CHAPITRE 7

VARIABLES ALÉATOIRES INTÉGRABLES
MOMENTS D'UNE V.A.
LOI FORTE des GRANDS NOMBRES

(I) VARIABLES ALÉATOIRES INTÉGRABLES - ESPACE $\mathcal{L}^1$:

Soit Y une v.a. simple définie sur $(\Omega, \mathcal{F}, \mathbb{P})$:

$$Y = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} ,$$

où les valeurs α_i sont 2 à 2 distinctes et $A_i \in \mathcal{F}$

Définition : on appelle intégrale de Y sur Ω (relativement à la mesure de probabilité $\mathbb{P}$), le nombre

$$\int_\Omega Y \, d\mathbb{P} = \sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i).$$

Plus généralement, on appelle intégrale de Y sur l'ensemble $A \in \mathcal{F}$, le nombre

$$\int_A Y \, d\mathbb{P} = \int_\Omega \mathbb{1}_A \cdot Y \, d\mathbb{P} = \sum_{i=1}^{n} \alpha_i \mathbb{P}(A \cap A_i)$$

Définition : Soit X une v.a. à valeurs dans $\mathbb{R}_+$ et $A \in \mathcal{F}$. On appelle intégrale de X sur A, le le nombre (éventuellement $+\infty$) défini par

$$\int_A X \, d\mathbb{P} = \operatorname{Sup} \int_A Y \, d\mathbb{P},$$

où le sup est pris sur l'ensemble de toutes les v.a. simples Y telles que $0 \leq Y \leq X$.

Soit maintenant X une v.a. réelle quelconque. On pose

$X^+ = \sup(X, 0)$ (i.e. $\forall \omega \in \Omega$, $X^+(\omega) = \sup(X(\omega), 0)$)

$X^- = -\inf(X, 0)$ (i.e. $\forall \omega \in \Omega$, $X^-(\omega) = -\inf(X(\omega), 0)$).

Alors X^+ et X^- sont des v.a. <u>positives</u> (au sens large) et : $X = X^+ - X^-$ et $|X| = X^+ + X^-$

<u>Définition</u>: Si les intégrales $\int_\Omega X^+ d\mathbb{P}$ et $\int_\Omega X^- d\mathbb{P}$ sont <u>finies</u> (ce qui équivaut à $\int_\Omega |X| \, d\mathbb{P} < +\infty$) on dit que X est <u>intégrable</u> (ou que $X \in \mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$). L'intégrale de X appelée aussi <u>espérance de X</u> est le nombre réel

$$\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P} := \int_\Omega X^+ d\mathbb{P} - \int_\Omega X^- d\mathbb{P}.$$

De plus pour tout $A \in \mathcal{F}$, on définit :

$$\int_A X \, d\mathbb{P} = \mathbb{E}(X \mathbb{1}_A)$$

Définition : On dit que deux v.a. X et Y sont égale $\mathbb{P}$-p.s. ($\mathbb{P}$ presque surement) et on écrira $X = Y$ p.s si $\mathbb{P}(X = Y) = 1$.

De même une propriété est dite vraie p.s. si elle est vraie avec probabilité 1.

Théorème : Soient X et Y deux v.a. intégrables telles que $X = Y$ p.s. alors

$$\forall A \in \mathcal{F}, \int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$$

Réciproquement : si X et Y sont des v.a. intégrables t.q. $\forall A \in \mathcal{C}$, $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$. Alors $X = Y$ p.s.

Théorème (critère de domination) : Soient X et Y des v.a. telles que $|X| \leq Y$ p.s. (i.e $\mathbb{P}(|X| \leq Y) = 1$)
Si $Y \in \mathcal{L}^1$ alors $X \in \mathcal{L}^1$

2 Les Propriétés fondamentals de l'intégrales

1) Positivité : si X est une v.a. ≥ 0 sur $A \in \mathcal{F}$, alors $\int_A X \, d\mathbb{P} \geq 0$.

(on notera que pour $X \geq 0$ sur A, l'intégrale $\int_A X \, d\mathbb{P}$ existe toujours, éventuellement elle vaut $+\infty$)

2) <u>Linéarité</u> : si X et Y sont dans $\mathcal{L}^1$, alors $\forall \alpha, \beta \in \mathbb{R}$, $\alpha X + \beta Y \in \mathcal{L}^1$ et on a :

$$\forall A \in \mathcal{F}, \quad \int_A (\alpha X + \beta Y) \, d\mathbb{P} = \alpha \int_A X \, d\mathbb{P} + \beta \int_A Y \, d\mathbb{P}.$$

3) <u>Sigma-additivité</u> : Soit X une v.a. positive ou une v.a. appartenant à $\mathcal{L}^1$. Soient $(A_n) \in \mathcal{F}$ une suite d'événements <u>2 à 2 disjoints</u>. Alors

$$\int_{\bigcup_n A_n} X \, d\mathbb{P} = \sum_n \int_{A_n} X \, d\mathbb{P}.$$

4) <u>Monotonie</u> : Soient X_1, X_2, X des v.a. positives ou des v.a. appartenant toutes les trois à $\mathcal{L}^1$.

Si $X_1 \leq X \leq X_2$ p.s., alors

$$\int_\Omega X_1 \, d\mathbb{P} \leq \int_\Omega X \, d\mathbb{P} \leq \int_\Omega X_2 \, d\mathbb{P}.$$

5) <u>Inégalité du module</u> : Si $X \in \mathcal{L}^1$, alors

$$\left| \int_\Omega X \, d\mathbb{P} \right| \leq \int_\Omega |X| \, d\mathbb{P}$$

6) <u>Théorème de convergence monotone</u> : Soient (X_n) une suite de v.a. ≥ 0 et X une v.a. ≥ 0. On suppose $X_n \nearrow X$ p.s. Alors

$$\int_\Omega X_n \, d\mathbb{P} \nearrow \int_\Omega X \, d\mathbb{P} \text{ (éventuellement } = +\infty)$$

($\nearrow$ signifie converge en croissant)

7) <u>Théorème de convergence dominée</u> : Soient (X_n), $n \in \mathbb{N}$, X et Y des v.a. On suppose

a) $Y \geq 0$ et $Y \in \mathcal{L}^1$

b) $\lim_n X_n = X$ p.s.

c) $\forall n \in \mathbb{N}$, $|X_n| \leq Y$ (domination)

<u>Alors</u> $X \in \mathcal{L}^1$ <u>et</u> :

1) $\lim_{n \to \infty} \int_\Omega |X_n - X| \, d\mathbb{P} = 0$

2) $\lim_{n \to \infty} \int_\Omega X_n \, d\mathbb{P} = \int_\Omega X \, d\mathbb{P}$

(2) est une conséquence de 1))

8) <u>Théorème de Lebesgue sur les séries</u> : Si (X_n) est une suite de v.a. telles que

$$\sum_n \int_\Omega |X_n| \, d\mathbb{P} < +\infty$$

alors,

1) $\sum_n |X_n| < +\infty$ p.s.

2) La v.a. $X = \sum_n X_n$ (définie p.s.) est intégrable et satisfait

$$\int_\Omega \Big(\sum_n X_n\Big) d\mathbb{P} = \sum_n \int_\Omega X_n \, d\mathbb{P}.$$

(précision $X = \sum_n X_n(\omega)$ si cette série converge et $X(\omega) = 0$ sinon)

9) <u>Le lemme de Fatou</u> : Soit (X_n) une suite de v.a. ≥ 0. Alors

$$\int_\Omega (\liminf X_n) d\mathbb{P} \leq \liminf_{n \to +\infty} \int_\Omega X_n \, d\mathbb{P}$$

(les deux membres sont éventuellement $+\infty$).

10) <u>L'espace L^1</u> :

<u>Définition</u> : On appelle $L^1 = L^1(\Omega, \mathcal{F}, \mathbb{P})$ l'espace quotient de $\mathcal{L}^1$ par la relation d'équivalence « égalité p.s. »

Autrement dit si $X = Y$ p.s., on considère que X et Y coïncident dans L^1 i.e. représentent la même v.a.

Norme sur L^1 : On définit une norme en posant pour tout $X \in L^1$:

$$\|X\|_1 = \int_\Omega |X| \, d\mathbb{P}$$

Théorème : Avec cette norme, L^1 est un espace de Banach (i.e. toute suite de Cauchy au sens de cette norme, est convergente).

(II) CALCUL EFFECTIF de l'ESPÉRANCE - FORMULE DU TRANSFERT.

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_1)$ une v.a. et μ_X sa loi de probabilité (c'est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}_1)$).

On considère l'espace $\mathcal{L}^1(\mu_X) = \mathcal{L}^1(\mathbb{R}, \mathcal{B}_1, \mu_X)$ l'espace des fonctions boréliennes $\varphi : \mathbb{R} \to \mathbb{R}$ qui sont intégrables relativement à la mesure μ_X. On note alors $\int_\mathbb{R} \varphi \, d\mu_X := \int_\mathbb{R} \varphi(x) \, d\mu_X(x)$ l'intégrale de φ

Théorème (du transfert) Soit X une v.a. et $\varphi : (\mathbb{R}, \mathcal{B}_1) \to (\mathbb{R}, \mathcal{B}_1)$ une fonction borélienne. Alors $\varphi \circ X \in \mathcal{L}^1(\mathbb{P})$ si et seulement si $\varphi \in \mathcal{L}^1(\mu_X)$ et dans ce cas, on a : $\int_\Omega \varphi \circ X \, d\mathbb{P} = \int_\mathbb{R} \varphi(x) \, d\mu_X(x)$

dém^n : Soit $B \in \mathcal{B}_1$. Démontrons le théorème pour la fonction $\varphi = \mathbb{1}_B$. Notons que $\mathbb{1}_B \circ X = \mathbb{1}_{[X \in B]}$

On a :

$$\mu_X(B) = \mathbb{P}(X \in B)$$

mais aussi $\mu_X(B) = \int_{\mathbb{R}} \mathbb{1}_B(x) d\mu_X(x)$ et

$$\mathbb{P}(X \in B) = \int_{\Omega} \mathbb{1}_{[X \in B]} d\mathbb{P} = \int_{\Omega} \mathbb{1}_B \circ X \, d\mathbb{P}$$

Donc $\int_{\mathbb{R}} \mathbb{1}_B(x) \, d\mu_X(x) = \int_{\Omega} \mathbb{1}_B \circ X \, d\mathbb{P}$. Ainsi $\mathbb{1}_B$ est μ_X intégrable $\Longleftrightarrow$ $\mathbb{1}_B \circ X$ est $\mathbb{P}$-intégrable. D'où le théorème si $\varphi = \mathbb{1}_B$.

Par linéarité de l'intégrale, le théorème est encore vrai lorsque φ est une combinaison linéaire d'indicatrices d'éléments de $\mathcal{B}_1$.

Soit alors $\varphi \geq 0$ dans $\mathcal{L}^1(\mu_X)$. Il existe une suite croissante (φ_n) de fonctions simples telle que $\varphi_n \nearrow \varphi$ (voir fin du § sur la mesurabilité du chapitre 6). D'après l'étape précédente,

$$\forall n \in \mathbb{N}, \quad \int_{\mathbb{R}} \varphi_n(x) \, d\mu_X(x) = \int_{\Omega} \varphi_n \circ X \, d\mathbb{P}$$

Par le théorème de convergence monotone, on passe à la limite quand $n \to +\infty$ et on obtient :

$$\int_{\mathbb{R}} \varphi(x)\, d\mu_X(x) = \int_{\Omega} \varphi \circ X \, d\mathbb{P} \quad (*)$$

et le théorème en résulte si $\varphi \geq 0$.

Si φ est de signe quelconque, on utilise la décomposition $\varphi = \varphi^+ - \varphi^-$ et le fait qu'alors $\varphi \circ X = \varphi^+ \circ X - \varphi^- \circ X$. On applique ensuite $(*)$ pour en déduire le théorème dans le cas général.

<u>CQFD</u>.

<u>Corollaire 1</u> : $X \in \mathcal{L}^1 \iff \int_{\mathbb{R}} |x|\, d\mu_X(x) < +\infty$ et dans ce cas :

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \, d\mu_X(x)$$

En particulier si X a une densité f, alors

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx .$$

<u>dém^n</u> : appliquer le théorème avec $\varphi(x) = x$.
Le cas particulier en résulte aussitôt puisque $d\mu_X(x) = f(x)dx$.

__Remarque :__ L'importance du théorème du transfert provient du fait que pour calculer l'espérance de la v.a. $Y = \varphi(X)$, on n'a pas besoin de déterminer la loi μ_Y de Y.

__Définition :__ On dit que la v.a. X a un moment d'ordre n si $X^n \in \mathcal{L}^1$ (i.e. $\int_\Omega |X|^n d\mathbb{P} < +\infty$).

Dans ce cas le moment d'ordre n est le nombre

$$\mathbb{E}(X^n).$$

__Corollaire 2 :__ X a un moment d'ordre n si et seulement si $\int_{\mathbb{R}} |x|^n d\mu_X(x) < +\infty$. Dans ce cas :

$$\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n d\mu_X(x).$$

En particulier si X a une densité f, on a

$$\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n f(x) dx.$$

__dém^n :__ résulte du théorème avec $\varphi(x) = x^n$.

__Théorème du transfert pour les vecteurs aléatoires__

Soit $X = (X_1, \dots, X_d) : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}_d)$ un vecteur aléatoire et

$\varphi : (\mathbb{R}^d, \mathcal{B}_d) \longrightarrow (\mathbb{R}, \mathcal{B}_1)$ une fonction borélienne à valeurs réelles.

Soit $Y = \varphi \circ X = \varphi(X) = \varphi(X_1, \dots, X_d)$; c'est une variable aléatoire.

<u>Théorème</u> : $Y \in \mathcal{L}^1 \iff \varphi \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_d, \mu_X)$

et dans ce cas :

$$\mathbb{E}(Y) = \int_{\mathbb{R}^d} \varphi(x_1, \dots, x_d)\, d\mu_X(x_1, \dots, x_n)$$

En particulier si X a une densité $f(x_1, \dots, x_n)$,

$$\mathbb{E}(Y) = \int_{\mathbb{R}^d} \varphi(x_1, \dots, x_n)\, f(x_1, \dots, x_n)\, dx_1 \dots dx_n .$$

<u>dém^n</u> : analogue au cas $d = 1$.

<u>Application</u> : Espérance du produit de n v.a. indépendantes.

<u>Corollaire 3</u> : Soient $X_1, \dots, X_n$ des v.a. de $\mathcal{L}^1$ qu'on suppose <u>indépendantes</u>. Alors la v.a. $Y = \prod_{i=1}^{n} X_i$ appartient à $\mathcal{L}^1$ et on a

$$\mathbb{E}(Y) = \prod_{i=1}^{n} \mathbb{E}(X_i)$$

<u>dém^n</u> : supposons $n = 2$ pour alléger l'écriture

Comme X_1 et X_2 sont indépendantes, la loi du vecteur $X = (X_1, X_2)$ est $\mu_X = \mu_{X_1} \otimes \mu_{X_2}$. D'après le théorème du transfert appliqué à la fonction $\varphi(x_1, x_2) = x_1 x_2$, on a $Y \in \mathcal{L}^1 \iff \varphi$ intégrable pour la mesure $\mu_{X_1} \otimes \mu_{X_2}$ i.e. :

$$\iint_{\mathbb{R}^2} |x_1 x_2| \, d\mu_{X_1}(x_1) d\mu_{X_2}(x_2) < +\infty .$$

Mais cette condition est bien vérifié puisque l'intégrale double précédente vaut

$$\left(\int_{\mathbb{R}} |x_1| \, d\mu_{X_1}(x_1) \right) \left(\int_{\mathbb{R}} |x_2| \, d\mu_{X_2}(x) \right)$$

donc est finie par hypothèse (X_1 et $X_2 \in \mathcal{L}^1$).

Le théorème du transfert (p. 11) donne alors

$$\mathbb{E}(Y) = \iint_{\mathbb{R}^2} x_1 x_2 \, d\mu_{X_1}(x_1) d\mu_{X_2}(x_2)$$

$$= \left(\int_{\mathbb{R}} x_1 \, d\mu_{X_1}(x_1) \right) \left(\int_{\mathbb{R}} x_2 \, d\mu_{X_2}(x_2) \right) \quad \text{(Fubini)}$$

$$= \mathbb{E}(X_1) \mathbb{E}(X_2) \quad \underline{\text{CQFD}}.$$

III ESPACE $\mathcal{L}^2$, VARIANCE des V.A.

Soit $\mathcal{L}^2$ l'espace des v.a. ayant un moment d'ordre 2 (i.e. les v.a. X telles que $\mathbb{E}(X^2)<+\infty$).

Pour $X \in \mathcal{L}^2$, on pose

$$\|X\|_2 = \left(\mathbb{E}(X^2)\right)^{1/2} \left(= \left(\int_\Omega |X|^2 d\mathbb{P}\right)^{1/2}\right)$$

Si on considère comme égales deux v.a. égales p.s. c'est à dire si $X \in L^2$, $\|\cdot\|_2$ est une norme sur L^2

Cette norme dérive du produit scalaire

$$(X, Y) \longmapsto \mathbb{E}(XY)$$

Muni de ce produit scalaire L^2 est un espace de Hilbert. Rappelons qu'on a les propriétés suivantes:

<u>Inégalité de Cauchy Schwarz</u> :

$$\forall X, Y \in L^2, \quad \mathbb{E}(|XY|) \leq \mathbb{E}(X^2)^{1/2}\, \mathbb{E}(Y^2)^{1/2} \; (*)$$

<u>Inégalité de Minkowski</u>:

$$\forall X, Y \in L^2, \quad \|X+Y\|_2 \leq \|X\|_2 + \|Y\|_2$$

<u>Théorème de Pythagore</u>: si X et Y sont telles que $\mathbb{E}(XY) = 0$ (on dit que X et Y sont <u>orthogonales</u>), alors $\|X+Y\|_2^2 = \|X\|_2^2 + \|Y\|_2^2$.

(*) Noter que ceci implique que si $X \in \mathcal{L}^2$ alors $X \in \mathcal{L}^1$

<u>Définition</u> pour toute v.a. $X \in \mathcal{L}^2$, on appelle <u>variance</u> de X le nombre

$$\mathrm{Var}\, X = \mathbb{E}\left((X - \mathbb{E}(X))^2 \right) = \| X - \mathbb{E}(X) \|_2^2$$

(on notera que c'est le carré de la norme de la variable centrée $X - \mathbb{E}(X)$).

On appelle alors <u>écart-type</u> de X, le nombre

$$\sigma_X = (\mathrm{Var}\, X)^{1/2} \quad (= \| X - \mathbb{E}(X) \|_2 \ : \text{c'est une norme}).$$

> <u>Proposition</u> (calcul de Var X) : Si $X \in \mathcal{L}^2$,
> $$\mathrm{Var}\, X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

<u>dém</u>^{n} : développer le carré $(X - \mathbb{E}(X))^2$ et utiliser la linéarité de l'espérance.

> <u>Théorème</u> : Soient $X_1, \dots, X_n \in \mathcal{L}^2$ telles que les v.a. centrées $\tilde{X}_i = X_i - \mathbb{E}(X)$ soient <u>2 à 2 orthogonales</u>. Alors
> $$\mathrm{Var}\left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \mathrm{Var}\, X_i$$

<u>dém</u>^{n} : c'est le théorème de Pythagore appliqué aux v.a. $\tilde{X}_i$ <u>CQFD</u>.

> <u>Corollaire</u> : Si $X_1, \dots, X_n \in \mathcal{L}^2$ sont <u>indépendantes</u>. Alors $\mathrm{Var}\left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \mathrm{Var}\, X_i$

Exemples et exercices :

1) Si $X \in \mathcal{L}^2$, $\forall a, b \in \mathbb{R}$, $\mathrm{Var}(aX+b) = a^2 \,\mathrm{Var}\, X$

2) Si $X \in \mathcal{L}^2$ est telle que $\mathrm{Var}\, X = 0$, alors $X = \mathbb{E}(X)$ p.s.

3) Si $X \in \mathcal{L}^2$ et $Y = \dfrac{X - \mathbb{E}(X)}{\sigma_X}$ la v.a. centrée réduite associée, $\mathbb{E}(Y) = 0$ et $\mathrm{Var}\, Y = 1$.

4) Si X est de loi $N(m, \sigma^2)$, montrer que $X \in \mathcal{L}^2$, $\mathbb{E}(X) = m$ et $\mathrm{Var}\, X = \sigma^2$.

5) Calculer l'espérance et la variance d'une loi exponentielle de paramètre λ.

6) Soit X une v.a. de $\mathcal{L}^1$, montrer que pour tout $a > 0$, on a :

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

(inégalité de Markov). En déduire que si $X \in \mathcal{L}^2$, pour tout $a > 0$, on a

$$\mathbb{P}\left(\left| \frac{X - \mathbb{E}(X)}{\sigma_X} \right| \geq a \right) \leq \frac{1}{a^2}$$

(inégalité de Bienaymé-Tchebychev : B.T.)

7) Déduire de l'inégalité de B.T. le résultat suivant (loi faible des grands nombres) :
Soit (X_n) une suite de v.a. de $\mathcal{L}^2$ de même

espérance m et de même variance σ^2. On suppose que les v.a. centrées $\tilde{X}_n = X_n - \mathbb{E}(X_n)$ sont 2 à 2 orthogonales. Alors

$$\forall \varepsilon > 0, \lim_{n \to +\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - m \right| \geq \varepsilon \right) = 0$$

(Noter que si les X_n sont indépendantes, le résultat s'applique)

(indication: utiliser l'inégalité B.T. et la même méthode que dans le cas discret étudié au début du cours).

IV LA LOI FORTE des GRANDS NOMBRES

1) La convergence presque sûre : Soient (X_n) une suite de v.a. et X une v.a. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition : On dit que X_n converge presque-sûrement vers X si

$$\mathbb{P}\left(\{ \omega \in \Omega ; \lim_{n \to \infty} X_n(\omega) = X(\omega) \} \right) = 1$$

On écrit alors $\lim_{n \to \infty} X_n = X$ p.s.

Théorème : Si $\lim_{n \to \infty} X_n = X$ p.s alors on a aussi $\lim_{n \to \infty} X_n = X$ en probabilité (i.e. $\forall \delta > 0$, $\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \delta) = 0$)

(Autrement dit la convergence p.s. implique la convergence en probabilité.)

<u>dém^n</u>: Soit Ω_1 l'ensemble de probabilité 1 tel que $\forall \omega \in \Omega_1$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Alors :

$$\forall \omega \in \Omega_1, \forall \delta > 0, \exists n, \forall k \geq n, |X_k(\omega) - X(\omega)| \leq \delta$$

ce qui équivaut à :

$$\forall \omega \in \Omega_1, \forall \delta > 0, \exists n, \omega \in \bigcap_{k\geq n} [|X_k - X| \leq \delta]$$

$$\iff \forall \omega \in \Omega_1, \forall \delta > 0, \omega \in \bigcup_{n\geq 0} \left(\bigcap_{k \geq n} [|X_k - X| \leq \delta] \right)$$

Donc :

$$\forall \delta > 0, \Omega_1 \subset \liminf_n A_n \text{ où } A_n = [|X_n - X| \leq \delta]$$

En passant au complémentaire, on obtient

$$\limsup_n A_n^c \subset \Omega_1^c,$$

D'où $\mathbb{P}(\limsup_n A_n^c) = 0$. Donc

$$\mathbb{P}\Big(\bigcup_{k\geq n} A_k^c\Big) \longrightarrow 0 \text{ quand } n \to +\infty.$$

A fortiori, on a aussi $\mathbb{P}(A_n^c) = \mathbb{P}(|X_n - X| > \delta) \xrightarrow[n\to\infty]{} 0$

c'est à dire $X_n \to X$ en probabilité <u>CQFD</u>.

2) Un critère suffisant de convergence p.s.

Théorème : Soit (X_k) une suite de v.a. telles que

$$\forall \varepsilon > 0, \quad \sum_{k=1}^{+\infty} \mathbb{P}(|X_k| > \varepsilon) < +\infty$$

Alors : $\lim\limits_{k\to\infty} X_k = 0$ p.s.

On a besoin du lemme fondamental suivant d'une grande utilité dans toutes les questions concernant la convergence presque sûre.

Lemme (de Borel-Cantelli) : Soit (A_n) une suite d'événements tels que $\sum_n \mathbb{P}(A_n) < +\infty$. Alors

$$\mathbb{P}(\limsup_n A_n) = 0$$

(rappel : $\limsup_n A_n = \bigcap_n \bigcup_{k\geq n} A_k$: voir feuilles d'exercices)

dém^{n} du Lemme :

$$\mathbb{P}(\limsup A_n) = \mathbb{P}\left(\bigcap_n B_n\right) \quad \text{où } B_n = \bigcup_{k\geq n} A_k$$

forment une suite décroissante d'événements. Comme $\mathbb{P}$ est continue par limite décroissante, on a :

$$\mathbb{P}(\limsup_n A_n) = \lim_{n\to\infty} \mathbb{P}(B_n) \leq \lim_{n\to\infty} \sum_{k=n}^{+\infty} \mathbb{P}(A_k) = 0$$

(reste d'une série convergente) CQFD.

<u>dém^n du théorème</u> : D'après le lemme de B.C. :

$\forall \varepsilon > 0$, $\mathbb{P}\left(\limsup_k [|X_k| > \varepsilon]\right) = 0$

En passant au complémentaire, on obtient

$\forall \varepsilon > 0$, $\mathbb{P}\left(\liminf_k [|X_k| \leq \varepsilon]\right) = 1$.

Ainsi pour tout entier $n \geq 1$, l'événement

$$A_n = \liminf_k \left[|X_k| \leq \frac{1}{n}\right]$$

est de probabilité 1. Donc $\mathbb{P}\left(\bigcap_n A_n\right) = 1$

Mais si $\omega \in \bigcap_n A_n$, alors pour tout $n \in \mathbb{N}^*$, $|X_k(\omega)| \leq \frac{1}{n}$ pour tout k assez grand ce qui prouve que $\lim_{k \to \infty} X_k(\omega) = 0$, <u>CQFD</u>.

> <u>Corollaire</u> : Si pour tout $k \geq 1$, $X_k \in \mathcal{L}^p$ $(1 \leq p < \infty)$ est telle que $\sum_{k=1}^{+\infty} \mathbb{E}(|X_k|^p) < +\infty$. Alors
>
> $X_k \longrightarrow 0$ p.s.

<u>dém^n</u> : $\mathbb{P}(|X_k| > \varepsilon) = \mathbb{P}(|X_k|^p > \varepsilon^p) \leq \dfrac{\mathbb{E}(|X_k|^p)}{\varepsilon^p}$

d'après l'inégalité de Markov. Donc la série $\sum_{k=1}^{\infty} \mathbb{P}(|X_k| > \varepsilon)$ converge et le théorème s'applique <u>CQFD</u>.

3) <u>La loi forte des grands nombres</u>

<u>Définition</u> : Si X et $Y \in \mathcal{L}^2$, on appelle <u>covariance</u> des v.a. X et Y le nombre

$$\operatorname{cov}(X,Y) = \mathbb{E}\big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big)$$

(i.e. le produit scalaire des v.a. <u>centrées</u>).

<u>Théorème</u> (<u>loi forte des grands nombres</u>) Soit (X_k) une suite de v.a. de $\mathcal{L}^2$, de même espérance m et de même variance σ^2. On suppose de plus que les v.a. X_k sont 2 à 2 de covariance nulle. Alors

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow m \quad \text{p.s.} \quad (\text{si } n \to +\infty).$$

On appelle ce résultat loi forte des grands nombres car on connaissait déjà ce résultat pour la convergence en probabilité. La convergence p.s. est plus forte que la convergence en probabilité d'où la terminologie.

<u>dém^n</u> : Si on remplace les X_k par les v.a. centrées $X_k - m$, on se ramène à démontrer le théorème dans le cas $m = 0$, ce que nous supposerons :

Posons $S_n = X_1 + \dots + X_n$. On a :

$$E\left(\left(\frac{S_n}{n}\right)^2\right) \overset{(*)}{=} Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var\, S_n = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0$$

Ce qui montre que $\frac{S_n}{n} \xrightarrow[(n\to\infty)]{} 0$ dans L^2 mais ceci ne permet pas encore de conclure. Par contre si on pose

$$Y_n = \frac{S_{n^2}}{n^2},$$

l'égalité précédente $(*)$ montre que $\sum_{n=1}^{\infty} E(Y_n^2) < +\infty$.

Donc $\frac{S_{n^2}}{n^2} \xrightarrow[(n\to\infty)]{} 0$ p.s.

d'après le corollaire p.19. Pour montrer que $\frac{S_n}{n} \to 0$ procédons comme suit :

Pour tout entier n, il existe un entier $m (= m(n)) \geq 1$ tel que $m^2 \leq n < (m+1)^2$. Ainsi

$$\left|\frac{S_n}{n}\right| = \left|\frac{S_{m^2} + X_{m^2+1} + \cdots + X_n}{n}\right|$$

$$\leq \left|\frac{S_{m^2}}{n}\right| + \left|\frac{X_{m^2+1} + X_{m^2+2} + \cdots + X_n}{n}\right|$$

$$\leq \left|\frac{S_{m^2}}{m^2}\right| + \left|\frac{1}{n}\left(X_{m^2+1} + \cdots + X_n\right)\right|$$

Comme $\left|\frac{S_{m^2}}{m^2}\right| \to 0$ p.s $(n \to +\infty)$, il reste à prouver que

$$\left|\frac{1}{n}\left(X_{m^2+1} + \cdots + X_n\right)\right| \xrightarrow[(n\to\infty)]{} 0 \text{ p.s.}$$

On utilise encore une fois le corollaire p. 19 en notant que

$$\mathbb{E}\left(\left(\frac{X_{m^2+1}+\cdots+X_n}{n}\right)^2\right)=\frac{1}{n^2}\operatorname{Var}\left(X_{m^2+1}+\cdots+X_n\right)=$$

$$=\frac{1}{n^2}\sum_{k=m^2+1}^{n}\operatorname{Var}X_k$$

$$=\frac{(n-m^2)\sigma^2}{n^2}\leqslant\frac{(m+1)^2-m^2}{n^2}=\frac{2m+1}{n^2}\leqslant\frac{2\sqrt{n}+1}{n^2}\sim 2n^{-3/2}$$

qui est le terme général d'une série convergente.

Donc $\frac{1}{n}\left(X_{m^2+1}+\cdots+X_n\right)\longrightarrow 0$ p.s. (si $n\to+\infty$) CQFD.

Cas particulier : A fortiori le théorème s'applique quand les v.a. (X_k) sont indépendantes car alors elles sont de covariance nulle.

Remarque : Il existe des résultats portant aussi le nom de loi forte des grands nombres mais beaucoup plus difficiles. Par exemple :

Théorème (loi forte des grands nombres de Kolmogorov)

Soit (X_k) une suite de v.a. indépendantes et de même loi. On suppose $X_1\in\mathcal{L}^1$ (i.e. $\mathbb{E}(|X_1|)<+\infty$).

Alors :

$$\frac{1}{n}\sum_{k=1}^{n}X_k\longrightarrow\mathbb{E}(X_1)\ \text{p.s.}\ (\text{si } n\to+\infty)$$

(résultat admis).