Autour de la convergence en loi et théorème de Donsker.

Benjamin Bobbia

Université de Franche Comté

6 avril 2018

Étant données des variables aléatoires $(X_n)_{n\in\mathbb{N}}$ et X sur un espace Ω on aimerait étudier la convergence en loi des X_n vers X.
Si les variables sont réelles, les résultats sont bien connus, mais quel sens donner à cette convergence si on a affaire à une suite de processus, $X_n : x \mapsto X_n(x)$?
Classiquement on a, pour tout x

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i(x)-\mu(x)}{\sigma(x)} \overset{\mathcal{L}}{\to} \mathcal{N}(0,1).$$

Et le TCL multivarié fournit pour tout k entier et toute grille $(t_1, ..., t_k)$

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i(t_1)-\mu(t_1)}{\sigma(t_1)}, ..., \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i(t_k)-\mu(t_k)}{\sigma(t_k)}\right) \overset{\mathcal{L}}{\to} \mathcal{N}(0,\Sigma_k)$$

avec Σ_k la matrice de covariance. A-t-on $X_n \overset{\mathcal{L}}{\to} X$?

Définition

Soient $(P_n)_{n\in\mathbb{N}}$ et P des mesures de probabilité. On dit que la suite $(P_n)_{n\in\mathbb{N}}$ converge faiblement vers P si

$$\int_\Omega f dP_n \underset{n\to\infty}{\longrightarrow} \int_\Omega f dP$$

Pour toute fonction f continue bornée à valeurs dans $\mathbb{R}$.
On le note $P_n \Rightarrow P$.

Ce qui nous intéresse ici est de savoir si, étant donnés des processus stochastiques $X_n \sim P_n$ et $X \sim P$, établir la convergence en loi des processus revient à montrer que $P_n \Rightarrow P$.

Théorème (Portmenteau)

Les assertions suivantes sont équivalentes

- $P_n \Rightarrow P$.
- $\int_\Omega f dP_n \underset{n\to\infty}{\longrightarrow} \int_\Omega f dP$ *pour toute* f *uniformément continue.*
- $\limsup P_n(F) \leq P(F)$ *pour tout* F *fermé.*
- $\liminf P_n(O) \geq P(O)$ *pour tout* O *ouvert.*
- $P_n(A) \to P(A)$ *pour tout* A *tel que* $P(\partial A) = 0$.

Théorème

La suite $(P_n)_{n\in\mathbb{N}}$ *converge faiblement vers* P *si et seulement si de toute sous-suite de* $(P_n)_{n\in\mathbb{N}}$ *on peut extraire une sous-suite qui converge faiblement vers* P.
On dit que la suite $(P_n)_{n\in\mathbb{N}}$ *est relativement compacte.*

Définition

- *Une mesure de probabilité P est dite tendue si pour tout $\epsilon > 0$ il existe un compact K tel que $P(K) > 1 - \epsilon$.*
- *Une suite de mesure de probabilité $(P_n)_{n\in\mathbb{N}}$ est dite tendue si pour tout $\epsilon > 0$ il existe un compact K tel que pour tout $n \in \mathbb{N}$, $P_n(K) > 1 - \epsilon$.*

Comment montrer qu'une mesure est tendue ?

- En construisant le compact "à la main".
- En utilisant le théorème d'Ascoli.
- Grâce à des arguments d'intégrabilité uniforme.

Théorème (Prohorov)

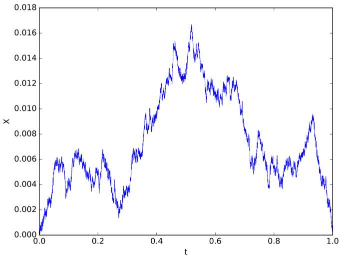
Dans un espace séparable,
la suite $(P_n)_{n\in\mathbb{N}}$ est relativement compacte si et seulement si elle est tendue.

Problématique : Étant donné un échantillon i.i.d $(x_1, ..., x_n)$ de réels, on aimerait bien connaître la loi du phénomène observé, c'est-à-dire, estimer sa densité. Ici on s'intéressera plutôt à son intégrale, c'est-à-dire sa fonction de répartition.

Définition

Un processus $(\mathbb{B}_t)_{t\geq 0}$ *est un mouvement brownien si*

- $(\mathbb{B}_t)_{t\geq 0}$ *est un processus gaussien.*
- $(\mathbb{B}_t)_{t\geq 0}$ *est presque-sûrement continue.*
- *Pour tout* s, t, $\mathbb{E}(\mathbb{B}_t) = 0$ *et* $\mathbb{E}(\mathbb{B}_t\mathbb{B}_s) = \min(s, t)$.



Cas particuliers :

1. Si $\mathbb{B}(0) = 0$, $\mathbb{B}$ est appelé mouvement brownien standard.
2. Si $\mathbb{B}(0) = \mathbb{B}(1) = 0$, $\mathbb{B}$ est appelé pont brownien.

Exemple de mouvement brownien :

$$t \mapsto tN_0 + \sum_{k=1}^{\infty} \frac{1}{\sqrt{2}\pi k}(N_k \cos(2\pi kt - 1) + N'_k \sin(2\pi kt))$$

avec N_k et N'_k des variables aléatoires normales centrées réduites.

Théorème (Donsker (1951))

Soit $(U_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires centrées, de variance σ^2 et de carré intégrable. Considérons, pour tout $t \in [0,1]$,

$$X_n(t) = \frac{1}{\sigma\sqrt{n}}\left(\sum_{k=1}^{\lfloor nt\rfloor} U_k + (nt - \lfloor nt\rfloor)U_{\lfloor nt\rfloor+1}\right)$$

alors

$$X_n \Rightarrow \mathbb{B}$$

avec $\mathbb{B}$ un mouvement brownien.

Cas des processus empiriques

Soit $(X_n)_{n\in\mathbb{N}}$ une suite i.i.d de loi uniforme sur $[0,1]$. On appelle F la fonction de répartition des X_n et on note F_n la fonction de répartition empirique. Pour tout $t \in [0,1]$, $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$. En posant $\mathbb{G}_n(t) = \sqrt{n}(F_n(t) - F(t))$ on a

$$\mathbb{G}_n \Rightarrow \mathbb{G}$$

avec $\mathbb{G}$ un pont brownien.
Traiter le cas uniforme est suffisant car on peut montrer que, si les X_i ne sont pas uniforme alors $\mathbb{G}_n(t) \overset{\mathcal{L}}{=} \mathbb{U}_n(F(t))$. Avec $\mathbb{U}_n$ le processus empirique sur un échantillon uniformes.

Merci de votre attention ! !