Groupe de lecture :
Convergence étroite et convergence en loi

Ulysse Herbach et Lucas Da Ros

5 mars 2013

Diverses caractérisations de la convergence étroite des mesures de probabilité ont été obtenues dans un cadre assez général. Le but de ce groupe de lecture est de formuler, à partir de ces caractérisations, quelques résultats purement probabilistes sur la convergence en loi des variables aléatoires.

Référence

P. Billingsley, *Convergence of Probability Measures*, Wiley (1999), p. 20-32.

Sommaire

1 Rappels sur la convergence étroite

1.1 Définitions

Définition 1.1 (Convergence étroite)**.** *Soient $(P_n)_{n\in\mathbb{N}}$ et P des mesures de probabilité définies sur un même espace topologique $(E,\mathcal{T})$ (i.e. sur sa tribu borélienne). On note $\mathcal{C}_b(E)$ l'ensemble des fonctions continues bornées à valeurs réelles sur E. On dit que la suite $(P_n)_{n\in\mathbb{N}}$ **converge étroitement** vers P, et on note $P_n \rightharpoonup P$, lorsque*

$$\forall \phi \in \mathcal{C}_b(E), \quad \int_E \phi\, dP_n \xrightarrow[n\to\infty]{} \int_E \phi\, dP.$$

Définition 1.2 (Mesure-image)**.** *Soient $(E,\mathcal{A})$ et $(F,\mathcal{B})$ deux espaces mesurables, $f : E \to F$ une application mesurable et μ une mesure positive sur $\mathcal{A}$. La **mesure-image** de μ par f, notée $\mu\circ f^{-1}$ est la mesure positive définie sur $\mathcal{B}$ par :*

$$\mu\circ f^{-1}(B) = \mu(f^{-1}(B)) = \mu(\{x\in E : f(x)\in B\}).$$

*Lorsque X est une variable aléatoire (v.a.) définie sur un espace de probabilité $(\Omega,\mathcal{A},\mathbb{P})$ et à valeurs dans F, la mesure-image de $\mathbb{P}$ par X est une mesure de probabilité : on l'appelle **loi** de X et on la note souvent P_X. On utilisera alors indifféremment les notations :*

$$P_X(B) = \mathbb{P}\circ X^{-1}(B) = \mathbb{P}(\{\omega\in\Omega : X(\omega)\in B\}) = \mathbb{P}(X\in B).$$

Définition 1.3 (Ensemble de μ-continuité)**.** *Soit $(E,\mathcal{T})$ un espace topologique, $\mathcal{A}$ sa tribu borélienne et μ une mesure positive sur $\mathcal{A}$. On dit qu'un borélien $B\in\mathcal{A}$ est un **ensemble de μ-continuité** si $\mu(\partial B) = 0$, où ∂B désigne la frontière de B. Si X est une v.a., on dit que B est de X-continuité si B est de P_X-continuité où P_X est la loi de X.*

1.2 Caractérisations de la convergence étroite

On rappelle deux résultats qui ont été établis lors du groupe de lecture précédent. Il s'agit de diverses caractérisations de la convergence étroite qui nous seront utiles pour la suite.

Théorème 1.1 (Portmanteau)**.** *Soient $(P_n)_{n\in\mathbb{N}}$ et P des mesures de probabilité définies sur un même espace métrique (S,d). Les assertions suivantes sont équivalentes :*

(i) $P_n \rightharpoonup P$;

(ii) $\limsup_n P_n(F) \le P(F)$ *pour tout fermé F de S* ;

(iii) $\liminf_n P_n(U) \ge P(U)$ *pour tout ouvert U de S* ;

(iv) $\lim_n P_n(B) = P(B)$ *pour tout ensemble de P-continuité $B\subset S$.*

Théorème 1.2 (Helly-Bray)**.** *Soient $(P_n)_{n\in\mathbb{N}}$ et P des mesures de probabilité définies sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ et $(F_n)_{n\in\mathbb{N}}$, F leurs fonctions de répartition respectives. Alors*

$$(P_n \rightharpoonup P) \iff (F_n(t)\to F(t) \text{ en tout point } t \text{ où } F \text{ est continue}).$$

1.3 Composition par une fonction mesurable

On considère à présent (S,d) et (S',d') deux espaces métriques et on note $\mathcal{S}$ et $\mathcal{S}'$ leurs tribus boréliennes respectives. Soient $(P_n)_{n\in\mathbb{N}}$ et P des mesures de probabilité sur $\mathcal{S}$ et $h : (S,\mathcal{S}) \to (S',\mathcal{S}')$ une application mesurable. On note D_h l'ensemble des points de discontinuité de h ($D_h \in \mathcal{S}$, cf. remarque ci-dessous). On a alors le théorème de "transfert" (*Mapping Theorem*) suivant :

Théorème 1.3. *Si $P_n \rightharpoonup P$ et $P(D_h) = 0$, alors $P_n \circ h^{-1} \rightharpoonup P \circ h^{-1}$.*

Preuve. Montrons dans un premier temps que ${D_h}^c \cap \overline{h^{-1}(F)} \subset h^{-1}(\overline{F})$ pour tout $F \subset S'$. Soit $x \in {D_h}^c \cap \overline{h^{-1}(F)}$: il existe (x_n) une suite dans $h^{-1}(F)$ telle que $x_n \to x$. Comme h est continue en x, on a $h(x_n) \to h(x)$. Or pour tout $n \in \mathbb{N}$, $h(x_n) \in F$, donc $h(x) \in \overline{F}$, d'où $x \in h^{-1}(\overline{F})$.

On applique ensuite deux fois le point (ii) du théorème 1.1 : une fois pour passer de la convergence en loi à la lim sup, et une fois pour le contraire. Soit F un fermé de S', on a

$$\begin{aligned}
\limsup_n P_n(h^{-1}(F)) &\leq \limsup_n P_n(\overline{h^{-1}(F)}) \quad \left(\text{car } h^{-1}(F) \subset \overline{h^{-1}(F)}\right) \\
&\leq P(\overline{h^{-1}(F)}) \\
&= P({D_h}^c \cap \overline{h^{-1}(F)}) \quad (\text{car } P({D_h}^c) = 1) \\
&\leq P(h^{-1}(\overline{F})) \\
&= P(h^{-1}(F)).
\end{aligned}$$

□

Remarque. Pour toute fonction $h : S \to S'$, l'ensemble D_h des points de discontinuité de h est dans $\mathcal{S}$. En effet, on sait grâce à l'étude de la notion d'*oscillation d'une fonction* que l'ensemble C_h des points de continuité de h est une intersection dénombrable d'ouverts. Ainsi, par définition de la tribu borélienne $\mathcal{S}$, on a $C_h \in \mathcal{S}$ puis enfin $D_h = {C_h}^c \in \mathcal{S}$.

2 Convergence en loi : traduction

2.1 Notations

Il s'agit de traduire les différentes caractérisations de la convergence étroite en termes de lois. Dans toute la suite, on considèrera :

- (S,d) et (S',d') deux espaces métriques ;
- $\mathcal{S}$ et $\mathcal{S}'$ leurs tribus boréliennes respectives ;
- X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans S ;
- $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. respectivement définies sur $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ et à valeurs dans S ;
- $(P_{X_n})_{n\in\mathbb{N}}$ et P_X les lois respectives de $(X_n)_{n\in\mathbb{N}}$ et X.

2.2 Définition

Définition 2.1. *On dit que la suite* (X_n) ***converge en loi*** *vers* X *et on note* $X_n \xrightarrow{\mathcal{L}} X$*, lorsque* $P_{X_n} \rightharpoonup P_X$*.*

Remarque. *Bien qu'elle ne soit qu'une autre manière de formuler la convergence étroite, la convergence en loi permettra souvent une écriture plus concise et plus intuitive.*

2.3 Traduction des théorèmes précédents

Les deux théorèmes suivants sont une simple transcription des théorèmes 1.1 et 1.2 dans le vocabulaire des variables aléatoires.

Théorème 2.1. *Les assertions suivantes sont équivalentes :*

(i) $X_n \xrightarrow{\mathcal{L}} X$;

(ii) $\limsup_n \mathbb{P}_n(X_n \in F) \leq \mathbb{P}(X \in F)$ *pour tout fermé* $F \subset S$;

(iii) $\liminf_n \mathbb{P}_n(X_n \in U) \geq \mathbb{P}(X \in U)$ *pour tout ouvert* $U \subset S$;

(iv) $\mathbb{P}_n(X_n \in B) \longrightarrow \mathbb{P}(X \in B)$ *pour tout ensemble de* X*-continuité* $B \subset S$*.*

Théorème 2.2. *Soient* $(X_n)_{n\in\mathbb{N}}$ *et* X *des v.a.r et* $(F_n)_{n\in\mathbb{N}}$, F *leurs fonctions de répartition respectives. Alors*

$$\left(X_n \xrightarrow{\mathcal{L}} X\right) \iff (F_n(t) \to F(t) \text{ en tout point } t \text{ où } F \text{ est continue}).$$

On a également la propriété suivante, utile pour la suite :

Propriété 2.1. *Si* $h : (S, \mathcal{S}) \to (S', \mathcal{S}')$ *est une application mesurable, alors* $h(X)$ *est une variable aléatoire de loi* $P_X \circ h^{-1}$*.*

Preuve. Les applications h et X sont mesurables, donc $h(X) := h \circ X$ est une application mesurable $\Omega \to S'$. De plus, pour tout $B \in \mathcal{S}'$,

$$\begin{aligned}
\mathbb{P}(h(X) \in B) &= \mathbb{P}((h \circ X)^{-1}(B)) \\
&= \mathbb{P}(\{w \in \Omega : h(X(w)) \in B\}) \\
&= \mathbb{P}(\{w \in \Omega : X(w) \in h^{-1}(B)\}) \\
&= P_X(h^{-1}(B))
\end{aligned}$$

donc la loi de $h(X)$ est $P_X \circ h^{-1}$. □

On a alors une traduction immédiate du *Mapping Theorem* (1.3) :

Théorème 2.3. *Si* $X_n \xrightarrow{\mathcal{L}} X$ *et* $\mathbb{P}(X \in D_h) = 0$*, alors* $h(X_n) \xrightarrow{\mathcal{L}} h(X)$*.*

3 Résultats de convergence en loi

On va maintenant donner quelques résultats de convergence en loi qui utilisent les théorèmes précédents. Il convient de noter que si (X, Y) est une v.a. à valeurs dans $S \times S$, alors par continuité $d(X, Y)$ est également une v.a., définie par $d(X, Y)(\omega) = d(X(\omega), Y(\omega))$, $\forall \omega \in \Omega$.

3.1 Utilisation de la métrique

On utilisera dans la suite le résultat bien connu suivant : si $X_n \xrightarrow{\mathcal{L}} a$ où a est une constante, alors (X_n) converge en probabilité vers a, c'est-à-dire $\mathbb{P}(d(X_n, a) > \varepsilon) \xrightarrow[n\to\infty]{} 0$, $\forall \varepsilon > 0$.

Théorème 3.1. *Soient $((X_n, Y_n))_{n\in\mathbb{N}}$ une suite de v.a. dans $S \times S$ et X une v.a. dans S. Si $X_n \xrightarrow{\mathcal{L}} X$ et $d(X_n, Y_n) \xrightarrow{\mathcal{L}} 0$, alors $Y_n \xrightarrow{\mathcal{L}} X$.*

Preuve. On va encore utiliser deux fois le point (ii) du théorème 2.1. Soient F un fermé de S et $\varepsilon > 0$, on pose $F_\varepsilon = \{x \in S : d(x, F) \leq \varepsilon\}$.

F_ε est fermé dans S : si $F_\varepsilon \ni x_n \to x \in S$ (i.e. $d(x_n, x) \to 0$), alors $\forall n \in \mathbb{N}$, $d(x_n, F) \leq \varepsilon$ donc

$$\begin{aligned} d(x, F) &= \inf_{y\in F}(d(y, x)) \\ &\leq \inf_{y\in F}(d(y, x_n)) + d(x_n, x) \\ &\leq \underbrace{\varepsilon + d(x_n, x)}_{\xrightarrow[n\to\infty]{} 0} \end{aligned}$$

d'où $d(x, F) \leq \varepsilon$ et $x \in F_\varepsilon$. On a ensuite

$$\begin{aligned} \mathbb{P}(Y_n \in F) &= \mathbb{P}(Y_n \in F,\, d(X_n, Y_n) \leq \varepsilon) + \mathbb{P}(Y_n \in F,\, d(X_n, Y_n) > \varepsilon) \\ &\leq \mathbb{P}(X_n \in F_\varepsilon) + \mathbb{P}(d(X_n, Y_n) > \varepsilon). \end{aligned}$$

Or $\mathbb{P}(d(X_n, Y_n) > \varepsilon) \xrightarrow[n\to\infty]{} 0$ (car $d(X_n, Y_n) \xrightarrow{\mathcal{L}} 0$), donc $\limsup_n \mathbb{P}(d(X_n, Y_n) > \varepsilon) = 0$, d'où

$$\limsup_n \mathbb{P}(Y_n \in F) \leq \limsup_n \mathbb{P}(X_n \in F_\varepsilon) \leq \mathbb{P}(X \in F_\varepsilon), \tag{3.1}$$

la dernière inégalité découlant du théorème 2.1.

Enfin, comme F est fermé on a $\lim_{\varepsilon\to 0} \mathbb{P}(X \in F_\varepsilon) = \mathbb{P}(\bigcap_{\varepsilon>0} \{X \in F_\varepsilon\}) = \mathbb{P}(X \in F)$, donc en faisant tendre ε vers 0 dans (3.1) on obtient

$$\limsup_n \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F)$$

d'où $Y_n \xrightarrow{\mathcal{L}} X$ par le théorème 2.1. □

Corollaire. *Supposons que $((X, Y_n))_{n\in\mathbb{N}}$ sont des v.a. à valeurs dans $S \times S$. Si $d(X, Y_n) \xrightarrow{\mathcal{L}} 0$, alors $Y_n \xrightarrow{\mathcal{L}} X$.*

Preuve. On applique le théorème précédent avec $\forall n \in \mathbb{N}$, $X_n = X$. □

Théorème 3.2. *Soient $(X_{k,n}, Y_n)_{k,n\in\mathbb{N}}$ des v.a. dans $S \times S$. On suppose qu'il existe X et $(Z_k)_{k\in\mathbb{N}}$ des v.a. dans S telles que*

$$\begin{cases} Z_k \xrightarrow[k\to\infty]{\mathcal{L}} X \\ \forall k \in \mathbb{N},\ X_{k,n} \xrightarrow[n\to\infty]{\mathcal{L}} Z_k \\ \forall \varepsilon > 0,\ \left(\limsup_n \mathbb{P}(d(X_{k,n}, Y_n) \geq \varepsilon)\right) \xrightarrow[k\to\infty]{} 0. \end{cases}$$

Alors $Y_n \xrightarrow{\mathcal{L}} X$.

3.2 Comportement asymptotique en moyenne

Lemme 1. *Soit X une v.a.r. positive. Alors*

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)dt.$$

Théorème 3.3. *Soient $(X_n)_{n\in\mathbb{N}}$ et X des v.a. à valeurs réelles. Si $X_n \xrightarrow{\mathcal{L}} X$, alors $\mathbb{E}(|X|) \leq \liminf_n \mathbb{E}(|X_n|)$.*

Définition 3.1. *Une suite $(X_n)_{n\in\mathbb{N}}$ est dite* ***uniformément intégrable*** *si*

$$\sup_n \Big(\mathbb{E}(|X_n|\mathbb{1}_{|X_n|\geq a})\Big) := \sup_n \left(\int_{|X_n|\geq a} |X_n| d\mathbb{P}_n\right) \xrightarrow[a\to\infty]{} 0.$$

Il est clair qu'une suite uniformément bornée est uniformément intégrable. De plus, si $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable, alors $\sup_n \mathbb{E}(|X_n|) < \infty$.

Si la convergence en loi des v. a. fait intervenir dans sa définition les fonctions continues bornées, on ne sait rien a priori de leur convergence en moyenne (cela revient à composer par l'identité qui n'est pas bornée). L'uniforme intégrabilité permet de trancher grâce au théorème suivant.

Théorème 3.4. *Soient $(X_n)_{n\in\mathbb{N}}$ et X des v.a. à valeurs réelles. On suppose que $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable et que $X_n \xrightarrow{\mathcal{L}} X$. Alors X est intégrable et $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

De plus, dans le cas de variables positives, on constate que la condition d'uniforme intégrabilité est nécessaire pour avoir convergence en moyenne.

Théorème 3.5. *Soient $(X_n)_{n\in\mathbb{N}}$ et X des v.a.r. positives et intégrables. Si $X_n \xrightarrow{\mathcal{L}} X$ et $\mathbb{E}(X_n) \to \mathbb{E}(X)$, alors $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.*