

Convergences des suites de variables aléatoires

Prépa Agreg Option A

Nathan Huguenin

Introduction

Largement inspiré du livre de M-L. Chabanol et J-J. Ruch.

1 Convergences en probabilité, dans L^p , et presque sûre

On commence par définir les différentes notions de convergence de la suite (X_n) de v.a. vers X où les X_n et X sont définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^d$.

Definition 1.1. On dit que (X_n) converge presque sûrement (p.s.) vers X si

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\}) = 1.$$

Remark 1.2. Par définition de la limite,

$$\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\} = \bigcap_{p \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \frac{1}{p}\}.$$

Donc $X_n \rightarrow X$ p.s. si et seulement si

$$\forall p > 0 \quad \mathbb{P}(\liminf |X_n - X| > \frac{1}{p}) = 1$$

ou encore

$$\forall p > 0 \quad \mathbb{P}(\limsup |X_n - X| < \frac{1}{p}) = 0.$$

Le lemme de Borel-Cantelli permet de montrer le résultat suivant.

Lemma 1.3. • Si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$, alors $X_n \rightarrow X$ p.s.

- Si les X_n sont indépendantes, alors la suite (X_n) tend vers 0 p.s. si et seulement si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) < \infty$.

Exercice 1.4. Faire la preuve du lemme.

Correction. Pour $p \geq 1$ on note $A_{n,p}$ l'événement $|X_n - X| \geq \frac{1}{p}$. Alors par Borel-Cantelli, $\sum \mathbb{P}(A_{n,p}) < \infty$ implique $\mathbb{P}(\limsup_n A_{n,p}) = 0$, ce qui est équivalent à la convergence p.s. par la remarque au-dessus. Réciproquement si $\sum \mathbb{P}(|X_n| \geq \varepsilon) = \infty$ et si les X_n sont indépendantes, on déduit de Borel-Cantelli que pour presque tout ω , l'ensemble des n t.q. $|X_n| \geq \varepsilon$ est infini, donc X_n ne peut pas converger p.s. vers 0. $\square$

Definition 1.5. On dit que (X_n) converge en probabilité vers X si

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Proposition 1.6. *L'espace $L^0_{\mathbb{R}^d}(\Omega, \mathcal{F}, \mathbb{P})$ des v.a. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, modulo l'égalité p.s., est complet pour la distance $d(X, Y) = \mathbb{E}[|X - Y| \wedge 1]$, qui est compatible avec la convergence en probabilité :*

$$X_n \rightarrow X \text{ en proba.} \Leftrightarrow d(X_n, X) \rightarrow 0.$$

Proof. On omet la preuve que d est bien une distance sur L^0 . Si (X_n) converge en proba vers X alors pour tout $\varepsilon > 0$ suffisamment petit on a

$$d(X_n, X) \leq \mathbb{E}[|X_n - X| \mathbf{1}_{|X_n - X| \leq \varepsilon}] + \mathbb{E}[|X_n - X| \wedge 1 \mathbf{1}_{|X_n - X| > \varepsilon}] \leq \varepsilon + \mathbb{P}(|X_n - X| \geq \varepsilon).$$

Ainsi $\limsup d(X_n, X) \leq \varepsilon$ et on a bien $d(X_n, X) \rightarrow 0$. Réciproquement si $d(X_n, X) \rightarrow 0$ alors pour tout $0 < \varepsilon < 1$,

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon} d(X_n, X) \rightarrow 0.$$

Soit maintenant (X_n) une suite de Cauchy pour d . Il existe une sous-suite $(Y_k) = (X_{n_k})$ t.q. pour tout $k \geq 1$ $d(Y_k, Y_{k+1}) \leq 2^{-k}$. Alors $\sum d(Y_k, Y_{k+1}) < \infty$, donc $\sum |Y_k - Y_{k+1}| \wedge 1 < \infty$ p.s. par Fubini-Tonelli, et donc $\sum |Y_k - Y_{k+1}| < \infty$. Soit $X = Y_1 + \sum (Y_{k+1} - Y_k)$, alors $Y_k \rightarrow X$ p.s., puis $d(X_n, X) \rightarrow 0$. $\square$

Corollary 1.7 (Corollaire de la preuve). *Si (X_n) converge en probabilité vers X , il existe une sous-suite qui converge p.s. vers X .*

Definition 1.8. *Si $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, on dit que (X_n) converge dans L^p vers X si*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Exercice 1.9. *Soit $c \in \mathbb{R}^d$. Montrer que $X_n \rightarrow c$ dans L^2 ssi $\mathbb{E}[X_n] \rightarrow c$ et $\text{Var}[X_n] \rightarrow 0$.*

Proof. $\mathbb{E}[(X_n - c)^2] = \text{Var}[X_n] + (\mathbb{E}[X_n] - c)^2$. $\square$

Proposition 1.10. *Si (X_n) converge p.s. ou dans L^p vers X , elle converge aussi en probabilité vers X .*

Exercice 1.11. *Prouver la proposition.*

Correction. Si (X_n) converge p.s. vers X , alors pour tout $\varepsilon > 0$, $\mathbf{1}_{|X_n - X| > \varepsilon}$ converge p.s. vers 0, et est majorée par 1, donc $\mathbb{P}(|X_n - X| \geq \varepsilon) \rightarrow 0$ par convergence dominée. Pour la deuxième implication, c'est simplement l'inégalité de Markov. $\square$

Proposition 1.12. *Si (X_n) converge en probabilité vers X et s'il existe $1 < q < \infty$ t.q. (X_n) soit bornée dans L^q , alors pour tout $1 < p < q$, (X_n) converge dans L^p vers X .*

Proof. Soit $1 < p < q$. Il existe M t.q. pour tout n , $\mathbb{E}[|X_n|^q] \leq M$. Par le lemme de Fatou appliqué à une sous-suite qui converge p.s., on a aussi $\mathbb{E}[|X|^q] \leq M$. On utilise en suite l'inégalité de Hölder : pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}[|X_n - X|^p] &= \mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| > \varepsilon}] + \mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| \leq \varepsilon}] \\ &\leq (\mathbb{E}[|X_n - X|^q])^{\frac{p}{q}} (\mathbb{P}(|X_n - X| > \varepsilon))^{\frac{1-p}{q}} + \varepsilon^p \\ &\leq (2M)^{\frac{p}{q}} (\mathbb{P}(|X_n - X| > \varepsilon))^{\frac{1-p}{q}} + \varepsilon^p. \end{aligned}$$

D'où quand $n \rightarrow \infty$, $\limsup \mathbb{E}[|X_n - X|^p] \leq \varepsilon^p$, et donc $\mathbb{E}[|X_n - X|^p] \rightarrow 0$. $\square$

Proposition 1.13. *Les fonctions continues préservent les convergences p.s. et en proba. La somme préserve les trois types de convergence.*

Remark 1.14. *C'est faux pour la convergence L^p en général.*

2 Convergence en loi

Definition 2.1. Une suite (X_n) de v.a. à valeurs dans $\mathbb{R}^d$ converge en loi vers X si pour toute fonction φ continue bornée sur $\mathbb{R}^d$, $\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)]$.

Ici, on ne suppose plus les X_n définies sur le même espace probabilisé. D'ailleurs, seule la loi de proba de "la" limite X est définie de manière unique.

Theorem 2.2 (Théorème de Lévy). La convergence en loi est équivalente à la convergence ponctuelle des fonctions caractéristiques.

Pour la preuve, on pourra se référer au Chabanol-Ruch (c'est une preuve d'analyse de Fourier).

Theorem 2.3. Soient (X_n) et X des v.a. de fonctions de répartition F_{X_n} et F_X respectivement. La suite (X_n) converge en loi vers X ssi $F_{X_n}(t) \rightarrow F_X(t)$ pour tout t où F_X est continue.

Pour la preuve, cf Chabanol-Ruch. Dans le cas discret, cela donne le résultat suivant.

Proposition 2.4. (X_n) converge en loi vers X ssi pour tout $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.

Et dans le cas des v.a. à densité :

Proposition 2.5 (Lemme de Scheffé). Soient X_n et X des v.a. de densités respectives f_n et f . On suppose que $f_n(x) \rightarrow f(x)$ pour presque tout x . Alors $f_n \rightarrow f$ dans L^1 et X_n converge en loi vers X .

Proposition 2.6. La convergence en probabilité implique la convergence en loi.

Exercice 2.7. Le prouver.

Correction. Si φ est continue bornée, alors pour tout $\varepsilon > 0$, $|\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leq \varepsilon + 2\|\varphi\|_\infty \mathbb{P}(|X_n - X| > \varepsilon)$, d'où $0 \leq \limsup |\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leq \varepsilon$. $\square$

Exercice 2.8. Prouver le résultat suivant : la convergence en loi vers une constante implique la convergence en probabilité.

Correction. Dans le cas $d = 1$,

$$\mathbb{P}(|X_n - a| > \varepsilon) = \mathbb{P}(X_n \geq a + \varepsilon) + \mathbb{P}(X_n \leq a - \varepsilon) \leq 1 - F_n(a + \varepsilon) + F_n(a - \varepsilon).$$

Or F_n est continue en $a \pm \varepsilon$, d'où $\lim 1 - F_n(a + \varepsilon) + F_n(a - \varepsilon) = 0$. $\square$

Proposition 2.9. Les fonctions continues préservent la convergence en loi.

Theorem 2.10 (Lemme de Slutsky). Soient (X_n) et (Y_n) deux suites de v.a. définies sur le même espace probabilisé. On suppose que X_n converge en loi vers X et que Y_n converge en loi vers une constante a . Alors le couple (X_n, Y_n) converge en loi vers (X, a) . En particulier, $X_n Y_n \rightarrow aX$ et $X_n + Y_n \rightarrow X + a$ en loi.

Pour la preuve, voir le livre. Ce résultat est particulièrement utilisé en Statistique.

3 Loi des grands nombres

La loi des grands nombres "signifie" que l'espérance d'une variable aléatoire formalise la moyenne des résultats obtenus sur un grand nombre de répétitions de la même expérience. Les différents énoncés correspondent à des convergences plus ou moins fortes, au prix d'hypothèses plus ou moins coûteuses. On commence par deux versions faciles à démontrer.

Exercice 3.1 (Loi faible des grands nombres). Prouver le résultat suivant. Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. Si $\mathbb{E}[X_1^2] < \infty$ alors $\frac{1}{n} \sum_{k=1}^n X_k$ converge dans L^2 vers $\mathbb{E}[X_1]$.

Proof. $\mathbb{E}[(\bar{X}_n - \mathbb{E}[X_1])^2] = \frac{1}{n} \text{Var}(X_1) \rightarrow 0.$ $\square$

Exercice 3.2 (LGN L^4). *Prouver la loi forte des grands nombres dans le cas où les X_n ont un moment d'ordre 4.*

Proof. On peut supposer que les v.a. sont centrées. Ainsi,

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{n}(X_1 + \dots, X_n)\right)^4\right] &= \frac{1}{n^4} \sum_{1 \leq i_1, \dots, i_4 \leq n} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] \\ &= \frac{1}{n^4} (n\mathbb{E}[X_1^4] + 3n(n-1)\mathbb{E}[X_1^2 X_2^2]) \\ &\leq \frac{C}{n^4}. \end{aligned}$$

Ainsi $\sum \mathbb{E}[(\frac{1}{n}(X_1 + \dots + X_n))^4] < \infty$ et par Fubini-Tonelli, $\sum (\frac{1}{n}(X_1 + \dots + X_n))^4 < \infty$ p.s. ce qui entraîne le résultat. $\square$

Notons que les hypothèses pour la loi faible peuvent être allégées ou changées. On énonce maintenant la loi forte des grands nombres, dont la démonstration est compliquée (cf par exemple Chabanol-Ruch).

Theorem 3.3 (Loi forte des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. Alors $\frac{1}{n} \sum_{k=1}^n X_k$ converge p.s. si et seulement si $\mathbb{E}[|X_1|] < +\infty$. Si l'une des deux conditions est vérifiée alors la limite est $\mathbb{E}[X_1]$.*

Un corollaire immédiat est le suivant, qui illustre par exemple le fait que si on répète un lancer de dé non pipé un grand nombre de fois, la fréquence d'apparition du 1 (ou d'une autre face), se rapproche de $\frac{1}{6}$.

Corollary 3.4. *Si (A_n) est une suite d'événements indépendants de même probabilité, alors $\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{A_k}$ converge p.s. vers $\mathbb{P}(A_1)$.*

Le théorème de Glivenko-Cantelli est une conséquence du corollaire précédent. Il est notamment utile pour illustrer une convergence en loi : il affirme que la fonction de répartition empirique d'un échantillon i.i.d. de loi donnée converge vers la fonction de répartition théorique de la loi en question. Rappelons la définition de la fonction de répartition empirique.

Definition 3.5. *Soit (X_n) une suite de v.a. i.i.d. La fonction de répartition empirique du vecteur aléatoire $(X_1, \dots, X_n)$ est la fonction (aléatoire) définie sur $\mathbb{R}$ par*

$$F_n(x) = \sum_{k=1}^n \mathbb{1}_{X_k \leq x}.$$

Theorem 3.6 (Théorème de Glivenko-Cantelli). *Soit (X_n) une suite de v.a. i.i.d. de même fonction de répartition F . Alors la fonction de répartition empirique F_n converge presque sûrement uniformément vers la fonction de répartition F :*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Pour la preuve, on se réfère au Chabanol-Ruch par exemple. On verra dans la dernière section comment implémenter en pratique ce théorème pour illustrer une convergence en loi par ordinateur.

4 Théorème central limite

Theorem 4.1 (TCL 1D). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles i.i.d. dans L^2 . On note $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \text{Var}[X_1]$. Alors on a la convergence en loi*

$$\frac{1}{\sigma\sqrt{n}} \left(\sum_{k=1}^n X_k - n\mu \right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Proof. Une idée de développement, voir le Queffélec-Zuily par exemple... $\square$

Ce théorème suggère que la moyenne empirique des variables aléatoires converge vers son espérance (c'est la loi des grands nombres) à vitesse $O(\sqrt{n})$. Pour la preuve on peut se référer par exemple au Queffélec-Zuily.

A noter qu'une version plus faible de ce résultat, où les X_n sont des variables de Bernoulli, est enseignée en Terminale S : c'est le théorème de Moivre-Laplace.

Corollary 4.2. *Soit (S_n) une suite de v.a. binomiales de paramètres n, p . Alors*

$$\frac{1}{\sqrt{np(1-p)}}(S_n - np) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

en loi.

Exercice 4.3. *Démontrer le théorème de Moivre-Laplace directement.*

Proof. On peut supposer que les variables sont centrées. On pose $Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$. Par indépendance des X_k , la fonction caractéristique de Z_n est

$$\Phi_{Z_n}(u) = \Phi_{X_1}(u/\sqrt{n})^n.$$

Comme X_1 est L^2 , Φ_{X_1} est deux fois dérivable en 0, et

$$\Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n}\right)$$

puis en passant à la limite,

$$\lim_{n \rightarrow \infty} \Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right)^n = \exp\left(-\frac{\sigma^2 u^2}{2}\right).$$

$\square$

Un corollaire immédiat du TCL est le résultat suivant, qui permet d'obtenir des intervalles de confiance asymptotiques pour l'espérance, comme ce sera expliqué dans la partie statistiques.

Corollary 4.4. *Sous les hypothèses du TCL, pour $a > 0$,*

$$\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-a}^a \exp\left(-\frac{x^2}{2}\right) dx$$

où $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$.

La version multidimensionnelle fait intervenir les vecteurs Gaussiens.

Theorem 4.5 (TCL multidimensionnel). *Soit (X_n) une suite de vecteurs aléatoires de $\mathbb{R}^d$ i.i.d. dans L^2 . On note $\mu = \mathbb{E}[X_1]$, Γ la matrice de covariance, et $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors la suite $(\sqrt{n}(\bar{X}_n - \mu))_n$ converge en loi vers un vecteur Gaussien de loi $\mathcal{N}(0, \Gamma)$.*

Ce théorème est utile en statistiques notamment,

5 Illustration des théorèmes de convergence

5.1 Convergence presque sûre

Pour illustrer la convergence p.s. (loi des grands nombres), on simule un nombre suffisant de v.a. i.i.d., et on calcule à chaque fois la moyenne empirique, qu'on trace (en abscisses le nombre n de v.a. simulées et en ordonnées la valeur de $\frac{X_1 + \dots + X_n}{n}$). On effectue plusieurs simulations de la suite. Voici un exemple de code pour des v.a. uniformes sur $[0, 1]$ i.i.d.

```

# Paramètres
Nsimu = 5 # Nombre de simulations
n = 100 # Nombre de réalisations par simulations

# Réalisations des v.a. et calcul de la moyenne empirique
X = rd.random((Nsimu, n))
k = range(1,n+1)
X_bar = np.cumsum(X, axis=1)/k

plt.figure()

# Tracé graphique
plt.title("Loi des grands nombres")
plt.plot(k, X_bar.T)

# Une ligne pour la limite
plt.hlines(0.5, 1, n, label="Limite", color='k', linestyle='--')

# Des légendes
plt.xlabel("k")
plt.ylabel("X")

# Une grille en pointillés
plt.grid()

# Taille des axes
plt.axis([0, n, 0, 1]) # définit la taille des axes [xmin, xmax, ymin, ymax]

# Une légende avec les labels des tracés si besoin
plt.legend()

plt.show()

```

5.2 Convergence en loi

Il y a plusieurs façons d'illustrer la convergence en loi (TCL). A chaque fois, il s'agit de comparer l'échantillon obtenu sur un grand nombre de simulations avec le résultat théorique. Voici trois exemples différents.

5.2.1 Comparaison des fonctions de répartition

```

# Paramètres
Nsimu = 1000
n = 100

# Réalisations des v.a.
X = rd.random((Nsimu, n))
std = 1/np.sqrt(12)
X_bar = np.sum(X, axis=1)/n
Z = np.sqrt(n)*(X_bar-1/2)/std # Variables centrées réduites

# Calcul des fréquences cumulées croissantes
table = np.unique(Z, return_counts=True)

```

```

frequence_cum = np.cumsum(table[1])/Nsimu

# Tracé fdr empirique/fdr théorique
plt.figure()
plt.title("Théorème central limite")
plt.step(table[0], frequence_cum, label="Empirique")
plt.plot(table[0], st.norm.cdf(table[0]), label="Théorique", color='k', linestyle='--')
plt.xlabel("Z")
plt.ylabel("Probabilité")
plt.legend()
plt.show()

```

5.2.2 Histogramme

```

# Paramètres
Nsimu = 1000
n = 100

# Les réalisations
X = rd.random(Nsimu, n)
std = 1/np.sqrt(12)
X_bar = np.sum(X, axis=1)/n
Z = np.sqrt(n)*(X_bar-1/2)/std

# Valeurs en abscisses
x = np.linspace(min(Z), max(Z), num=1000)

# Tracé histogramme des valeurs/densité théorique
plt.figure()
plt.title("Théorème central limite")
plt.hist(Z, density=True, label="Empirique")
plt.plot(x, st.norm.pdf(x), label="Théorique", color='k', linestyle='--')
plt.xlabel("Z")
plt.ylabel("Probabilité")
plt.legend()
plt.show()

```

5.2.3 Diagramme en bâtons

```

# Paramètres
Nsimu = 10000
n = 10

# Réalisations de v.a. de Bernoulli
X = rd.binomial(n, 1/2, size=Nsimu)
table = np.unique(X, return_counts=True)
valeur = table[0]
frequence = table[1]/Nsimu

# Tracé fréquences empiriques/théoriques
plt.figure()
plt.title("Comparaison fréquences empiriques/théoriques")
plt.bar(valeur, frequence, label="Fréquence empirique") # Diagramme en bâtons

```

```
plt.vlines(valeur, [0]*(n+1), st.binom(n, 1/2).pmf(valeur),  
           color='black', linestyle='--', label="Probabilité théorique")  
plt.legend(loc="lower right")  
plt.xlabel("Position")  
plt.ylabel('Fréquence')  
plt.show()
```