Convergences

Djalil Chafaï*

14 février 2013

Ces notes rassemblent sous forme condensée des notions d'intégration probabiliste [1]. Soient $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, X, Y des v.a.r. définies sur un même espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$, de lois respectives μ_n, ν_n, μ, ν, et de fonction de répartitions respectives F_n, G_n, F, G.

Convergence presque sûre. On dit que $X_n \xrightarrow{\text{p.s.}} X$ lorsque

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$$

ce qui signifie $\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$. C'est la notion de convergence qui apparaît dans la loi forte des grands nombres.

Convergence en probabilité. On dit que $X_n \xrightarrow{\mathbb{P}} X$ lorsque

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

ce qui signifie $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$. C'est la notion de convergence qui apparaît dans la loi faible des grands nombres.

Convergence en moyenne. Soit $p \in [1, \infty[$. On dit que $X_n \xrightarrow{\mathrm{L}^p} X$ lorsque

$$X \in \mathrm{L}^p \quad \text{et} \quad \lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0,$$

qui signifie que $X_n \to X$ dans L^p. Les cas les plus utiles sont $p = 1$ et $p = 2$ (et $p = \infty$? !).

Convergence en loi. Les propriétés suivantes sont équivalentes et on dit alors que $X_n \xrightarrow{\text{loi}} X$, ou encore $X_n \xrightarrow{\text{loi}} \mu$, ou encore $\mu_n \xrightarrow{\text{étr.}} \mu$ (convergence étroite). C'est la notion de convergence qui apparaît dans le théorème central limite.

1. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f : \mathbb{R} \to \mathbb{R}$ continue et bornée ;
2. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f : \mathbb{R} \to \mathbb{R}$ $\mathcal{C}^\infty$ et à support compact ;
3. **(sur $\mathbb{Z}$)** $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f = \mathbf{1}_{\{x\}}$, $x \in \mathbb{Z}$,
   c'est-à-dire que $\mathbb{P}(X_n = x) \to \mathbb{P}(X = x)$ pour tout $x \in \mathbb{Z}$;
4. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f = \mathbf{1}_{]-\infty,x]}$ avec x point de continuité de F,
   c'est-à-dire que $F_n(x) \to F(x)$ dès que F est continue en x ;
5. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f = e^{it\bullet}$, $t \in \mathbb{R}$;
6. **(sur $\mathbb{R}_+$)** $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f = e^{-t\bullet}$, $t \geq 0$;
7. **(sur $\mathbb{N}$)** $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ pour toute $f = s^\bullet = e^{\ln(s)\bullet}$, $s \in]0, 1[$.

---

*`http://djalil.chafai.net/`

1. Il s'agit ici de l'intégrale de Lebesgue, adossée à la théorie de la mesure. Cette théorie permet d'englober les lois et les variables aléatoires discrètes et continues dans le même formalisme cohérent, puissant, et efficace. Malheureusement, le programme de l'agrégation a du retard en la matière !

Contrairement aux autres modes de convergence, la convergence en loi ne dépend pas de la loi des couples (X_n, X), et ne fait intervenir que les lois marginales : la loi de X_n (pour tout n) et la loi de X. Les quatre dernières propriétés expriment respectivement une convergence ponctuelle des fonctions de répartition, des transformées de Fourier (ou fonctions caractéristiques), des transformées de Laplace, et des fonctions génératrices.

Sommes de variables indépendantes. Les transformées de Fourier, de Laplace, et les fonctions génératrices sont particulièrement utiles pour établir la convergence en loi de sommes de variables aléatoires indépendantes : l'exponentielle transforme somme en produit, puis l'indépendance transforme espérance de produit en produit d'espérances.

Relations entre les convergences. L'inégalité de Hölder permet d'établir que la convergence en moyenne d'ordre $q \geq 1$ implique la convergence en moyenne d'ordre $1 \leq p \leq q$ car en notant $1/r = 1/p - 1/q$:

$$\mathbb{E}(|X_n - X|^p)^{1/p} \leq \mathbb{E}(|X_n - X|^q)^{1/q}\mathbb{E}(1^r)^{1/r} = \mathbb{E}(|X_n - X|^q)^{1/q}.$$

L'inégalité de Markov permet d'établir que la convergence en moyenne d'ordre $p \geq 1$ entraîne la convergence en probabilités :

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|^p)}{\varepsilon^p}.$$

On passe de la convergence presque sûre à la convergence en probabilité en traduisant tout d'abord les quantificateurs $\forall$ et $\exists$ par $\cap$ et $\cup$ comme suit :

$$\{\lim_{n\to\infty} X_n = X\} = \cap_{k=1}^{\infty} \cup_{m=1}^{\infty} \cap_{n\geq m}\{|X_n - X| \leq 1/k\},$$

puis en observant que par monotonie, pour tout k fixé :

$$\begin{aligned}
1 &= \mathbb{P}(\lim_{n\to\infty} X_n = X) \\
&\leq \mathbb{P}(\cup_{m=1}^{\infty} \cap_{n\geq m} \{|X_n - X| \leq 1/k\}) \\
&\to \lim_{m\to\infty} \mathbb{P}(\cap_{n\geq m}\{|X_n - X| \leq 1/k\}) \\
&\leq \varliminf_{n\to\infty} \mathbb{P}(|X_n - X| \leq 1/k).
\end{aligned}$$

Si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\text{loi}} X$ car pour toute fonction f continue et périodique ou à support compact, la fonction f est uniformément continue par le théorème de Heine, et donc pour tout $\eta > 0$, il existe $\varepsilon > 0$ tel que $|f(x) - f(y)| \leq \eta$ si $|x - y| \leq \varepsilon$, d'où

$$\begin{aligned}
|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq \mathbb{E}(|f(X_n) - f(X)|) \\
&= \mathbb{E}(|f(X_n) - f(X)|\mathbf{1}_{|X_n - X|<\varepsilon}) + \mathbb{E}(|f(X_n) - f(X)|\mathbf{1}_{|X_n - X|\geq\varepsilon}) \\
&\leq \eta + 2\,\|f\|_\infty\, \mathbb{P}(|X_n - X| \geq \varepsilon).
\end{aligned}$$

On peut donc choisir η assez petit, ce qui donne ε, puis n assez grand.

$$\begin{array}{ccccc}
 & & \boxed{\text{CV en moyenne } \mathrm{L}^p} & & \\
 & & \Downarrow & & \\
 & & \boxed{\text{CV en moyenne } \mathrm{L}^1} & & \\
 & & \Downarrow & & \\
\boxed{\text{CV presque sûre}} & \Rightarrow & \boxed{\text{CV en probabilité}} & \Rightarrow & \boxed{\text{CV en loi}}
\end{array}$$

Si X est constante (et égale à c) alors la convergence en loi entraîne la convergence en probabilité : si $f_{c,\varepsilon} : \mathbb{R} \mapsto \mathbb{R}_+$ est continue bornée avec $f(c) = 1$ et $\mathbf{1}_{]c-\varepsilon,c+\varepsilon[} \geq f_{c,\varepsilon}$ alors

$$\mathbb{P}(|X_n - X| < \varepsilon) = \mathbb{E}(\mathbf{1}_{]c-\varepsilon,c+\varepsilon[}(X_n)) \geq \mathbb{E}(f_{c,\varepsilon}(X_n)) \to \mathbb{E}(f_{c,\varepsilon}(c)) = 1.$$

Si $\sup_n |X_n|$ est intégrable alors la convergence presque sûre entraîne la convergence en moyenne (par le théorème de convergence dominée). Le théorème de convergence dominée permet également de déduire la convergence en loi de la convergence presque sûre sans passer par la convergence en probabilité. Enfin, au delà de la convergence dominée, une suite converge en moyenne d'ordre 1 ssi elle est uniformément intégrable (UI) et converge en probabilité.

Contrexemples.

1. *Suite convergeant en probabilité mais pas presque sûrement :* Il suffit de considérer une suite tournante sur $[0, 1]$. Posons $X_{m+k} = \mathbf{1}_{[k,k+1[/m}(U)$ pour $0 \leq k < m$ et $m \geq 1$, où U est uniforme sur $[0, 1]$. On a alors $X_n \xrightarrow{\mathbb{P}} 0$ mais $(X_n)_{n\geq 1}$ ne converge pas p.s. car « tous les ω voient passer des 0 et des 1 une infinité de fois » ;
2. *Suite convergeant en loi mais pas en probabilités.* Soit X_n uniforme sur $[0, 1]$ pour tout $n \geq 1$, et X uniforme sur $[0, 1]$ et indépendante de $(X_n)_{n\geq 1}$. On a $X_n \xrightarrow{\text{loi}} X$ car $\mathcal{L}(X_n) = \mathcal{L}(X)$ pour tout $n \geq 1$. Mais si $\varepsilon > 0$ alors $\mathbb{P}(|X_n - X| \geq \varepsilon)$ est non nul et ne dépend pas de n, et donc $(X_n)_{n\geq 1}$ ne converge pas en probabilité vers X ;
3. *Suite convergeant en probabilité mais pas en moyenne.* On pose $X_n = n\mathbf{1}_{[0,1/n]}(U)$ où U est une v.a.r. uniforme sur $[0, 1]$. On a $\mathbb{P}(|X_n - 0| \geq \varepsilon) = \mathbb{P}(U \leq 1/n) = 1/n \to 0$ pour tout $\varepsilon > 0$ et donc $X_n \xrightarrow{\mathbb{P}} 0$. Cependant, $\mathbb{E}(|X_n - 0|) = n \int_0^1 \mathbf{1}_{[0,1/n]}(u)\, du = 1 \not\to 0$ et donc $(X_n)_{n\geq 1}$ ne converge pas en moyenne vers 0.

Stabilités des convergences. Les convergences presque sûre, en probabilité, et en moyenne sont stables par combinaisons linéaires finies. Les convergences presque sûres, en probabilité et en loi sont stables par composition avec une fonction continue.

Extension aux vecteurs aléatoires. Les trois types de convergence s'étendent naturellement aux vecteurs aléatoires. Pour la convergence en probabilité ou en moyenne, on peut choisir une distance ou une norme. Pour la convergence en loi, on utilise un produit scalaire dans la transformée de Fourier (fonction caractéristique) : itX devient $i\langle t, X\rangle$.

Théorème de convergence monotone. Si $(X_n)_{n\geq 1}$ est positive et croissante alors

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

dans $\mathbb{R}_+ \cup \{+\infty\}$, et en particulier, $\mathbb{E}(\lim_{n\to\infty} X_n) < \infty$ ssi $\lim_{n\to\infty} \mathbb{E}(X_n) < \infty$. On utilise souvent ce lemme avec le fait suivant : si $X \geq 0$ alors $\mathbb{E}(X) < \infty$ entraîne $\mathbb{P}(X < \infty) = 1$. On en déduit ainsi une preuve très sympathique du premier lemme de Borel-Cantelli :

$$\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E}\mathbf{1}_{A_n} = \mathbb{E}\sum_n \mathbf{1}_{A_n} \quad \text{et} \quad \left\{\sum_n \mathbf{1}_{A_n} = \infty\right\} = \overline{\lim_n} A_n.$$

La même méthode permet d'établir très rapidement une loi forte des grands nombres pour des variables aléatoires indépendantes et bornées dans L^4 (pas forcément de même loi). En effet, si $S_n := n^{-1}(X_1 + \cdots + X_n)$ et $\sup_n \mathbb{E}(X_n^4) < \infty$ alors $\mathbb{E}(S_n^4) = \mathcal{O}(n^{-2})$ (développer la puissance 4), d'où $\mathbb{E}(\sum_n S_n^4) = \sum_n \mathbb{E}(S_n^4) < \infty$, donc $\sum_n S_n^4 < \infty$ p.s. et ainsi $S_n \to 0$ p.s. !

Lemme de Fatou. Si $(X_n)_{n\geq 1}$ est positive alors

$$\underline{\lim}_{n\to\infty} \mathbb{E}(X_n) \geq \mathbb{E}(\underline{\lim}_{n\to\infty} X_n).$$

Ce lemme est utile pour les suites qui sont ni monotones ni bornées. Pour se souvenir de du sens de l'inégalité, retenir que tout se passe comme si $\underline{\lim}_n$ était concave.

Théorème de convergence dominée. Si $X_n \xrightarrow{\text{p.s.}} X$ et $\sup_n |X_n| \leq Y$, $\mathbb{E}(Y) < \infty$, alors

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

La condition de domination est un critère d'intégrabilité uniforme facile à vérifier.

Lemme de Scheffé. Si $X_n, X \in \mathrm{L}^1$, $X_n \xrightarrow{\text{p.s.}} X$ alors $X_n \xrightarrow{\mathrm{L}^1} X$ ssi $\mathbb{E}(|X_n|) \to \mathbb{E}(|X|)$.

Lemme de Slutsky. Si $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\text{loi}} Y$ et Y constante alors $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$. En particulier, $X_nY_n \xrightarrow{\text{loi}} XY$; $X_n + Y_n \xrightarrow{\text{loi}} X + Y$; $X_n/Y_n \xrightarrow{\text{loi}} X/Y$ si $Y \neq 0$. Le lemme de Slutsky permet notamment de remplacer une moyenne ou une variance par un estimateur empirique dans un théorème de convergence en loi du même type que le théorème central limite, lié par exemple à un estimateur, ce qui s'avère pratique pour fabriquer une région de confiance ou un test d'hypothèse statistique. Pour établir le lemme de Slutsky, on observe que comme Y est constante, on a pour tous $s, t \in \mathbb{R}$ et tout $n \geq 1$,

$$\mathbb{E}(e^{itX_n+isY_n}) - \mathbb{E}(e^{itX+isY}) = \mathbb{E}(e^{isY})(\mathbb{E}(e^{itX_n+is(Y_n-Y)}) - \mathbb{E}(e^{itX})),$$

mais aussi que $Y_n \xrightarrow{\mathbb{P}} Y$, et on utilise le fait que $X_n \xrightarrow{\text{loi}} X$ et une continuité uniforme.

Intégrabilité uniforme ou équi-intégrabilité. Pour toute famille de variables aléatoires $(X_i)_{i\in I}$ finie ou infinie, dénombrable ou pas, les propriétés suivantes sont équivalentes et on dit dans ce cas que la famille est *uniformément intégrable* (UI) :

1. $\lim_{R\to+\infty} \sup_{i\in I} \mathbb{E}(|X_i|\mathbf{1}_{\{|X_i|\geq R\}}) = 0$;
2. [2] $\forall\varepsilon > 0$, $\exists\delta > 0$, $\forall A \in \mathcal{A}$, $\mathbb{P}(A) \leq \delta \Rightarrow \sup_{i\in I} \mathbb{E}(|X_i|\mathbf{1}_A) \leq \varepsilon$;
3. [3] il existe une fonction $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ positive croissante et convexe telle que

$$\lim_{x\to+\infty} \frac{\varphi(x)}{x} = +\infty \quad \text{et} \quad \sup_{i\in I} \mathbb{E}(\varphi(|X_i|)) < \infty.$$

Une famille UI ne peut être constituée que de v.a.r. intégrables. Une famille finie de v.a.r. est UI ssi les v.a.r. qui la constituent sont toutes intégrables. En vertu du troisième critère, une famille bornée dans L^p avec $p > 1$ est toujours UI, car il suffit de prendre $\varphi(x) = |x|^p$. Enfin, une famille dominée dans L^1 (au sens où $\sup_{i\in I} |X_i| \leq Y$ avec $\mathbb{E}(Y) < \infty$) est toujours UI. Grâce au troisième critère, domination dans L^1 entraîne bornitude dans « L^{1+} » ! En particulier, si une v.a.r. X est intégrable alors $\varphi(X)$ est intégrable pour une fonction φ surlinéaire qui dépend toutefois de X. Pour comprendre ce phénomène, il suffit de penser aux séries de Riemann : si $\sum_n (1/n^\alpha) < \infty$ alors $\sum_n (1/n^{\alpha-\varepsilon}) < \infty$ dès que $0 < \varepsilon < \alpha - 1$, car la condition de convergence des séries de Riemann est « ouverte » plutôt que « fermée ».

2. Critère « $\varepsilon - \delta$ ».
3. Critère de Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin. Ceci n'est pas un canular.