Rappels et compléments sur la convergence en loi

Lionel Truquet,
lionel.truquet@ensai.fr

1. Critères pour la convergence en loi

2. La méthode delta

Outline

Notations

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

- On considèrera des vecteurs de variables aléatoires $X = (X_1, \dots, X_k)$ de $\mathbb{R}^k$. On parle de vecteur aléatoire ou de variables aléatoire à valeurs dans $\mathbb{R}^k$.
- On notera d une distance sur $\mathbb{R}^k$, d sera équivalente à la distance euclidienne. La distance euclidienne est défini par $d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$.
- On rappelle qu'une suite de variable aléatoire $(X_n)_{n\geq 0}$ à valeurs dans $\mathbb{R}^k$ converge en probabilité vers une variable aléatoire X à valeurs dans $\mathbb{R}^k$ si pour tout $\epsilon > 0$,

  $$\lim_{n\to\infty} \mathbb{P}\left(d(X_n, X) > \epsilon\right) = 0.$$

  On notera $X_n \xrightarrow{P} X$.
- Une suite de variable aléatoire $(X_n)_{n\geq 0}$ à valeurs dans $\mathbb{R}^k$ converge presque sûrement (p.s.) vers une variable aléatoire X à valeurs dans $\mathbb{R}^k$ s'il existe $\Omega_1 \in \mathcal{A}$ tel que $\mathbb{P}(\Omega_1) = 1$ et

  $$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega_1.$$

Notations

- La convergence presque sûre entraine la convergence en probabilité mais le contraire n'est pas vrai. On peut montrer qu'une condition suffisante de convergence presque sûre est

$$\sum_{n\geq 0} \mathbb{P}\left(d(X_n, X) > \epsilon\right) < \infty, \quad \epsilon > 0.$$

- Pour $x, y \in \mathbb{R}^k$, on notera $x \leq y$ lorsque $x_i \leq y_i$ pour $1 \leq i \leq k$.

- Pour une variable aléatoire X à valeurs dans $\mathbb{R}^k$, la fonction définie sur $\mathbb{R}^k$, $x \mapsto \mathbb{P}\left(X \leq x\right)$ est appelée fonction de répartition.

Le lemme de Portmanteau

Lemma 1

Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires toutes à valeurs dans $\mathbb{R}^k$ et X une autre variables aléatoire à valeurs dans $\mathbb{R}^k$. Les conditions suivantes sont équivalentes.

1. $\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$ *en tout point de continuité de* $x \mapsto \mathbb{P}(X \leq x)$.
2. $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ *pour toute fonction* f *continue et bornée.*
3. $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ *pour toute fonction* f *Lipschitz et bornée.*
4. $\liminf_{n\to\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X)$ *pour toute fonction* f *positive et continue.*
5. $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ *pour tout ensemble ouvert* G.
6. $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *pour tout ensemble fermé* F.
7. $\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ *pour tout borélien* B *tel que* $\mathbb{P}(X \in \delta B) = 0$ *où* $\delta B = \overline{B} \setminus B^\circ$ *est la frontière de* B.

Definition 1

Lorsque l'une des conditions du lemme précédent est vérifiée, on dit que $(X_n)_{n\geq 0}$ *convergence en loi vers* X *et on note* $X_n \Rightarrow X$.

Théorème de l'application continue

Theorem 1

Soit $g : \mathbb{R}^k \to \mathbb{R}^m$ *continue en chaque point de* C *tel que* $\mathbb{P}(X \in C) = 1$. *On a alors les propriétés suivantes.*

1. *Si* $X_n \Rightarrow X$ *alors* $g(X_n) \Rightarrow g(X)$.
2. *Si* $X_n \xrightarrow{P} X$, *alors* $g(X_n) \xrightarrow{P} g(X)$.
3. *Si* $(X_n)_{n\geq 0}$ *converge vers* X *p.s., alors* $(g(X_n))_{n\geq 0}$ *converge* $g(X)$ *p.s.*

Tension et compacité

Definition 2

On dit qu'une suite de variables aléatoires $(X_n)_{n\geq 0}$ *est tendue si pour tout* $\epsilon > 0$*, il existe* $M > 0$ *tel que* $\sup_{n\geq 0} \mathbb{P}\left(d(X_n, 0) > M\right) \leq \epsilon$*.*

- Une suite constante est toujours tendue.
- Si la suite prend ses valeurs dans un sous-ensemble compact de $\mathbb{R}^d$ alors elle est tendue.

Theorem 2 (de Prohorov)

1. *Si* $X_n \Rightarrow X$ *alors* $(X_n)_{n\geq 0}$ *est tendue.*
2. *Si* $(X_n)_{n\geq 0}$ *est tendue alors il existe une sous-suite* $(X_{n_j})_{j\geq 0}$ *qui converge en loi.*

- Si on prend une suite de variables aléatoires à valeurs dans un même espace fini E, alors il existe une sous-suite qui converge en loi (démonstration directe ?).
- Le théorème ci-dessous montre que tout suite qui prend ses valeurs dans un sous-ensemble compact K de $\mathbb{R}^k$ admet automatiquement une sous-suite qui converge en loi.

Convergence en loi pour les couples

- On rappelle que la convergence p.s. entraîne la convergence en probabilité qui elle-même entraîne la convergence en loi.
- On notera cependant que la propriété de convergence en loi ne concerne que les distributions de probabilités. Souvent la variable aléatoire limite X n'est là que pour des raisons de commodité (pour représenter une loi de probabilité) et n'est pas nécessairement définie sur le même espace de probabilité.

Theorem 3

1. *Si $X_n \Rightarrow c$ où c est une constance alors $(X_n)_{n\geq 0}$ converge vers c en probabilité.*
2. *Si $X_n \Rightarrow X$ et $d(X_n, Y_n) \xrightarrow{P} 0$ alors $Y_n \Rightarrow X$.*
3. *Si $X_n \Rightarrow X$ et $Y_n \xrightarrow{P} c$ pour une constance c, alors $(X_n, Y_n) \Rightarrow (X, c)$.*

On prendra garde que $X_n \Rightarrow X$ et $Y_n \Rightarrow Y$ n'entraine pas $(X_n, Y_n) \Rightarrow (X, Y)$ en général (alors que c'est vrai pour la convergence en probabilité et la convergence p.s.).

Lemme de Slutsky et applications

Lemma 2

Si $X_n \Rightarrow X$ *et* $Y_n \Rightarrow c$ *où* c *est une constante, alors les propriétés suivantes sont vérifiées.*

1. $X_n + Y_n \Rightarrow X + c.$
2. $Y_n X_n \Rightarrow cX.$
3. $Y_n^{-1} X_n \Rightarrow c^{-1} X$ *provided* $c \neq 0.$

Les propriétés précédentes s'appliquent aussi bien à des vecteurs qu'à des matrices lorsque les sommes ou produits ont un sens.

Un exemple important d'application est la convergence d'une statistique après renormalisation par l'écart type estimé. Si $\sqrt{n}\,(T_n - \theta) \Rightarrow \mathcal{N}(0, \sigma^2)$ et $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ alors

$$\sqrt{n}\frac{T_n - \theta}{\hat{\sigma}} \Rightarrow \mathcal{N}(0, 1).$$

De plus si $q_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$ alors $[T_n - q_{1-\alpha/2}\hat{\sigma}/\sqrt{n}, T_n + q_{1-\alpha/2}\hat{\sigma}/\sqrt{n}]$ est un intervalle de confiance de niveau asymptotique $1 - \alpha$.

Convergence en loi et convergence des moments

$X_n \Rightarrow X$ et même $\lim_{n\to\infty} X_n = X$ p.s. ne guarantit pas que $\mathbb{E}X_n \to \mathbb{E}X$. Exemple : $X_n = 2^n Y_1 \cdots Y_n$ avec $(Y_n)_{n\geq 1}$ suite de v.a. i.i.d. toutes de loi $\mathcal{B}(1/2)$.

Definition 3

La suite $(X_n)_{n\in\mathbb{N}}$ est dite uniformément intégrable si

$$\lim_{M\to+\infty} \sup_{n\in\mathbb{N}} \mathbb{E}\left[\mathbb{1}_{\|X_n\|\geq M}\|X_n\|\right] = 0.$$

- On pourra vérifier que la suite donnée dans l'exemple ci-dessus n'est pas uniformément intégrable.
- Un critère qui guarantit l'uniforme intégrabilité est l'existence d'un nombre $p > 1$ tel que $\sup_{n\in\mathbb{N}} \mathbb{E}\|X_n\|^p < \infty$.

Convergence en loi et convergence des moments

Theorem 4

Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires telle que $X_n \Rightarrow X$. Soit $f : \mathbb{R}^k \to \mathbb{R}$ mesurable et continue en tout point de $C \in \mathcal{B}\left(\mathbb{R}^k\right)$ tel que $\mathbb{P}(X \in C) = 1$. Alors $\lim_{n\to+\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ dès lors que la suite $(f(X_n))_{n\geq 0}$ est uniformément intégrable.

Exemple d'application : si $X_n \Rightarrow X$ et $\sup_{n\geq 0} \mathbb{E}|X_n|^p < +\infty$ pour un réel $p > 0$, alors $\lim_{n\to+\infty} \mathbb{E}X_n^q = \mathbb{E}X^q$ pour tout $q < p$.

Theorem 5

Si $\mathbb{E}X_n^p \to \mathbb{E}X^p$ pour tout entier $p \geq 0$ et si la distribution de X est uniquement déterminée par ses moments alors $X_n \Rightarrow X$.

Par exemple, la loi $\mathcal{N}(0,1)$ est uniquement déterminée par ses moments $m_{2p+1} = 0$ et $m_{2p} = \frac{(2p)!}{2^p p!}$. La loi gaussienne standard est la seule loi de probabilité ayant ses moments.

o et O stochastique

- On note $X_n = o_P(1)$ si $X_n \xrightarrow{P} 0$ et $X_n = O_P(1)$ si $(X_n)_{n\geq 0}$ est tendue.
- On alors les règles suivantes : $o_P(1) + o_P(1) = o_P(1)$,

$$o_P(1) + O_P(1) = O_P(1), \quad O_P(1)o_P(1) = o_P(1), \quad (1 + o_P(1))^{-1} = O_P(1).$$

- Si R_n à valeurs réelles, on dit que

$$X_n = o_P(R_n) \text{ si } X_n = R_n Y_n \text{ et } Y_n = o_P(1),$$

$$X_n = O_P(R_n) \text{ si } X_n = R_n Y_n \text{ et } Y_n = O_P(1).$$

- On obtient de nouvelles règles

$$o_P(R_n) = R_n o_P(1), \quad O_P(R_n) = R_n O_P(1), \quad o_P\left(O_P(1)\right) = o_P(1).$$

o et O stochastique

A titre d'exercice, on pourra démontrer le résultat suivant

Lemma 3

Soient $R : \mathbb{R}^k \to \mathbb{R}$ une fonction telle que $R(0) = 0$ et $(X_n)_{n\geq 0}$ une suite telle que $X_n = o_P(1)$. Pour tout $p > 0$, on a

1. *si $R(h) = o\left(\|h\|^p\right)$ lorsque $h \to 0$, alors $R(X_n) = o_P\left(\|X_n\|^p\right)$;*
2. *si $R(h) = O\left(\|h\|^p\right)$ lorsque $h \to 0$, alors $R(X_n) = O_P\left(\|X_n\|^p\right)$.*

On pourra poser $g(h) = R(h)/\|h\|^p$ si $h \neq 0$ et $g(0) = 0$ et étudier la suite $g(X_n)$.

Fonctions caractéristiques

Pour une variable aléatoire X à valeurs dans $\mathbb{R}^k$, on définit sa fonction caractéristique ϕ_X par

$$\phi_X(t) = \mathbb{E}\left[e^{it'X}\right] = \mathbb{E}\left[e^{i\sum_{j=1}^k t_j X_j}\right].$$

Theorem 6 (de continuité de Lévy)

1. *On a l'équivalence*

$$X_n \Rightarrow X \Leftrightarrow \forall t \in \mathbb{R}^k, \quad \lim_{n\to+\infty} \phi_{X_n}(t) = \phi_X(t).$$

2. *De plus si ϕ_{X_n} converge point par point vers une fonction ϕ continue en 0 alors ϕ est forcément la fonction caractéristique d'une variable X et de plus $X_n \Rightarrow X$.*

Outline

Développement limité

- On suppose une convergence en loi du type $\sqrt{n}\,(T_n-\theta)\Rightarrow T$.
- On s'intéresse à une convergence en loi pour $\phi(T_n)$ (e.g. T_n^2) pour une fonction donnée ϕ.
- Lorsque ϕ est dérivable, on peut utiliser un développement limité :

$$\sqrt{n}\,(\phi(T_n)-\phi(\theta)) = \sqrt{n}\phi'(\theta)(T_n-\theta)+\sqrt{n}R_n.$$

- Si $\sqrt{n}R_n=o_{\mathbb{P}}(1)$, on peut déduire du lemme de Slutsky la convergence

$$\sqrt{n}\,(\phi(T_n)-\phi(\theta))\Rightarrow \phi'(\theta)T.$$

Résultat

- Si $\phi : \mathbb{R}^k \to \mathbb{R}$, on pose $D\phi(\theta) = (\partial_1\phi(\theta), \ldots, \partial_k\phi(\theta))$ (le vecteur ligne formé des dérivées partielles de ϕ au point θ).

- Si $\phi : \mathbb{R}^k \to \mathbb{R}^m$ avec $\phi = (\phi_1, \ldots, \phi_m)'$, on pose

$$D\phi(\theta) = \begin{pmatrix} \partial_1\phi_1(\theta) & \cdots & \partial_k\phi_1(\theta) \\ \vdots & \ddots & \vdots \\ \partial_1\phi_m(\theta) & \cdots & \partial_k\phi_m(\theta) \end{pmatrix}.$$

Theorem 7

Soit $\phi : O \subset \mathbb{R}^k \to \mathbb{R}^m$ une fonction différentiable au point θ. Soit également T_n une variable aléatoire à valeurs dans O et telle que $r_n(T_n - \theta) \Rightarrow T$ pour une suite $(r_n)_{n\geq 0}$ telle que $r_n \to +\infty$. Alors

$$r_n\left(\phi(T_n) - \phi(\theta)\right) \Rightarrow D\phi(\theta) \cdot T.$$

Exemples

- Si $X_1, X_2, \ldots$ i.i.d. et telles que $\mu_4 < +\infty$ et $\mu_1 = 0$ avec $\mu_i = \mathbb{E}X_1^i$ pour $i \in \mathbb{N}^*$. Alors si $S^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$,
$$\sqrt{n}\left(S^2 - \mu_2\right) \Rightarrow \mathcal{N}\left(0, \mu_4 - \mu_2^2\right).$$
- Montrer que si (X_1, Y_1) est un vecteur gaussien tel que $\rho = corr(X_1, Y_1)$, montrer que
$$\sqrt{n}\,(\rho_n - \rho) \Rightarrow \mathcal{N}\left(0, (1-\rho^2)^2\right),$$
où ρ_n est le coefficient de corrélation empirique construit à partir de $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d.
- On peut en théorie utiliser des DL d'ordre supérieur si la dérivée d'ordre 1 est nulle. Par exemple, si $\mathbb{E}(X_1) = 0 = 1 - \mathbb{E}X_1^2$, déterminer la loi asymptotique de $\cos\left(\overline{X}_n\right)$ à partir d'un DL à l'ordre 2.