Interlude : Couplages

Un couplage de deux variables aléatoires X et Y (pas forcément définies sur le même espace de proba) est une variable aléatoire (X', Y') telle que $X' \overset{loi}{=} X$, $Y' \overset{loi}{=} Y$.

Exemple Soit $n \geq 1$ un entier, et $0 \leq p \leq q \leq 1$. Alors pour tout entier $0 \leq k \leq n$,

$$\mathbb{P}(\mathrm{Bin}(n,p) \geq k) \leq \mathbb{P}(\mathrm{Bin}(n,q) \geq k)$$

Preuve : On construit un couplage dans lequel $\mathrm{Bin}(n,p) \leq \mathrm{Bin}(n,q)$ en considérant $(U_i)_{1 \leq i \leq n}$ iid sur $[0,1]$ et en posant $X' = \sum_{i=1}^{n} \mathbb{1}_{U_i \leq p}$, $Y' = \sum_{i=1}^{n} \mathbb{1}_{U_i \leq q}$.

~

Plan : 1) Théorème de représentation de Skorokhod

2) Distance en variation totale

3) Distance de Lévy-Prokhorov

4) Monotonie, ordre stochastique

Références pour un bestiaire de distances sur $\mathcal{M}_1(E)$:

- Alison L. Gibbs, Francis Edwards Su, <u>On choosing and bounding probability metrics</u>
- Svante Janson, <u>Probability distances</u>

1) Théorème de représentation de Skorokhod

Théorème Soit (E,d) séparable, $(X_i)_{i \geq 1}$, X des v.a à valeurs dans E telles que $X_n \xrightarrow{loi} X$. On peut définir des v.a $(X'_i)_{i \geq 1}$, X' sur un même espace de proba tq $\forall n \geq 1$, $X_n \overset{loi}{=} X'_n$, $X \overset{loi}{=} X'$ et $X'_n \xrightarrow{p.s} X'$.

Voir feuille d'exos pour la preuve dans le cas $E = \mathbb{R}$

En pratique, ce théorème est utilisé pour se ramener à un "raisonnement déterministe" à partir d'une convergence en loi (on fait "comme si" $X_n \to X$ de manière déterministe)

⚠ Toute information sur les lois jointes des X_i est perdue chez les X_i'.

Application Soit $X, X_2, X_1, \dots$ des v.a réelles tq $X_n \xrightarrow{\text{loi}} X$. On suppose que $\lim_{a \to \infty} \sup_{n \geq 1} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geq a}] = 0$ $(*)$

Alors $\mathbb{E}[X_n] \longrightarrow \mathbb{E}[X]$

Preuve : D'après le thm de représentation de Skorokhod, on peut supposer que $X_n \xrightarrow{p.s} X$ et que $(*)$ (car c'est une propriété des lois individuelles des X_n).

Alors $X_n \xrightarrow{\mathbb{P}} X$ et la conclusion découle du fait que

$$X_n \xrightarrow{\mathbb{P}} X, \ (X_n)_{n \geq 1} \text{ UI} \implies X_n \xrightarrow{L^1} X$$

~

Pour la preuve, on admet l'existence de v.a indépendantes ayant des lois fixées et on utilise le lemme suivant

Lemme Soit (E, d) séparable et $\mu \in \mathcal{M}_1(E)$. Pour tout $\varepsilon > 0$, il existe une partition $E = \coprod_{i \geq 1} A_i$ en ensembles mesurables tel que

- $\forall i \geq 1$, $\mathrm{diam}(A_i) \leq \varepsilon$
- $\forall i \geq 1$, $\mu(\partial A_i) = 0$.

Preuve : Soit $(x_i)_{i\geq 1}$ une suite dénombrable dense. Pour tout $i\geq 1$,

$\{r\in[\frac{\varepsilon}{4},\frac{\varepsilon}{2}] : \mu(\partial B(x_i,r))>0\}$ est dénombrable.

On choisit alors $\frac{\varepsilon}{2}\leq r_i\leq\varepsilon$ tq $\mu(\underbrace{\partial B(x_i,r_i)}_{B_i})=0$

On définit par récurrence $A_1=B_1, \dots, A_i=B_1^c\cap\dots\cap B_{i-1}^c\cap B_i$

et on vérifie que $\mu(\partial A_i)=0$ (en effet, $\partial A=\partial(A^c)$ et $\partial(A\cup B)\subset\partial A\cup\partial B$, donc

$\mu(\partial A)=0 \Rightarrow \mu(\partial A^c)=0$ et $\mu(\partial A)=\mu(\partial B)=0 \Rightarrow \mu(\partial(A\cup B))=0$)

~

Preuve du théorème de représentation de Skorokhod.

Notons μ_n la loi de X_n, μ la loi de X

Étape 1 : Soit $p\geq 1$ et $(A_i^p)_{i\geq 1}$ une partition donnée par le lemme pour $\varepsilon_p=\frac{1}{2^p}$ avec μ

Soit k_p tel que $\mu\left(\bigcup_{i=k_p+1}^{\infty} A_i^p\right)<\varepsilon_p$

On pose $A_0^p=\bigcup_{i=k_p+1}^{\infty} A_i^p$, de sorte que $\mu(A_0^p)<\varepsilon_p$ et $\mu(\partial A_0^p)=0$.

En particulier, $A_0^p, A_1^p, \dots, A_{k_p}^p$ est une partition de E.

Par convergence en loi, on peut trouver a_p tel que

$$n\geq a_p \Rightarrow \mu_n(A_i^p)\geq(1-\varepsilon_p)\,\mu(A_i^p) \quad \text{pour } 0\leq i\leq k_p \quad (*)$$

En particulier, $\mu_n(A_i^p)=0 \Rightarrow \mu(A_i^p)=0$

Sans perte de généralité, on suppose $a_1<a_2<\dots$ et $n>a_1$.

Étape 2 L'idée est essentiellement de définir $X_n'=X_n\,|\,X_n\in A_i^p$ lorsque $X\in A_i^p$.

Il faut cependant ajuster à la marge pour que $X_n'\overset{\text{loi}}{=}X_n$.

On va définir X'_n comme suit pour $a_p \leq n < a_{p+1}$:

$$X'_n = \sum_{\substack{i=0,\\ \mu_n(A_i^p)>0}}^{k_p} \mathbb{1}_{\{U \leq 1-\varepsilon_p,\, X' \in A_i^p\}} Y_{n,i} + \mathbb{1}_{U > 1-\varepsilon_p} Z_n$$

avec $(Y_{n,i})_{n,i\geq 1}$, $(Z_n)_{n\geq 1}$, U, X' v.a ⫫ définies sur le même espace de proba tq :

- U uniforme sur $[0,1]$
- $X \overset{\text{loi}}{=} X'$
- $\text{Loi}(Y_{n,i}) = \mu_n(\cdot \mid A_i^p)$
- $\text{Loi}(Z_n)$ définie par

$$\mathbb{P}(Z_n \in A) = \frac{1}{\varepsilon_p} \sum_{\substack{i=0,\\ \mu_n(A_i^p)>0}}^{k_p} \mu_n(A \mid A_i^p) \left(\mu_n(A_i^p) - (1-\varepsilon_p)\, \mu(A_i^p) \right),$$

quantité qui définit bien une mesure positive par (*) et de masse totale 1 (car $\mu_n(A_i^p)=0 \Rightarrow \mu(A_i^p)=0$)

On vérifie que $X'_n \overset{\text{loi}}{=} X_n$:

$$\begin{aligned}
\mathbb{P}(X'_n \in A) &= \sum_{\substack{i=0,\\ \mu_n(A_i^p)>0}}^{k_p} \mathbb{P}(U \leq 1-\varepsilon_p,\, X' \in A_i^p,\, Y_{n,i} \in A) + \mathbb{P}(U > 1-\varepsilon_p,\, Z_n \in A) \\
&= (1-\varepsilon_p) \sum_{\substack{i=0,\\ \mu_n(A_i^p)>0}}^{k_p} \underbrace{\mathbb{P}(X' \in A_i^p)}_{\mu(A_i^p)} \mu_n(A \mid A_i^p) + \varepsilon_p \mathbb{P}(Z_n \in A) \\
&= \mu_n(A)
\end{aligned}$$

car $\mathbb{P}(Z_n \in A) = \frac{1}{\varepsilon_p}\left(\mu_n(A) - (1-\varepsilon_p) \sum_{\substack{i=0\\ \mu_n(A_i^p)>0}}^{k_p} \mu(A_i^p)\, \mu_n(A \mid A_i^p) \right)$

<u>Étape 3</u> On vérifie que $X'_n \xrightarrow{ps} X'$.

Soit $E_p = \{ X' \notin A_0^p,\, U \leq 1-\varepsilon_p \}$.

Si $a_p \leq n < a_{p+1}$ et si E_p est réalisé, $\exists\, 1 \leq i \leq k_p$ tq $X', X'_n \in A_i^p$

et donc $d(X', X'_n) \leq \varepsilon_p$.

Or $\mathbb{P}((E_p)^c) \leq \mathbb{P}(X \in A_0^p) + \mathbb{P}(U > 1-\varepsilon_p) \leq \frac{1}{2^{p-1}}$.

Donc $\sum_{p \geq 1} \mathbb{P}((E_p)^c) < \infty$.

Donc par Borel-Cantelli, ps, à partir d'un certain rang E_p est réalisé. Donc $X'_n \xrightarrow{p-s} X'$.

~

Remarque Lorsque E est polonais, il est possible de montrer que l'espace de proba commun peut être pris comme $[0,1]$ muni de la mesure de Lebesgue

2) Distance en variation totale

Soit $(S,\mathcal{B})$ un espace mesuré.

Définition Pour $\mu,\nu \in \mathcal{M}_1(S)$, on pose $d_{VT}(\mu,\nu) = \sup_{A\in\mathcal{B}} |\mu(A)-\nu(A)|$

Pour des variables aléatoires X, Y à valeurs dans S, on pose

$$d_{VT}(X,Y) = \sup_{A\in\mathcal{B}} |\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)|$$

On montre les résultats suivants :

- d_{VT} est une distance sur $\mathcal{M}_1(S)$ qui le rend complet
- $d_{LP} \leq d_{VT}$
- Si μ et ν sont absolument continues par rapport à une même σ-finie ρ sur $(S,\mathcal{B})$, alors $d_{VT}(\mu,\nu) = \frac{1}{2}\int_E \left|\frac{d\mu}{d\rho} - \frac{d\nu}{d\rho}\right| d\rho$

Ceci permet de montrer que :

$$d_{VT}(\mu,\nu) = \frac{1}{2}\sup\left\{ \left|\int f d\mu - \int f d\nu\right| : f : S\to\mathbb{R} \text{ mesurable}, \|f\|_\infty \leq 1\right\}$$

et que $d_{VT}(\mu,\nu) = \inf\left\{ \mathbb{P}(X'\neq Y') : (X',Y') \text{ couplage de } X,Y\right\}$,

avec l'inf qui est atteint.

- Lorsque S est dénombrable, muni de la distance discrète, $\mathcal{B} = \mathcal{P}(S)$ la convergence pour d_{VT} est équivalente à la convergence en loi (⚠ faux en général) et $d_{VT}(\mu,\nu) = \frac{1}{2}\sum_{x\in S} |\mu(x) - \nu(x)|$

Remarque : Une p-métrique sur S est une fonction δ à valeurs dans $[0,\infty]$ définie sur l'ensemble des lois jointes de variables aléatoires (X,Y) à valeurs dans S telle que pour des v.a X,Y,Z définies sur le même espace :

① Si $X=Y$ p.s, alors $\delta(X,Y)=0$ ② $\delta(X,Y)=\delta(Y,X)$ ③ $\delta(X,Y)\leq\delta(X,Z)+\delta(Z,Y)$.

La quantité $\hat{\delta}(X,Y)=\inf\{\delta(X',Y') : X'\overset{loi}{=}X,\ Y'\overset{loi}{=}Y\}$ est appelée <u>métrique minimale</u> associée à δ.

Ainsi, d_{VT} est la métrique minimale associée à $\delta(X,Y)=\mathbb{P}(X\neq Y)$.

Remarque Pour un ensemble $\mathcal{F}$ de fonctions mesurables $S\to\mathbb{R}$, on pose pour X,Y v.a à valeurs dans S, $d_{\mathcal{F}}(X,Y)=\sup\{|\mathbb{E}[f(X)]-\mathbb{E}[f(Y)]| : f\in\mathcal{F}\}$, appelée <u>métrique duale définie par $\mathcal{F}$</u>. Ainsi, d_{VT} est la métrique duale définie par l'une des 3 familles $\mathcal{F}=\{f:S\to\mathbb{R}$ mesurable, $\|f\|_\infty\leq\frac{1}{2}\}$, $\mathcal{F}=\{f:S\to\mathbb{R}$ mesurable, $0\leq f\leq 1\}$ ou $\mathcal{F}=\{\mathbb{1}_A : A\in\mathcal{B}(S)\}$

Référence : • Notes de cours "Probability Theory: The Coupling Method" de Frank den Hollander : http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf (Section 2)

3) Distance de Lévy-Prokhorov

Soit (E,d) métrique séparable. On rappelle que $d_{LP}(\mu,\nu)=\inf\{\varepsilon>0 : \forall A\in\mathcal{B}(E),\ \mu(A)\leq\nu(A^\varepsilon)+\varepsilon\}$ pour $\mu,\nu\in\mathcal{M}_1(E)$, où $A^\varepsilon=\{x\in E : d(x,A)<\varepsilon\}$.

Définition Pour des variables aléatoires X,Y définies sur le même espace de proba et à valeurs dans E, on pose $\alpha(X,Y)=\inf\{\varepsilon>0 : \mathbb{P}(d(X,Y)\geq\varepsilon)<\varepsilon\}$.

C'est la métrique de Ky-Fan, et on peut vérifier qu'elle métrise la convergence en probabilité.

Théorème (Strassen) Soit (E,d) séparable, $\mu,\nu \in \mathcal{M}_1(E)$. Alors

$$d_{LP}(\mu,\nu) = \inf\{\alpha(X,Y) : (X,Y) \text{ couplage avec } Loi(X)=\mu \text{ et } Loi(Y)=\nu\}.$$

Si μ et ν sont tendues (toujours le cas si E complet), alors l'inf est atteint

Preuve d'une inégalité: Soit (X,Y) un tel couplage. On mq $d_{LP}(\mu,\nu) \leq \alpha(X,Y)$.

Soit $\varepsilon > \alpha(X,Y)$, de sorte que $\mathbb{P}(d(X,Y)\geq\varepsilon) < \varepsilon$.

Alors $\mu(A) = \mathbb{P}(X\in A) \leq \mathbb{P}(X\in A, d(X,Y)<\varepsilon) + \mathbb{P}(X\in A, d(X,Y)\geq\varepsilon)$

$$\leq \mathbb{P}(Y\in A^\varepsilon) + \varepsilon = \nu(A^\varepsilon) + \varepsilon,$$

d'où $d_{LP}(\mu,\nu) \leq \varepsilon$

$\sim$

L'autre inégalité est délicate et peut se montrer en utilisant le lemme des mariages, cf Billingsley, Convergence of probability measures (2nd Ed), Thm 6.9

ou Dudley, Real Analysis and probability, Corollary 11.6.4.

Remarque Ainsi, d_{LP} est la métrique minimale associée à α.

On finit cette partie par un résultat utile de couplage "à la Skorokhod":

Théorème (Dudley) Soit (E,d) séparable, X_n, Y_n des v.a à valeurs dans E tq $d_{LP}(X_n,Y_n) \longrightarrow 0$

Alors il existe un espace de proba avec des v.a $(X'_n, Y'_n)_{n\geq 1}$ définies dessus tq

- $\forall n\geq 1$, $X'_n \overset{loi}{=} X_n$, $Y'_n \overset{loi}{=} Y_n$
- ps, $d(X'_n, Y'_n) \longrightarrow 0$

cf Dudley, Real Analysis and probability, Theorem 11.7.1

4) Monotonie, ordre stochastique

Ici (E,d) est un espace polonais.

Définition Une relation $\leq$ est un ordre partiel si $\forall x,y,z \in E$:

1) $x \leq x$ 2) $x \leq y, y \leq z \Rightarrow x \leq z$ 3) $x \leq y, y \leq x \Rightarrow x = y$

Exemples • $\mathbb{R}$ muni de l'ordre usuel

• $\{0,1\}^{\mathbb{Z}^2}$ avec $(a_i)_{i\in\mathbb{Z}^2} \leq (b_i)_{i\in\mathbb{Z}^2}$ si $a_i = 1 \Rightarrow b_i = 1$ (si $a_i = 1$, on dit que le site i est occupé, vide sinon)

Définition Un événement A est dit croissant si $x \in A \Rightarrow \{y \in E : x \leq y\} \in A$.

Exemple Pour $E = \{0,1\}^{\mathbb{Z}^2}$, $A = \{\exists$ chemin infini constitué de sites occupés adjacents$\}$ est croissant

Définition Soit $\mu, \nu \in \mathcal{M}_1(E)$. On dit que ν domine stochastiquement μ et on note $\mu \leq \nu$ si $\forall A \in \mathcal{B}(E)$ croissant, $\mu(A) \leq \nu(A)$, ou, de manière équivalente

$$\int_E f\,d\mu \leq \int_E f\,d\nu \quad \forall f : E \to \mathbb{R} \text{ mesurable, bornée, croissante.}$$

(Par croissante on veut dire $x \leq y \Rightarrow f(x) \leq f(y)$)

Exemple Pour $E = \mathbb{R}$, $\mu \leq \nu \Leftrightarrow \forall x \in \mathbb{R}, \ \mu(]-\infty, x]) \leq \nu(]-\infty, x])$.

Théorème (Strassen) On suppose que $\{(x,y) \in E^2 : x \leq y\}$ est fermé. Soit $\mu, \nu \in \mathcal{M}_1(E)$. Alors (1) $\Leftrightarrow$ (2)

(1) $\mu \leq \nu$

(2) $\exists$ couplage (X,Y) tq $\mathcal{L}oi(X) = \mu$, $\mathcal{L}oi(X) = \nu$ et $\mathbb{P}(X \leq X') = 1$

La preuve repose sur de l'analyse fonctionnelle, cf

Lindvall, <u>Lectures on the Coupling Method</u>, chapter IV

ou Lindvall, <u>On Strassen's theorem on stochastic domination</u>, ECP 4 (1999), 57-59