

I - Introduction

La thèse telle qu'introduite aux chapitres 1 et 2 se justifie rigoureusement à l'aide de l'intuition selon laquelle un nombre infini de répétitions d'une expérience permet à la fréquence de réalisation d'un événement de converger vers sa probabilité de réalisation et être démontrée.

Cette loi des grands nombres est donc un résultat fondamental des probabilités.

Dans ce contexte, on considère un espace d'états Ω muni de la tribu \mathcal{F} et de la probabilité \mathbb{P} . On veut étudier la répartition des valeurs d'une v.a. X de loi \mathbb{P}_X , réalisées au cours d'une succession de n expériences aléatoires \mathcal{E} .

Ces résultats sont modélisés par une suite X_1, \dots, X_n de v.a. \mathbb{I} et de même loi.

On s'intéresse en particulier au comportement de la moyenne empirique pd $n \rightarrow +\infty$, c'est-à-dire comme définir $\lim_{n \rightarrow +\infty} \frac{1}{n} (X_1 + \dots + X_n)$, sachant que chaque X_i est une fonction?

II - Modes de convergence de variables aléatoires.

Nous distinguons à l'aide de la convergence en loi de toutes les autres formes de convergence. En effet, ces dernières conduisent une proximité des variables aléatoires elles-mêmes.

Pour étudier ces différentes formes de convergence, on considère une suite $(X_n)_{n \geq 1}$ de v.a. définies sur un espace de probabilité (Ω, \mathcal{F}, P) , et à valeurs dans \mathbb{R}^d . On considère aussi une v.a. limite X sur le même espace.

① - Convergence presque sûre

Définition: La suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X s'il existe un ensemble $N \in \mathcal{F}$ de probabilité nulle tel que

$$X_n(\omega) \rightarrow X(\omega) \text{ quand } n \rightarrow +\infty \text{ pour tout } \omega \notin N.$$

→ Notation: on note en général $X_n \xrightarrow[n \rightarrow \infty]{} X$ p.s., ou $X_n \xrightarrow{\text{p.s.}} X$.

→ La convergence presque sûre est la plus proche de la convergence simple des fonctions. Sauf p.s., nous permettons à certains ω de ne pas vérifier $X_n(\omega) \rightarrow X(\omega)$ si la proba. de réalisation de l'ensemble de ces ω est nulle (ensemble négligeable).

② - Convergence en probabilité

Définition: La suite $(X_n)_{n \geq 1}$ converge en probabilité vers X si $\forall \epsilon > 0$, on a

$$P(|X_n - X| \geq \epsilon) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Notation: $X_n \xrightarrow{P} X$.

③ - Convergence en moyenne

Définition: La suite $(X_n)_{n \geq 1}$ converge en moyenne vers X si

- X_n et X sont dans L^1
- $E[|X_n - X|] \rightarrow 0$ quand $n \rightarrow +\infty$.

→ Notation: $X_n \xrightarrow{L^1} X$.

(4) Relations entre modes de convergence.

(2)

Ces convergences ne sont bien évidemment pas équivalentes, et impliquent des notions différentes. Prenons quelques exemples pour le montrer.

→ Exemple 1: soit $(X_n)_n$ une suite de v.a. de Bernoulli telle que

$$P(X_n=1) = \frac{1}{n} \quad \text{et} \quad P(X_n=0) = 1 - \frac{1}{n}.$$

• Pour tout $\varepsilon \in]0, 1[$, la probabilité $P(|X_n| > \varepsilon) = \frac{1}{n}$ tend vers 0 qd $n \rightarrow +\infty$.

⇒ la suite $(X_n)_n$ tend vers $X=0$ en probabilité.

• Comme $E[X_n] = \frac{1}{n}$, elle tend aussi en moyenne vers $X=0$.

→ Exemple 2: on considère maintenant une suite $(Y_n)_n$ de v.a. de Bernoulli k.p.:

$$P(Y_n = n^2) = \frac{1}{n} \quad \text{et} \quad P(Y_n = 0) = 1 - \frac{1}{n}.$$

• De la même manière que précédemment, $P(|Y_n| > \varepsilon) = \frac{1}{n} \xrightarrow{n \rightarrow +\infty} 0$

On a donc $Y_n \xrightarrow{P} Y=0$.

• Par contre $E[Y_n] = 0 \times (1 - \frac{1}{n}) + n^2 \times \frac{1}{n} = n \Rightarrow (Y_n)_n$ ne converge pas en moyenne vers 0 (ni aucune autre limite finie).

→ Exemple 3: soit U une v.a. uniforme sur $[0, 1]$.

On pose $Z_n = \mathbb{1}_{\{U \leq \frac{1}{n}\}}$. Alors $P(Z_n=1) = \frac{1}{n}$ et $P(Z_n=0) = 1 - \frac{1}{n}$.

• $Z_n \xrightarrow{P} 0$.

• De plus, $Z_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. En effet, si ω est fixé, alors dès que $U(\omega) > 0$ (ce qui est vrai avec proba 1), il existe n_0 tel que $U(\omega) > \frac{1}{n_0}$, et donc tel que $Z_n(\omega) = 0, \forall n \geq n_0$.

Théorème: Supposons que les X_n et X soient intégrables.

Si $X_n \xrightarrow{L^1} X$ alors $X_n \xrightarrow{P} X$.

Preuve: immédiate en utilisant l'inégalité de Markov. En effet, $\forall \epsilon > 0$,

$$P(|X_n - X| \geq \epsilon) \leq \frac{E[|X_n - X|]}{\epsilon}$$

On voit de suite que la convergence en proba vers X découle de la convergence en moyenne vers X (le membre de droite tend vers 0).

→ Δ: La réciproque est fautive, cf exemple 2.

Théorème: (de Lebesgue, ou de CV dominée)

Si les v.a. X_n convergent presque sûrement sur Ω vers une limite X , et
si $\forall n, |X_n| \leq Z$ avec $Z \in L^1$

Alors X_n et X sont intégrables et on a $E[|X_n - X|] \xrightarrow{n \rightarrow +\infty} 0$.

Preuve: admis (difficile).

→ La CV p.s. peut donc dans certains cas de domination entraîner la CV en moyen.

→ On a en particulier également selon ce théorème: $E[X_n] \xrightarrow{n \rightarrow +\infty} E[X]$, si les hypothèses sont respectées.

→ Δ: la réciproque est fautive. Dans l'exemple 1, supposons les X_n II.

Puisque $P(X_n > \epsilon) = \frac{1}{n}$, la série de terme général $P(X_n > \epsilon) = \frac{1}{n}$ diverge.

Par ailleurs, les événements $A_n = \{X_n > \epsilon\}$ sont II. Par Borel-Cantelli, on en déduit que pour presque tout ω , une infinité de A_n sont réalisées ($X_n(\omega) > \epsilon$).

⇒ La suite ne peut pas CV vers 0. Ainsi, la suite $(X_n)_n$ est bornée par 1.

⇒ elle tend en moyenne vers 0, mais pas presque sûrement.

(3)

Par contre, si on prend la suite $(V_n)_n$ avec, pour $\alpha > 1$,

$$P(V_n=1) = \frac{1}{n^\alpha} = 1 - P(V_n=0).$$

Puisque $\alpha > 1$, la série de terme général $P(V_n \geq \varepsilon)$ converge, et donc par Borel-Cantelli, on sait que pour presque tout ω , un nombre fini au plus de $V_n(\omega)$ seront supérieurs à $\varepsilon \Rightarrow (V_n)_n$ converge presque sûrement vers 0.

→ L'hypothèse de domination est nécessaire - Si par exemple, on prend la suite $(T_n)_n$ définie par $P(T_n=n^2) = \frac{1}{n\sqrt{n}} = 1 - P(T_n=0)$.

On montre comme ci-dessus que $T_n \xrightarrow{p.s.} 0$.

Mais $E[T_n] = \sqrt{n}$, et donc on n'a pas la convergence en moyenne!

Proposition: Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{P} X$.

Preuve: Soit $A_{n,\varepsilon} = \{\omega, |X_n(\omega) - X(\omega)| \geq \varepsilon\}$. Supposons que $X_n \xrightarrow{n \rightarrow \infty} X$ p.s., et soit N l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \rightarrow X(\omega)$. Si $\omega \notin N$, alors $\omega \notin A_{n,\varepsilon}$, $\forall n \geq n_0$ (où n_0 dépend de ω et ε). Ceci implique que les v.a. $Y_{n,\varepsilon} = 1_{N^c \cap A_{n,\varepsilon}}$ tendent simplement vers 0 pd $n \rightarrow \infty$. Comme on a aussi $0 \leq Y_{n,\varepsilon} \leq 1$, le théorème de Lebesgue (v.dominée) entraîne que $E[Y_{n,\varepsilon}] \xrightarrow{n \rightarrow \infty} 0$. Mais $P(A_{n,\varepsilon}) \leq P(N^c \cap A_{n,\varepsilon}) + P(N) = P(N^c \cap A_{n,\varepsilon}) = E[Y_{n,\varepsilon}] \xrightarrow{n \rightarrow \infty} 0$ et on en déduit la convergence en proba.

Proposition: S'il existe une constante α telle que $|X_n| \leq \alpha$ presque sûrement, il y a équivalence entre $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{L^1} X$.

Preuve: Sachant que si $X_n \xrightarrow{L^1} X$ alors $X_n \xrightarrow{P} X$, il suffit de montrer que si $|X| \leq a$ p.s. alors $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L^1} X$.

• Comme $|X| \leq a$, nous avons $\{|X| > a + \varepsilon\} \subset A_{n, \varepsilon}$, et donc $P(|X| > a + \varepsilon) \leq P(A_{n, \varepsilon})$. En faisant tendre $n \rightarrow +\infty$, on en déduit que $P(|X| > a + \varepsilon) = 0$. Ceci est vrai $\forall \varepsilon$, et donc $P(|X| > a) = 0$.

• Comme $|X_n| \leq a$, nous avons aussi $|X_n - X| \leq \varepsilon + (|X_n| + |X|) 1_{A_{n, \varepsilon}} \leq \varepsilon + 2a 1_{A_{n, \varepsilon}}$ sur l'ensemble $\{|X| \leq a\}$, qui est de proba 1. Donc il vient

$$E[|X_n - X|] \leq \varepsilon + 2a P(A_{n, \varepsilon})$$

En utilisant la définition de la CV en probabilité, on obtient

$\limsup_n E[|X_n - X|] \leq \varepsilon$, et comme ε est arbitrairement proche de 0, on obtient la CV en moyenne. CQFD.

Proposition: Soit f une fonction continue de \mathbb{R}^d dans \mathbb{R} .

a) Si $X_n \xrightarrow{p.s.} X$, alors $f(X_n) \xrightarrow{p.s.} f(X)$.

b) Si $X_n \xrightarrow{P} X$, alors $f(X_n) \xrightarrow{P} f(X)$.

Preuve: a) est évident. Pour b) on peut d'abord remarquer que si $K > 0$ et $\varepsilon > 0$,

$$\{|f(X_n) - f(X)| \geq \varepsilon\} \subset \{|X| > K\} \cup \{|X| \leq K, |f(X_n) - f(X)| \geq \varepsilon\} \quad (1)$$

La fonction est uniformément continue sur $\{x: |x| \leq 2K\}$, donc $\exists \eta > 0$ t.p. $|x - y| < \eta$ et $|x| \leq 2K, |y| \leq 2K$ impliquent que $|f(x) - f(y)| < \varepsilon$. Ainsi, en choisissant η suffisamment petit, $|x| \leq K$ entraîne $|y| \leq 2K$. L'équation (1) implique $\{|f(X_n) - f(X)| \geq \varepsilon\} \subset \{|X| > K\} \cup \{|X_n - X| \geq \eta\}$ donc $P(|f(X_n) - f(X)| \geq \varepsilon) \leq P(|X| > K) + P(|X_n - X| \geq \eta)$. D'après l'hypothèse b), on a $\limsup_n P(|f(X_n) - f(X)| \geq \varepsilon) \leq P(|X| > K)$. Enfin, $\lim_{K \rightarrow +\infty} P(|X| > K) = 0$ (grâce au théo de CV dominée), et donc le limsup est nul \Rightarrow on a le résultat.

III - La loi des grands nombres.

(4)

On considère de nouveau une suite $(X_n)_{n \geq 1}$ de v.a. réelles iid.

Soit M_n la moyenne des n premières v.a., donc

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

L'objectif est de montrer que M_n converge vers l'espérance des v.a. X_n lorsque cette dernière existe.

Énonçons dans un premier temps la loi des grands nombres pour des v.a. de carré intégrable.

Théorème: Soit $(X_n)_n$ une suite de v.a. iid, et de carré intégrable ($X_n \in L^2$).

général Soit $m = E[X_n]$ leur moyenne commune.

Alors la suite $(M_n)_n$ définie par $M_n = \frac{1}{n} (X_1 + \dots + X_n)$ converge vers m , presque sûrement et en moyenne, quand $n \rightarrow +\infty$.

Elle converge donc aussi en probabilité.

On a même plus que la convergence en moyenne, puisque $E[(M_n - m)^2] \xrightarrow{n \rightarrow +\infty} 0$.

Preuve: On note σ^2 la variance commune à tous les X_n ($X_n \in L^2$). Par

linéarité de l'espérance, $E[M_n] = E\left[\frac{1}{n} (X_1 + \dots + X_n)\right] = \frac{1}{n} E[nX_1] = m$

et $E[(M_n - m)^2] = \text{Var}(M_n) = \text{Var}\left(\frac{1}{n} (X_1 + \dots + X_n)\right) = \frac{\sigma^2}{n}$ (on a donc la dernière assertion du théorème).

• Comme $E[|Y|]^2 \leq E[Y^2]$, on en déduit que $E[|M_n - m|] \xrightarrow{n \rightarrow +\infty} 0$ grâce au point précédent. On obtient donc la CV en moyenne - (et donc aussi en proba).

- Reste à démontrer la convergence presque sûre: supposons pour simplifier que $m=0$ (ce qui reviendrait à remplacer X_n par X_{n-m} , et donc M_n par M_{n-m}).

→ Montrons d'abord que la sous-suite $(M_{n^2})_n$ converge presque sûrement vers 0:

d'après l'inégalité de Bienaymé-Chebyshev, pour $q \in \mathbb{N}^*$,

$$P(|M_{n^2}| \geq \frac{1}{q}) \leq \frac{\text{Var}(M_{n^2})}{(1/q)^2} = \frac{\sigma^2 q^2}{n^2} \quad (\text{on a remplacé } n \text{ par } n^2, \text{ et } \sigma^2 \text{ par } \sigma^2 q^2)$$

Donc si $A_{n,q} = \{|M_{n^2}| \geq \frac{1}{q}\}$, nous obtenons $\sum_{n \geq 1} P(A_{n,q}) < +\infty$.

Posons ensuite $B_{n,q} = \bigcup_{m \geq n} A_{m,q}$ et $C_q = \bigcap_{n \geq 1} B_{n,q} = \limsup_n A_{n,q}$. En appliquant le lemme de Borel-Cantelli, on montre alors que $P(C_q) = 0$.

Par suite, si nous posons $N = \bigcup_{q \in \mathbb{N}^*} C_q$, nous obtenons $P(N) \leq \sum_{q=1}^{\infty} P(C_q) = 0$.

Maintenant, si $w \notin N$, alors $w \in \bigcap_{q \in \mathbb{N}^*} (C_q)^c$. Ainsi, $w \notin C_q$ pour tout $q \geq 1$, donc aussi $w \notin B_{n,q}$ pour n assez grand (car $B_{n,q}$ est décroissant en n).

Cela signifie que $\forall w \notin N, \forall q \geq 1$, il existe n assez grand tel que

$M_{k^2}(w) \leq \frac{1}{q}$ dès que $k \geq n$. En d'autres mots, $M_{n^2}(w) \rightarrow 0$ si $w \notin N$, donc par définition $M_{n^2} \xrightarrow{p.s.} 0$.

→ Montrons maintenant que la suite $(M_n)_n$ tend presque sûrement vers 0:

$\forall n \in \mathbb{N}^*$, notons $p(n)$ l'entier tel que $p(n)^2 \leq n < (p(n)+1)^2$. Alors

$$M_n = \frac{p(n)^2}{n} M_{p(n)^2} = \frac{1}{n} \sum_{p=p(n)^2+1}^n X_p$$

Puisque les v.a. de la somme sont II,

$$E\left[\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right)^2\right] = E\left[\left(\frac{1}{n} \sum_{p=p(n)^2+1}^n X_p\right)^2\right] = \frac{n - p(n)^2}{n^2} \sigma^2 \leq \frac{2p(n)+1}{n^2} \sigma^2$$

car $p(n) \leq \sqrt{n}$. $\leq \frac{2\sqrt{n}+1}{n^2} \sigma^2$

On applique de nouveau l'inégalité de Bienaymé-Tchebychev:

(5)

$$P\left(\left|M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right| \geq a\right) \leq \frac{1}{a^2} \text{Var}\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right) \text{ avec}$$

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} \text{ une v.a. centrée, donc } \text{Var}\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right) = E\left[\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right)^2\right]$$

$$\text{d'où } P\left(\left|M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right| \geq a\right) \leq \frac{2\sqrt{n}+1}{n^2} \frac{\sigma^2}{a^2}.$$

Comme la série $\sum_n \frac{2\sqrt{n}+1}{n^2}$ converge, le même raisonnement que précédemment (pour la sous-suite) montre que $M_n - \frac{p(n)^2}{n} M_{p(n)^2} \xrightarrow{P.S.} 0$.

Or $M_{p(n)^2} \xrightarrow{P.S.} 0$, et $\frac{p(n)^2}{n} \rightarrow 1$. On en déduit que $M_n \xrightarrow{P.S.} 0$. CQFD
on rappelle que cela représente $(M_n - m)$.

①. Loi faible des grands nombres.

Théorème: Soient des v.a. X_n indépendantes, de même loi et intégrables.

Notons $m = E(X_n)$ leur moyenne commune. Alors la moyenne empirique

$\frac{1}{n}(X_1 + \dots + X_n)$ converge en probabilité vers, quand $n \rightarrow +\infty$.
(moyenne)

→ Nous avons vu que la convergence en moyenne implique la convergence en probabilité.

→ Cette loi faible est assez ^{peu} utile et informel: elle permet en général d'obtenir certains contrôles d'erreurs (dans des preuves par exemple).

② - la loi forte des grands nombres.

Théorème : Soient des v.e. X_n iid et intégrables. On note $m = E[X_n]$ leur moyenne commune. Alors la moyenne empirique $\frac{1}{n} (X_1 + \dots + X_n)$ converge presque sûrement vers m , quand $n \rightarrow +\infty$.

→ Ce résultat est plus fort et donne une information sur la trajectoire $n \rightarrow X_n(\omega)$, pour presque tout ω .

→ Il faut que les v.e. soient ii : en effet, si on prend par exemple toutes les X_n égales à X une v.e., alors $M_n = X$, qui ne converge vers m que si X est constante égale à m .

→ Un exemple illustratif :

Considérons un jeu de Pile ou Face infini, donc une suite $(X_n)_n$ de v.e. ne prenant que des valeurs 0 et 1. L'espace d'états de cette suite est $\Omega = \{0, 1\}^{\mathbb{N}}$. Ainsi, un point ω est une suite numérique x_1, \dots, x_n de 0 et de 1. Chaque suite est possible.

Soit \mathbb{P}_p la probabilité sous laquelle les X_n sont ii, de même loi $B(p)$, $p \in]0, 1[$. La loi des grands nombres nous dit que pour toute suite $(x_n)_n$ en dehors d'un ensemble de proba nulle, la moyenne $\frac{1}{n} (x_1 + \dots + x_n)$ tend vers p quand $n \rightarrow +\infty$.

Mais il existe évidemment bc de suites ne vérifiant pas cette propriété, par ex $x_n = 0 \forall n$.

Si on considère maintenant la probabilité \mathbb{P}_q définie de la même manière pour $q \neq p$, l'ensemble de probabilité 1 pour \mathbb{P}_p des suites $(x_n)_n$ qui tendent vers p , devient sous \mathbb{P}_q un ensemble de proba. nulle (donc un ensemble négligeable).

On remarque aussi que la proba. de chaque suite particulière est nulle. (elle vaut $\lim_{k \rightarrow \infty} p^{k_1} (1-p)^{k_2}$, $k_1 + k_2 = k$).

(6)

⇒ Cet exemple montre que l'étude de la convergence des variables aléatoires nécessite l'introduction de la convergence presque sûre, car on n'a généralement pas la convergence simple (i.e. pour tout ω).

IV - Résumé des liens entre modes de convergence.

On peut résumer les différentes implications entre modes de convergence à travers le schéma suivant :



