pour toute fonction mesurable et bornée φ, on a*

$$\mathbf{E}(\varphi(X,Y)|\mathcal{B}) = \Phi(Y), \quad \Phi(y) := \int \varphi(x,y) P_X(dx),$$

où P_X désigne la loi de X.

Preuve du Théorème 1.7. Soit $Z \in \mathcal{L}^\infty(\mathcal{B})$ et notons $P_{(X,Y,Z)}$ la loi du triplet (X,Y,Z). Comme X est indépendante de (Y,Z), on a $P_{(X,Y,Z)} = P_X \otimes P_{(Y,Z)}$. Grâce au théorème de Fubini, on a

$$\begin{aligned}
\mathbf{E}(\varphi(X,Y)Z) &= \int \varphi(x,y) z P_X \otimes P_{(Y,Z)}(dx,dy,dz) \\
&= \int z\Big(\int \varphi(x,y) P_X(dx)\Big) P_{(Y,Z)}(dy,dz) \\
&= \int z\Phi(Y) P_{(Y,Z)}(dy,dz) = \mathbf{E}(\Phi(Y)Z).
\end{aligned}$$

Cela démontre bien le résultat annoncé. □

Exercice 1.8. *Soient $\mathcal{B}$ et $\mathcal{C}$ deux sous-tribus et X une var $\mathcal{L}^1$. On suppose $\mathcal{C} \perp\!\!\!\perp \sigma(\sigma(X), \mathcal{B})$.*
a) Montrer que $\mathbf{E}(X\mathbf{1}_{B\cap C}) = \mathbf{E}(X\mathbf{1}_B)\mathbf{P}(C)$, pour tout $B \in \mathcal{B}$ et $C \in \mathcal{C}$.
b) Montrer que $\mathcal{D} := \{B \cap C;\, B \in \mathcal{B},\, C \in \mathcal{C}\}$ est une algèbre et que $\sigma(\mathcal{D}) = \sigma(\mathcal{B}, \mathcal{C})$.
c) En déduire que $\mathbf{E}(X|\sigma(\mathcal{B},\mathcal{C})) = \mathbf{E}(X|\sigma(\mathcal{B}))$.

2 Calculs d'espérance conditionnelle

• *Le cas discret.* Soit X une va à valeurs dans un espace dénombrable E et soit $Y \in \mathcal{L}^1(\mathcal{A})$. On introduit l'ensemble

$$E' := \{x \in E;\, P(X = x) > 0\},$$

la fonction

$$\varphi(x) := \frac{\mathbf{E}(Y\,\mathbf{1}_{X=x})}{P(X=x)} \text{ si } x \in E', \;\; := 0 \text{ sinon,}$$

et enfin la var $\varphi(X)$. Pour $Z := \mathbf{1}_{X=a}$, $a \in E'$, on calcule

$$\begin{aligned}
\mathbf{E}(\varphi(X)Z) &= \mathbf{E}\big(\sum_{x\in E'} \varphi(x')\mathbf{1}_{X=x'}\,\mathbf{1}_{X=a}\big) \\
&= \varphi(a)\,\mathbf{E}\big(\mathbf{1}_{X=a}\big) = \mathbf{E}(Y\,\mathbf{1}_{X=a}) = \mathbf{E}(YZ).
\end{aligned}$$

Comme E est discret, cela suffit pour dire que cette identité est vraie pour tout Z qui est $\sigma(X)$-mesurable. On vient donc de démontrer que $E(Y|X) = \varphi(X)$.

Exemple. On prend $\Omega := \{1,\dots,6\}$ et $P(\{\omega\}) = 1/6$ pour tout $\omega\in\Omega$. On définit

$$X(\omega) := \begin{cases} 1 & \text{si } \omega \text{ est impair,}\\ 0 & \text{si } \omega \text{ est pair,}\end{cases}$$

et $Y(\omega) = \omega$. Alors

$$\mathbf{E}(Y|X)(\omega) = \varphi(X(\omega)) = \begin{cases} 3 & \text{si } \omega \text{ est impair,}\\ 4 & \text{si } \omega \text{ est pair,}\end{cases}$$

puisque par exemple

$$\varphi(1) = \frac{\mathbf{E}(\sum_\omega \omega\,\mathbf{1}_{X(\omega)=1})}{P(X=1)} = \frac{\sum_{\omega\text{ impair}} \omega P(\{\omega\})}{1/2} = \frac{1+3+5}{3}.$$

• *Le cas continu.* Soient X et Y deux v.a.r. telles que le couple (X,Y) a pour densité $f_{X,Y}(x,y)$. On observe en particulier que la var Y a pour densité

$$f_Y(y) := \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy$$

et on pose

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathbf{1}_{f_Y(y)>0}.$$

Pour tout $\varphi,\psi\in\mathcal{L}^\infty(\mathbb{R})$, on calcule

$$\begin{aligned}\mathbf{E}(\varphi(X)\psi(Y)) &= \int_{\mathbb{R}^2} \varphi(x)\psi(y)\,f_{X,Y}(x,y)\,dxdy\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x)\,f_{X|Y}(x|y)dx\,\psi(y)f_Y(y)\mathbf{1}_{f_Y(y)>0}dy\\ &= \int_{\mathbb{R}} \Phi(y)\,\psi(y)f_Y(y)\,dy = \mathbf{E}(\Phi(Y)\psi(Y)),\end{aligned}$$

où on a posé

$$\Psi(y) := \int_{\mathbb{R}} \varphi(x)f_{X,Y}(x,y)\,dx.$$

On a ainsi démontré que $\mathbf{E}(\varphi(X)|Y) = \Phi(Y)$.

3 Asymptotique

Théorème 3.1. *Soient (X_n) une suite de var de $\mathcal{L}^2$, $(\mathcal{F}_n)$ une suite croissante de sous-tribus telle que $\mathcal{F}_n\supset\sigma(X_1,\dots,X_n)$. On suppose que*

$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n,\ \forall\, n\ge 1,\quad et\quad \sup_{n\ge1}\mathbf{E}(X_n^2)<\infty. \tag{15}$$

Alors, il existe $X_\infty\in\mathcal{L}^2$ telle que $X_n\to X_\infty$ au sens $\mathcal{L}^2$ et p.s. lorsque $n\to\infty$.

Preuve du Théorème 3.1. **(1)** Pour $m\ge n+2$ et à l'aide de (8), on a

$$\mathbf{E}(X_m|\mathcal{F}_n) = \mathbf{E}(\mathbf{E}(X_m|\mathcal{F}_{m-1})|\mathcal{F}_n) = \mathbf{E}(X_{m-1}|\mathcal{F}_n),$$

puis par récurrence $\mathbf{E}(X_m|\mathcal{F}_n) = X_n$ pour tout $m\ge n$. On en déduit

$$\begin{aligned}\mathbf{E}((X_m-X_n)^2|\mathcal{F}_n) &= \mathbf{E}(X_m^2|\mathcal{F}_n) - 2\mathbf{E}(X_mX_n|\mathcal{F}_n) + \mathbf{E}(X_n^2|\mathcal{F}_n)\\ &= \mathbf{E}(X_m^2|\mathcal{F}_n) - 2X_n\mathbf{E}(X_m|\mathcal{F}_n) + X_n^2\\ &= \mathbf{E}(X_m^2|\mathcal{F}_n) - X_n^2,\end{aligned}$$

en utilisant deux fois (7) à la deuxième égalité. En prenant l'espérance de cette dernière expression et en utilisant (5), on obtient

$$\mathbf{E}((X_m - X_n)^2) = \mathbf{E}(X_m^2) - \mathbf{E}(X_n^2), \tag{16}$$

soit donc en particulier

$$\mathbf{E}((X_{k+1} - X_k)^2) = \mathbf{E}(X_{k+1}^2) - \mathbf{E}(X_k^2).$$

En sommant cette dernière identité, on a également

$$\sum_{k=n}^{m-1} \mathbf{E}((X_{k+1} - X_k)^2) = \mathbf{E}(X_m^2) - \mathbf{E}(X_n^2). \tag{17}$$

(2) Grâce à la deuxième condition dans (15) et à cette dernière relation, on obtient

$$\sum_{k=1}^{\infty} \mathbf{E}((X_{k+1} - X_k)^2) = \lim_{m\to\infty} \sum_{k=1}^{m} \mathbf{E}((X_{k+1} - X_k)^2) \leq \sup_{m\geq 1} \mathbf{E}(X_m^2) < \infty.$$

En combinant (16), (17) et cette dernière borne, on en déduit

$$\mathbf{E}((X_m - X_n)^2) = \sum_{k=n}^{m-1} \mathbf{E}((X_{k+1} - X_k)^2) \leq \sum_{k=n}^{\infty} \mathbf{E}((X_{k+1} - X_k)^2) \to 0 \tag{18}$$

lorsque $m \geq n \to \infty$. Cela signifie que (X_n) est une suite de Cauchy dans $\mathcal{L}^2$, et il existe donc $X_\infty \in L^2$ telle que $X_n \to X_\infty$ en norme $\mathcal{L}^2$.
(3) Pour $\varepsilon > 0$, on définit

$$A_1 := \{|X_1| \geq \varepsilon\}, \quad A_\ell := \{|X_1| < \varepsilon, \ldots, |X_{\ell-1}| < \varepsilon, |X_\ell| \geq \varepsilon\} \text{ si } \ell \geq 2.$$

En reprenant **(1)**, pour $n \geq \ell$, on calcule

$$\begin{aligned}
\mathbf{E}(X_n^2 \mathbf{1}_{A_\ell}) &= \mathbf{E}((X_n - X_\ell)^2 \mathbf{1}_{A_\ell}) - 2\mathbf{E}((X_n - X_\ell)X_\ell \mathbf{1}_{A_\ell}) + \mathbf{E}(X_\ell^2 \mathbf{1}_{A_\ell}) \\
&= \mathbf{E}((X_n - X_\ell)^2 \mathbf{1}_{A_\ell}) + \mathbf{E}(X_\ell^2 \mathbf{1}_{A_\ell}) \\
&\geq \mathbf{E}(X_\ell^2 \mathbf{1}_{A_\ell}) \geq \varepsilon^2 \mathbf{P}(A_\ell),
\end{aligned}$$

où on a utilisé la première relation de (15) à la seconde ligne et la définition de A_ℓ dans la dernière inégalité. En observant que

$$\mathbf{P}(\max_{1\leq \ell\leq n} |X_\ell| \geq \varepsilon) = \mathbf{P}(A_1 \cup \cdots \cup A_n) = \sum_{\ell=1}^{n} \mathbf{P}(A_\ell),$$

par définition des (A_ℓ) et parce que ces ensembles sont disjoints, on en déduit

$$\mathbf{P}(\max_{1\leq \ell\leq n} |X_\ell| \geq \varepsilon) \leq \sum_{\ell=1}^{n} \frac{1}{\varepsilon^2} \mathbf{E}(X_n^2 \mathbf{1}_{A_\ell}) = \frac{1}{\varepsilon^2} \mathbf{E}(X_n^2 \sum_{\ell=1}^{n} \mathbf{1}_{A_\ell}).$$

Utilisant encore une fois que les (A_ℓ) sont disjoints, on a démontré l'inégalité de Kolmogorov

$$\mathbf{P}(\max_{1\leq \ell\leq n} |X_\ell| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{E}(X_n^2).$$

(4) Exactement comme à l'étape **(3)** (on applique le raisonnement précédent à la suite de var $M_k := X_{p+k} - X_p$, $k \geq 1$), on démontre

$$\mathbf{P}(\max_{p\leq q\leq N} |X_q - X_p| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{E}((X_N - X_p)^2).$$

D'après l'étape **(2)**, pour tout $p \geq 1$, il existe n_p tel que

$$\mathbf{E}((X_m - X_{n_p})^2) \leq \frac{1}{2^p}, \quad \forall\, m \geq n_p.$$

En posant

$$B_p := \{ \sup_{k \geq n_p} |X_k - X_{n_p}| \geq \varepsilon \}, \quad B_p^N := \{ \max_{N \geq k \geq n_p} |X_k - X_{n_p}| \geq \varepsilon \},$$

on a

$$\mathbf{P}(B_p) = \lim_{N \to \infty} \mathbf{P}(B_p^N) \leq \frac{1}{\varepsilon^2} \frac{1}{2^p},$$

de sorte que

$$\sum_{p=1}^{\infty} \mathbf{P}(B_p) \leq \frac{1}{\varepsilon^2}.$$

D'après le lemme de Borel-Cantelli, presque sûrement, un évenement élémentaire $\{\omega\}$ n'est contenu que dans un nombre fini d'ensembles B_p. Soyons plus précis. Il existe $\Omega_\varepsilon \in \mathcal{A}$ tel que $\mathbf{P}(\Omega \backslash \Omega_\varepsilon) = 0$ et $\forall\, \omega \in \Omega_\varepsilon$, $\exists p$ tel que $\omega \notin B_p$, et donc $\exists n_p \geq 1$ tel que

$$\forall\, k, \ell \geq n_p, \quad |X_k(\omega) - X_\ell(\omega)| \leq 2 \sup_{i \geq n_p} |X_i(\omega) - X_{n_p}(\omega)| < 2\varepsilon.$$

On pose

$$\Omega' := \bigcap_{\varepsilon \in \mathbb{Q},\ \varepsilon > 0} \Omega_\varepsilon,$$

de sorte que l'on a encore $\mathbf{P}(\Omega \backslash \Omega') = 0$, et cette fois-ci, pour tout $\omega \in \Omega'$, pour tout $\varepsilon \in \mathbb{Q}_+^*$, il existe $N = N(\omega, \varepsilon) \geq 1$ tel que

$$\forall\, k, \ell \geq N, \quad |X_k(\omega) - X_\ell(\omega)| \leq 2\varepsilon.$$

Cela signifie que $(X_\ell(\omega))$ est une suite de Cauchy dans $\mathbb{R}$ et est donc convergente. □

Application 1. Soit (Y_n) une suite de var iid centrées $\mathcal{L}^2$ et $\alpha > 1/2$, alors

$$\text{ps} \qquad \sum_{k=1}^{\infty} \frac{Y_k}{k^\alpha} \qquad \text{converge.}$$

Application 2. Processus de branchement de Galton-Watson. On se donne $(\zeta_{n,k})_{n,k \in \mathbb{N}}$ une famille de va iid de même loi de probabilité μ supportée par $\mathbb{N}$ et telle que

$$q := \sum_{k=0}^{\infty} k^2 \mu(k) < \infty, \quad m := \sum_{k=0}^{\infty} k \mu(k) > 1.$$

Soient (Y_n) et (X_n) les suites de va définies (par récurrence) par

$$Y_0 := 1, \quad Y_{n+1} := \sum_{k=1}^{Y_n} \zeta_{n,k}, \quad X_n := \frac{Y_n}{m^n}.$$

Alors

$$(X_n) \quad \text{converge ps et dans } \mathcal{L}^2.$$