Master 1 de Mathématiques Année 2012-2013

Probabilités 1 cours de M. Gallardo

Chapitre 5 : Variables Aléatoires générales,
Loi forte des grands nombres
Notions de convergence des variables aléatoires.

Introduction : Jusqu'à présent nous n'avons considéré que des variables aléatoires discrètes ou ayant une densité de probabilité car nous n'avons travaillé que sur des problèmes liés aux lois de probabilité. Pour des problèmes plus généraux liés aux variables aléatoires elles-mêmes par exemple pour des questions de convergence presque sûre, nous avons besoin d'un formalisme plus perfectionné qui utilise la théorie de l'intégration sur l'espace probabilisé $(\Omega, \mathcal{E}, \mathbb{P})$ sur lequel sont définies les variables aléatoires.

I Variables aléatoires intégrables - Espace $\mathcal{L}^1$

Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace probabilisé

Définition 1 : On appelle variable aléatoire (v.a.) simple (on dit aussi étagée) toute v.a. de la forme

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

où les $\alpha_i (\in \mathbb{R})$ sont des constantes et les A_i des événements appartenant à la tribu $\mathcal{E}$.

Si $A \in \mathcal{E}$, $\mathbb{1}_A$ est la v.a. de Bernoulli définie par :

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in A^c \end{cases}$$

Une v.a. simple est donc une combinaison linéaire (finie) de v.a. de Bernoulli

Définition 2 Soit X une v.a. simple $X = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$

On appelle intégrale de X sur Ω (pour la probabilité $\mathbb{P}$) ou espérance de X (pour $\mathbb{P}$), le nombre :

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

Plus généralement l'intégrale de X sur l'ensemble $A \in \mathcal{E}$ est le nombre :

$$\mathbb{E}(\mathbb{1}_A X) = \int_A X \, d\mathbb{P} = \sum_{i=1}^{n} \alpha_i \mathbb{P}(A \cap A_i)$$

Définition 3 Soit X une v.a. à valeurs dans $\mathbb{R}_+$ et $A \in \mathcal{E}$. On appelle intégrale de X sur A, le nombre (éventuellement $+\infty$) défini par :

$$\int_A X \, d\mathbb{P} = \sup_{\substack{Y \text{ simple :} \\ 0 \le Y \le X}} \int_A Y \, d\mathbb{P}$$

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

Soit maintenant $X : \Omega \to \mathbb{R}$ une v.a. à valeurs réelles. On pose :

$X^+ = \mathrm{Max}(X,0)$ (i.e. $\forall \omega \in \Omega$, $X^+(\omega) = \mathrm{Max}(X(\omega),0)$)
et $X^- = -\mathrm{Min}(X,0)$ (i.e. $\forall \omega \in \Omega$, $X^-(\omega) = -\mathrm{Min}(X(\omega),0)$).
On a alors : $X = X^+ - X^-$ et $|X| = X^+ + X^-$

<u>Définition 4</u> : Soit $X : \Omega \to \mathbb{R}$ une v.a. à valeurs réelles telle que $\int_\Omega X^+ d\mathbb{P} < +\infty$ et $\int_\Omega X^- d\mathbb{P} < +\infty$ (ce qui équivaut à $\int_\Omega |X| d\mathbb{P} < +\infty$).
On dit que <u>X est intégrable</u> et <u>l'intégrale de X appelée aussi espérance de X</u> est le nombre

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

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

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

On note $\mathcal{L}^1(\Omega, \mathcal{C}, \mathbb{P}) := \mathcal{L}^1$ (lorsqu'il n'y a pas d'ambiguïté) l'ensemble des v.a. intégrables.

<u>Variables aléatoires égales presque-sûrement</u>

Deux v.a. X et Y sont dites <u>égales presque sûrement</u> s'il existe $A \in \mathcal{C}$ avec $\mathbb{P}(A) = 0$ tel que :

$$\forall \omega \in \Omega \setminus A \,(= A^c), \ X(\omega) = Y(\omega)$$

On écrit alors $X = Y$ $\mathbb{P}$-p.s. ou plus simplement $X = Y$ p.s.

On a de même l'inégalité presque sûre :

$X \leq Y$ $\mathbb{P}$-p.s. s'il existe $A \in \mathcal{C}$ avec $\mathbb{P}(A) = 0$ tel que $\forall \omega \in \Omega \setminus A$, $X(\omega) \leq Y(\omega)$.

On notera que l'égalité p.s. (resp. l'inégalité p.s) est transitive i.e.

$$X = Y \text{ p.s et } Y = Z \text{ p.s} \Rightarrow X = Z \text{ p.s}$$

(resp $X \leq Y$ p.s. et $Y \leq Z$ p.s $\Rightarrow X \leq Z$ p.s).

<u>Exercice</u> : si $X = Y$ p.s et si $X \in \mathcal{L}^1$ alors $Y \in \mathcal{L}^1$ et $\forall B \in \mathcal{C}$, $\mathbb{E}(\mathbb{1}_B \cdot X) = \mathbb{E}(\mathbb{1}_B \cdot Y)$.

<u>Propriétés de l'espérance</u>

(Nous listons sans démonstration les principales propriétés de l'espérance qui ne sont en fait qu'une réécriture des propriétés de l'intégrale vues en L3 dans le cours d'intégration).

1) (<u>linéarité</u>) : Si X et $Y \in \mathcal{L}^1$, alors $\forall \alpha, \beta \in \mathbb{R}$, $\alpha X + \beta Y \in \mathcal{L}^1$ et on a

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$$

2) (<u>σ-additivité</u>) : Soit $X \in \mathcal{L}^1$ et soient $(A_n) \in \mathcal{C}$ une suite d'événements 2 à 2 disjoints. Alors

$$\mathbb{E}(\mathbb{1}_{\cup A_n} \cdot X) = \sum_n \mathbb{E}(\mathbb{1}_{A_n} \cdot X)$$

3) <u>Monotonie</u> : Soient X_1, X_2, X des v.a. de $\mathcal{L}^1$

telles que $X_1 \leq X \leq X_2$ p.s. alors
$\mathbb{E}(X_1) \leq \mathbb{E}(X) \leq \mathbb{E}(X_2)$

4) (Inégalité du module) Si $X \in \mathcal{L}^1$, alors
$|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

5) (Thm de convergence monotone) Soient (X_n) une suite de v.a. positives et X une v.a. positive. On suppose que $X_n \nearrow X$ p.s. Alors

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

(pour une v.a. positive son intégrale vaut éventuellement $+\infty$. On parle alors d'intégrale mais pas d'espérance)

6) (Thm de convergence dominée) 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) $\lim_{n\to\infty} X_n = X$ p.s

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

$X \in \mathcal{L}^1$ et on a:

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|) = 0$$

et $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$

II Calcul Effectif de l'espérance : Formule du transfert.

Soit $X : (\Omega, \mathcal{C}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_1)$ une v.a.

($\mathcal{B}_1$ désigne la tribu de Borel de $\mathbb{R}$)

Définition : On appelle loi de la v.a. X, la mesure image μ_X de $\mathbb{P}$ par l'application X sur $\mathcal{B}_1$. Autrement dit :

$$\forall B \in \mathcal{B}_1, \ \mu_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

(Ne pas oublier qu'une v.a. $X : \Omega \to \mathbb{R}$ est une application mesurable i.e. telle que $\forall B \in \mathcal{B}_1$, $X^{-1}(B) := [X \in B] \in \mathcal{C}$).

Cas particulier :

1) Si la v.a. X est discrète : $X(\Omega) = \{x_n \, ; \, n \in \mathbb{N}\}$ et $p_n = \mathbb{P}(X = x_n)$, la mesure μ_X est donnée par

$$\mu_X = \sum_n p_n \delta_{x_n}$$

(δ_{x_n} désigne la mesure de Dirac au point x_n).

2) Si la v.a. X a une densité f (chapitre 4) on a :

$$\forall B \in \mathcal{B}_1, \ \mu_X(B) = \int_B f(x)\,dx$$

Notation : On note $\mathcal{L}^1(\mu_X) = \mathcal{L}^1(\mathbb{R}, \mathcal{B}_1, \mu_X)$ l'espace des fonctions boréliennes $\varphi : \mathbb{R} \to \mathbb{R}$ qui sont intégrables pour 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 mesurable. Alors $\varphi \circ X \in \mathcal{L}^1(\mathbb{P})$ si et seulement si $\varphi \in \mathcal{L}^1(\mu_X)$ et on a alors :

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

(*Remarque* la v.a. $\varphi \circ X$ est notée aussi $\varphi(X)$)

En particulier si $\varphi : x \mapsto x$ est l'identité, on a

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

démonstration : Dans le cours oral.

Corollaire : Soit $n \in \mathbb{N}^*$. La v.a. X a un moment d'ordre n (i.e. $X^n \in \mathcal{L}^1(\mathbb{P})$) si et seulement si $\int_{\mathbb{R}} |x|^n d\mu_X(x) < +\infty$ et on a alors :

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

Remarque 1 : Si $X : \Omega \to \mathbb{R}^d$ est un vecteur aléatoire i.e. une application mesurable de $(\Omega, \mathcal{C}, \mathbb{P})$ dans $(\mathbb{R}^d, \mathcal{B}_d)$ où $\mathcal{B}_d$ est la tribu de

Borel de $\mathbb{R}^d$, la loi de X est la mesure image μ_X de $\mathbb{P}$ par X sur $(\mathbb{R}^d, \mathcal{B}_d)$. Lorsque $B = A_1 \times A_2 \times \dots \times A_d$ est un pavé de $\mathbb{R}^d$ (où $A_i \in \mathcal{B}_1 \ \forall i = 1, \dots, d$),

$$\mu_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n)$$

où $X_1, X_2, \dots, X_d$ sont les coordonnées de X.

Comme dans le cas déjà vu aux chapitres 4 (et 3) les v.a. $X_1, X_2, \dots, X_d$ sont indépendantes si et seulement si

$$\mu_X = \mu_{X_1} \otimes \mu_{X_2} \otimes \dots \otimes \mu_{X_d}$$

(i.e. la loi du vecteur X est le produit tensoriel des lois marginales) Nous omettons les détails.

On peut alors énoncer un théorème du transfert pour les vecteurs aléatoires :

Soit $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire et $\varphi : (\mathbb{R}^d, \mathcal{B}_d) \to (\mathbb{R}, \mathcal{B}_1)$ une fonction mesurable à valeurs réelles. Posons $Y = \varphi \circ X \ (= \varphi(X))$. Alors Y est une v.a. réelle et on a

Théorème : $Y \in \mathcal{L}^1 \iff \varphi \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_d, \mu_X)$ et alors

$$\mathbb{E}(Y) = \int_{\mathbb{R}^d} \varphi(x_1, x_2, \dots, x_d) \, d\mu_X(x_1, x_2, \dots, x_d)$$

et si en particulier X a une densité $f(x_1, \dots, x_d)$

$$\boxed{\mathbb{E}(Y)=\int_{\mathbb{R}^d}\varphi(x_1,\dots,x_d)f(x_1,\dots,x_d)\,dx_1dx_2\dots dx_d}$$

Remarque 2 : les résultats sur l'espérance et les moments vus dans les chapitres précédents sont toujours valables par exemple :

a) Si $X_1,\dots,X_k \in \mathcal{L}^1$ et sont indépendantes alors la v.a. $P=X_1X_2\dots X_k\in\mathcal{L}^1$ et

$$\mathbb{E}(P)=\prod_{i=1}^{k}\mathbb{E}(X_i)$$

b) si $X_1,\dots,X_k$ ont un moment d'ordre 2 et sont indépendantes, alors $S=X_1+X_2+\dots+X_k$ a un moment d'ordre 2 et

$$\mathrm{Var}(S)=\sum_{i=1}^{k}\mathrm{Var}(X_k)$$

etc....

III Les espaces $\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$ et $L^1(\Omega,\mathcal{C},\mathbb{P})$

Les propriétés de l'intégrale (i.e. de l'espérance) p.4 montrent que $\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$ est un espace vectoriel et que l'application

$$X\longmapsto\mathbb{E}(X)$$

de $\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$ dans $\mathbb{R}$ est une forme linéaire.

On peut munir cet espace d'une semi-norme en

posant :

$$\forall X\in\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P}),\ \|X\|_1=\mathbb{E}(|X|)=\int_\Omega|X|\,d\mathbb{P}.$$

On a en effet :

$$\forall X,Y\in\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P}),\ \|X+Y\|_1\le\|X\|_1+\|Y\|_1$$

et

$$\forall\lambda\in\mathbb{R},\forall X\in\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P}),\ \|\lambda X\|_1=|\lambda|\,\|X\|_1$$

De plus $X=0$ implique $\|X\|_1=0$ Mais :

$\|X\|_1=0$ implique $X=0$ $\mathbb{P}$ p.s.

Donc $\|\cdot\|_1$ n'est pas à proprement parler une norme d'espace vectoriel. Mais si on change la notion d'égalité par l'égalité $\mathbb{P}$-p.s. c'est à dire si on considère comme égales deux v.a. X et Y telles que

$$X=Y\ \ \mathbb{P}\text{ p.s.},$$

alors l'espace $\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$ devient l'espace $L^1(\Omega,\mathcal{C},\mathbb{P})$ et $\|\cdot\|_1$ est une norme sur cet espace.

De plus :

> **Théorème** : $(L^1(\Omega,\mathcal{C},\mathbb{P}),\|\cdot\|_1)$ est un espace de Banach (i.e. un espace vectoriel normé complet)

démonstration : c'est un résultat d'intégration (voir le cours de L3 second semestre).

Convergence des suites de v.a. de $\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$

On dit qu'une suite (X_n) de v.a. de $L^1(\Omega, \mathcal{E}, \mathbb{P})$ converge dans L^1 vers une v.a. X si :

$$\lim_{n\to+\infty} \|X_n - X\|_1 = 0.$$

Remarque : La convergence $\mathbb{P}$-presque sure (i.e. $\forall \omega \in \Omega \setminus N$ (où $N \in \mathcal{E}$ et $\mathbb{P}(N)=0$),

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega))$$

n'implique pas la convergence dans L^1 (voir le cours d'intégration 2 du L3 pour des contre-exemples) mais, on a :

> Proposition : Soit (X_n) une suite de v.a. et X une v.a. telles que :
>
> H1) $\lim_{n\to+\infty} X_n = X$ $\mathbb{P}$ p.s
>
> H2) $\exists Y \in L^1(\Omega, \mathcal{E}, \mathbb{P})$ t.q.
>
> $\forall n \in \mathbb{N}$, $|X_n| \leq Y$ $\mathbb{P}$-p.s.
>
> Alors X_n converge vers X dans L^1.

Démonstration : c'est le théorème de convergence dominée (voir le cours d'intégration du L3 1er semestre)

> Proposition : Soit (X_n) une suite de v.a. qui converge dans L^1 vers une v.a. X. Alors (X_n) converge vers X en probabilité

Démonstration : on applique l'inégalité de Markov

$\left(\mathbb{P}(|Z| \geq \varepsilon) \leq \dfrac{\mathbb{E}(|Z|)}{\varepsilon} \text{ si } Z \in \mathcal{L}^1(\Omega, \mathcal{E}, \mathbb{P}) \right)$

à la v.a. $Z = X_n - X$. cqfd.

IV Les espaces $\mathcal{L}^2(\Omega, \mathcal{E}, \mathbb{P})$ et $L^2(\Omega, \mathcal{E}, \mathbb{P})$:

Notation : On note $\mathcal{L}^2(\Omega, \mathcal{E}, \mathbb{P})$ l'ensemble des v.a. définies sur Ω et ayant un moment d'ordre 2.

On sait que si X et Y ont un moment d'ordre 2, $\forall \lambda \in \mathbb{R}$, λX a un moment d'ordre 2 et $X+Y$ a un moment d'ordre 2. Il en résulte que $\mathcal{L}^2(\Omega, \mathcal{E}, \mathbb{P})$ est un espace vectoriel. Pour $X \in \mathcal{L}^2$, on pose

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

et pour X et $Y \in \mathcal{L}^2$,

$$\langle X, Y \rangle_2 = \mathbb{E}(XY).$$

Alors l'application $(X,Y) \longmapsto \langle X, Y \rangle_2$ est un semi-produit scalaire et $\|\cdot\|_2$ la semi-norme associée. Si on remplace l'égalité des v.a. par l'égalité $\mathbb{P}$-p.s (comme on l'a fait pour $\mathcal{L}^1(\Omega, \mathcal{E}, \mathbb{P})$) on obtient l'espace $L^2(\Omega, \mathcal{E}, \mathbb{P})$, l'application $\langle \cdot, \cdot \rangle_2$ est un produit scalaire et $\|\cdot\|_2$ la norme associée sur cet espace.

Rappelons quelques propriétés vues en L3 en calcul intégral (2e semestre) de l'espace

$\mathcal{L}^2(\Omega,\mathcal{C},\mathbb{P})$ ou $L^2(\Omega,\mathcal{C},\mathbb{P})$ qui est essentielle-
-ment le même (modulo l'égalité qui a été modifiée en égalité $\mathbb{P}$-p.s.) :

<u>Inégalité de Cauchy-Schwartz</u> :

$$\forall X,Y\in\mathcal{L}^2,\ \mathbb{E}(|XY|)\leq \mathbb{E}(X^2)^{1/2}\,\mathbb{E}(Y^2)^{1/2}\quad (C.S.)$$

<u>Inégalité de Minkowski</u> :

$$\forall X,Y\in\mathcal{L}^2,\ \|X+Y\|_2\leq\|X\|_2+\|Y\|_2$$

<u>Théorème de Pythagore</u> : Si X et $Y\in\mathcal{L}^2$ sont telles que $\langle X,Y\rangle_2=0$ (i.e. $\mathbb{E}(XY)=0$), alors

$$\|X+Y\|_2^2=\|X\|_2^2+\|Y\|_2^2$$

<u>Remarque</u> : (C.S) avec $Y=\mathbb{1}_\Omega$ implique que

$$\forall X\in\mathcal{L}^2,\ \mathbb{E}(|X|)\leq\mathbb{E}(X^2)^{1/2}.$$

Il en résulte que $\mathcal{L}^2(\Omega,\mathcal{C},\mathbb{P})\subset\mathcal{L}^1(\Omega,\mathcal{C},\mathbb{P})$

<u>Théorème</u> : L'espace $(L^2(\Omega,\mathcal{C},\mathbb{P}),\|\cdot\|_2)$ est un <u>espace de Hilbert</u> (i.e. un espace complet pour la norme $\|\cdot\|_2$)

<u>Convergence des suites de v.a. dans $\mathcal{L}^2(\Omega,\mathcal{C},\mathbb{P})$ (ou $L^2(\Omega,\mathcal{C},\mathbb{P})$)</u>

> <u>Définition</u> : On dit qu'une suite (X_n) de v.a. de L^2 <u>converge dans L^2</u> vers une v.a. X ($\in L^2$) si :
>
> $$\lim_{n\to+\infty}\|X_n-X\|_2=0$$

<u>propriétés de la convergence dans L^2</u> :

> <u>Proposition</u> : Soit (X_n) une suite de v.a. de L^2 et X une v.a. de L^2 telles que : $\lim_{n\to\infty}X_n=X$ dans L^2.
>
> Alors :
>
> 1) $\lim_{n\to\infty}X_n=X$ en probabilité
>
> 2) $\lim_{n\to\infty}X_n=X$ dans L^1

<u>démonstration</u> : 1) on utilise l'inégalité de Markov :

$$\mathbb{P}(|X_n-X|\geq\varepsilon)=\mathbb{P}(|X_n-X|^2\geq\varepsilon^2)\leq\frac{\mathbb{E}(|X_n-X|^2)}{\varepsilon^2}\quad(*)\qquad(\forall\varepsilon>0)$$

2) On utilise l'inégalité de (C.S)

<u>Remarque</u> : on notera que pour des v.a. de L^2, l'inégalité de Markov implique l'inégalité de Bienaymé-Tchebychev $(*)$

<u>Remarque probabiliste sur l'orthogonalité des v.a. de L^2</u> :

Si X et $Y\in L^2$ sont deux v.a. indépendantes alors les v.a. centrées $\tilde{X}=X-\mathbb{E}(X)$ et $\tilde{Y}=Y-\mathbb{E}(Y)$ sont orthogonales. Pour des v.a. centrées, l'indépendance implique l'orthogonalité mais la réciproque est fausse. Notons aussi que la covariance de 2 v.a. X et Y de L^2 est :

$$\mathrm{cov}(X,Y)=\langle\tilde{X},\tilde{Y}\rangle_2$$

V Notions de convergence presque sûre

Soit $(\Omega, \mathcal{C}, \mathbb{P})$ un espace probabilisé, soit (X_n) une suite de v.a. et X une v.a. définies sur Ω à valeurs réelles.

Définition : On dit que la suite (X_n) converge $\mathbb{P}$-presque sûrement vers X si :

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

On écrit alors $\lim_{n\to+\infty} X_n = X$ $\mathbb{P}$.p.s. (ou simplement p.s. s'il n'y a pas d'ambiguité sur la probabilité $\mathbb{P}$).

Théorème : Si $\lim_{n\to+\infty} X_n = X$ p.s., alors on a $\lim_{n\to+\infty} X_n = X$ 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).

dém^n : Soit Ω_1 l'événement de probabilité 1 t.q. $\forall \omega \in \Omega_1$, $\lim_{n\to+\infty} X_n(\omega) = X(\omega)$.

Alors :

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

ce qui équivaut à :

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

ce qui est encore équivalent à :

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

Donc :

$\forall \delta > 0, \; \Omega_1 \subset \liminf_n A_n, \quad (*)$

où $A_n = [|X_n - X| \leq \delta]$.

En passant au complémentaire dans l'inclusion (*) on obtient :

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

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

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

A fortiori (comme $\mathbb{P}(A_n^c) \leq P(\bigcup_{k\geq n} A_k^c)$), on a

$$\mathbb{P}(A_n^c) = \mathbb{P}(|X_n - X| > \delta) \to 0 \text{ si } n \to +\infty$$

cqfd.

Un critère suffisant de convergence p.s. :

Dire qu'une suite (X_n) converge p.s. vers une v.a. X équivaut à dire que $X_n - X$ converge vers 0 p.s. Pour alléger les énoncés on va donc considérer la convergence vers zéro p.s. et donner un critère suffisant pour que X_n (resp. $X_n - X$) converge vers 0 p.s quand $n \to +\infty$:

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

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

Alors: $\lim_{k\to\infty} X_k = 0$ 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'évènements 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:

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

forment une suite décroissante d'évènements. 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.

dém^{n} du théorème: D'après le lemme de B.C.:

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

En passant au complémentaire, on obtient

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

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

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

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

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 $\lim_{k\to\infty} X_k(\omega) = 0$, CQFD.

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

$$X_k \longrightarrow 0 \text{ p.s.}$$

dém^{n}: $\mathbb{P}(|X_k| > \varepsilon) = \mathbb{P}(|X_k|^p > \varepsilon^p) \leq \dfrac{E(|X_k|^p)}{\varepsilon^p}$ d'après l'inégalité de Markov. Donc la série $\sum_{k=1}^{\infty} \mathbb{P}(|X_k| > \varepsilon)$ converge et le théorème s'applique CQFD

VI) La loi forte des grands nombres

<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. centrée $X_k - m$, on se ramène à démontrer le théorème dans le cas $m=0$, ce que nous supposerons :

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

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

Ce qui montre que $\frac{S_n}{n} \underset{(n\to\infty)}{\longrightarrow} 0$ 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 $\sum_{n=1}^{\infty} \mathbb{E}(Y_n^2) < +\infty$.

Donc $\frac{S_{n^2}}{n^2} \underset{(n\to\infty)}{\longrightarrow} 0$ 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

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

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

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

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

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

qui est le terme général d'une série convergente.

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

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

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

> <u>Théorème</u> (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 <u>admis</u>).

III) <u>Lien de la convergence en loi avec les autres modes de convergence des suites de v.a.</u>

Soit (X_n) une suite de v.a. et X une v.a.

<u>Rappel</u> : On dit que $X_n\to X$ en loi (si $n\to+\infty$) si en tout point $x\in\mathbb{R}$ où $F_X(x)$ est continue, on a :

$$\lim_{n\to+\infty}F_{X_n}(x)=F_X(x)$$

(où $F_X(x)=\mathbb{P}(X\leq x)$ est la fonction de répartition de la v.a. X).

> <u>Théorème</u> : Si $X_n\to X$ en probabilité (qd $n\to+\infty$) alors $X_n\to X$ en loi (qd $n\to\infty$)
> i.e. la convergence en probabilité implique la convergence en loi.

> <u>Lemme</u> : Soient X et Y deux v.a. quelconques. On a :
> $\forall\eta>0,\ \forall x\in\mathbb{R}$,
> $$|F_X(x)-F_Y(x)|\leq F_X(x+\eta)-F_X(x-\eta)+\mathbb{P}(|X-Y|>\eta)$$

<u>dém$^{\text{ti}}$</u> : On a

$$[Y\leq x]=[Y\leq x, X\leq x+\eta]\cup[Y\leq x, X>x+\eta]$$
$$\subset[X\leq x+\eta]\cup[|X-Y|>\eta]$$

$$\Rightarrow F_Y(x)\leq F_X(x+\eta)+\mathbb{P}(|X-Y|>\eta)$$

de même, on a :

$F_X(x-\eta) \leq F_Y(x) + \mathbb{P}(|X-Y|>\eta)$

(remplacer X par Y et x par $x-\eta$)

$\Longrightarrow$

(*) $F_X(x-\eta) - \mathbb{P}(|X-Y|>\eta) \leq F_Y(x) \leq F_X(x+\eta) + \mathbb{P}(|X-Y|>\eta)$

Or trivialement, on a :

(**) $F_X(x-\eta) \leq F_X(x) \leq F_X(x+\eta)$ (F_X est f^on^ croissante)

($\Rightarrow -F_X(x+\eta) \leq -F_X(x) \leq -F(x-\eta)$)

On en déduit (en (*) et (**)) :

$$F_X(x-\eta) - F_X(x+\eta) - \mathbb{P}(|X-Y|>\eta) \leq F_Y(x) - F_X(x) \leq F_X(x+\eta) - F_X(x-\eta) + \mathbb{P}(|X-Y|>\eta)$$

D'où :

$$|F_Y(x) - F_X(x)| \leq F_X(x+\eta) - F_X(x-\eta) + \mathbb{P}(|X-Y|>\eta).$$

cqfd.

<u>dém du théorème</u> : avec $Y = X_n$, on a donc

$$|F_{X_n}(x) - F_X(x)| \leq F_X(x+\eta) - F_X(x-\eta) + \mathbb{P}(|X_n - X|>\eta)$$

Soit x un point de continuité de F_X. Alors

$\forall \varepsilon>0, \exists \eta$ t.q. $F_X(x+\eta) - F_X(x-\eta) \leq \varepsilon/2$, on garde ce η

Comme $X_n \to X$ en proba, $\exists N\ (\in \mathbb{N})$ tq

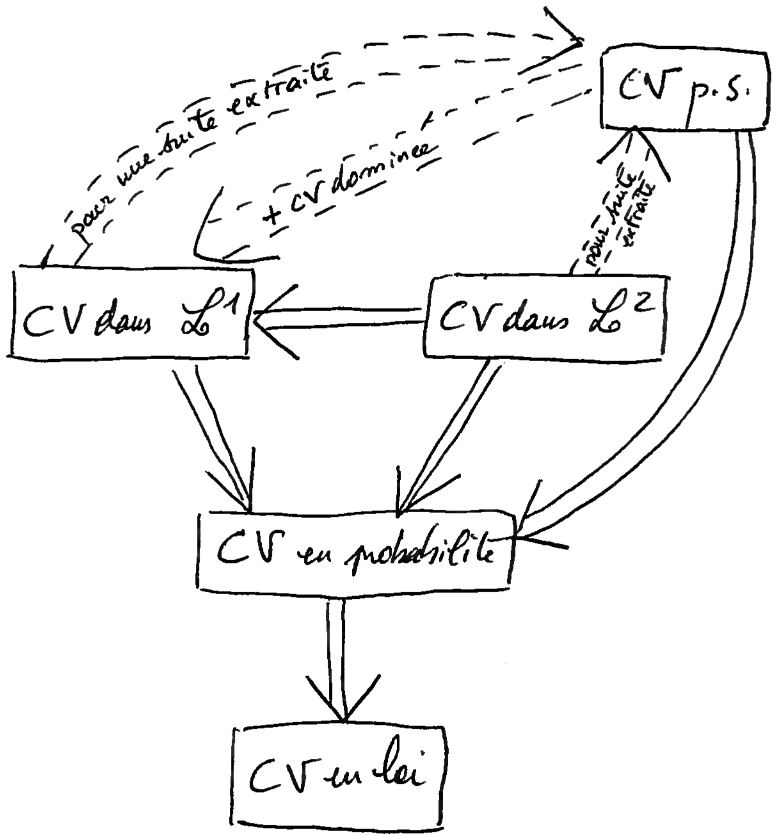
$n \geq N \Rightarrow \mathbb{P}(|X_n - X|>\eta) \leq \varepsilon/2$.

<u>Au total</u> : $\forall \varepsilon>0, \exists N (\in \mathbb{N})$ t.q. $n \geq N \Longrightarrow$

$$|F_{X_n}(x) - F_X(x)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

donc $X_n \to X$ en loi. cqfd.

<u>Récapitulation des différents modes de convergence pour une suite de variables aléatoires</u>



FIN du COURS de PROBABILITÉS du 1er semestre