MIA - **Notes de la Séance 7**

GABIOT Maxime & PLAS Jean-Pierre

4 Octobre 2022

Table des matières

1 Statistique inférentielle asymptotique (classique)

1.1 Modes de convergence

Il y a quatre modes de convergences différents.
On a $X_1, \ldots, X_n$ une suite de variables aléatoires réelles (v.a.r.) définies sur Ω. On va définir la limite $\lim_{n \to +\infty} X_n$. Or, comme il y a plusieurs types de convergences, on note plutôt :

1.1.1 Convergence presque sûre ou presque partout

$$\begin{aligned}
& X_n \to_{a.e.} X \sim P, && a.e. = \text{almost everywhere}, \\
& \equiv X_n \to_{p.s.} X \sim P, && p.s. = \text{presque sûr}, \\
& \equiv X_n \to_{a.s.} X \sim P, && a.s. = \text{almost sure}. \\
& \text{ssi } \exists N \subset \Omega, \text{ tq } P(N) = 0, \forall \omega \in \Omega \setminus N, \lim_{n \to +\infty} X_n(\omega) = X(\omega). \\
& \equiv \forall \varepsilon > 0, \lim_{n \to +\infty} P(\{\omega \mid \omega \in \Omega \text{ et } \exists m \geq n\} \|X_m(\omega) - X(\omega)\| > \varepsilon)
\end{aligned}$$

Un estimateur θ_n de θ est fortement consistant de θ si et seulement si la suite de var $(\theta_n)n \in \mathbb{N}^*$ converge presque sûrement vers θ.

1.1.2 Convergence en probabilité

$$\begin{aligned}
& X_n \to_p X \sim P \\
& \text{ssi } \forall \varepsilon > 0, \lim_{n \to +\infty} P(\|X_n - X\| \geq \varepsilon) = 0
\end{aligned}$$

Un estimateur θ_n de θ est faiblement consistant de θ si et seulement si la suite de var $(\theta_n)n \in \mathbb{N}^*$ converge en probabilité vers θ.

i.e. son biais tend vers 0 quand n tend vers l'infini.

1.1.3 Convergence en loi

$X_1, \ldots, X_n$ v.a.r. définies sur Ω. On Dit que (X_n) tend vers $X \sim P$ en distribution, écrit $X_n \to_d X$, ssi pour tout $x \in \mathbb{R}$ tel que F_x la fonction de répartition de X soit continue en x, alors $\lim_{n \to +\infty} F_{n,X}(x) = F_X(x)$ où $F_{n,X}$ est la fonction de répartition de X_n.

Aussi, elle diffère des 2 lois précédentes, car on ne regarde pas événement par événement, mais on regarde directement les fonctions de répartitions.

$$\text{Si } X_n \to_{a.e.} X \sim P \text{ alors } X_n \to_p X \sim P.$$
$$\text{Si } X_n \to_p X \sim P \text{ alors } X_n \to_d X \sim P.$$

La convergence presque sûre est plus forte que la convergence en probabilité. Cela veut dire que s'il y a convergence presque sûre, alors il y a convergence en probabilité (mais la réciproque n'est pas vraie). Il y a la même relation entre convergence en probabilité et convergence en loi.

1.1.4 Convergence en moyenne quadratique

Nous n'avons pas vu ce mode.

1.2 Applications des modes de convergence

1.2.1 Loi forte des grands nombres

$X_1, X_2, \ldots$ variables aléatoires i.i.d.

(The SLLN). Une condition nécessaire et suffisante pour l'existence d'une constante c telle que

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow_{a.s.} c$$

est que $\mathbb{E}[X_1] < \infty$, dans ce cas $c = \mathbb{E}[X_1]$.
On peut réécrire cette condition comme $X1$ a une moyenne ou $X1$ a une espérance $< \infty$.

Remarque :
Il existe aussi la loi faible des grands nombres qui repose sur une convergence en probabilité. Dans certains cas de figure, la loi faible s'applique, mais pas la forte.

1.2.2 Théorème de la limite centrale

(Multivariée) Soit $X_1, \ldots, X_n$ i.i.d. des k-vecteurs aléatoires avec $\Sigma = \operatorname{Var}(X_1)$ fini.
Alors

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \rightarrow_d N_k(0, \Sigma).$$

où en multipliant par $\frac{n}{n}$ on obtient $\sqrt{n}(\hat{\mu}_n - \mu)$ et où $\mathbb{E}[X_1]$ peut-être réécrit μ.

Pour n grand, $\hat{\mu}_n - \mu \sim$ approx $\mathcal{N}\left(0, \frac{\Sigma}{n}\right)$
Pour n grand, $\hat{\mu}_n \sim$ approx $\mathcal{N}\left(\mu, \frac{\Sigma}{n}\right)$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \\
&= \frac{n}{n} * \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \\
&= \frac{\sqrt{n}\sqrt{n}}{n\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \\
&= \frac{\sqrt{n}}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \\
&= \sqrt{n} * \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1]) \\
&= \sqrt{n}(\hat{\mu}_n - \mu)
\end{aligned}$$

1.2.3 Transformations continues

Théorème 1.10.
Soit $X, X_1, X_2, \ldots$ des k-vecteurs aléatoires définis sur un espace de probabilité
et g une fonction de $\left(\mathcal{R}^k\right)$ en $\left(\mathcal{R}^l\right)$.
Supposons que g est continue.
Alors
(i) $X_n \rightarrow_{\text{a.s.}} X$ implique $g(X_n) \rightarrow_{\text{a.s.}} g(X)$;
(ii) $X_n \rightarrow_p X$ implique $g(X_n) \rightarrow_p g(X)$;
(iii) $X_n \rightarrow_d X$ implique $g(X_n) \rightarrow_d g(X)$.

Exemple pour (i) :
$X_1, \ldots, X_n$ i.i.d. sur P avec $\mu = \mathbb{E}[P]$.
$\hat{\mu}_n(X_1, \ldots, X_n) \rightarrow_{a.e.} \mu$
$g : x \mapsto x^2$
$g(\hat{\mu}_n) \rightarrow_{a.s.} g(\mu)$
$\hat{\mu}_n^2 \rightarrow_{a.s.} \mu^2$

1.2.4 Théorème de Slutsky

Théorème 1.11 (Théorème de Slutsky).
Soit $X, X_1, X_2, \ldots, Y_1, Y_2, \ldots$ des variables aléatoires sur un espace de probabilité.
Supposons que $X_n \rightarrow_d X$ et $Y_n \rightarrow_p c$, où c est une constante réelle.
Alors
(i) $X_n + Y_n \rightarrow_d X + c$
(ii) $Y_n X_n \rightarrow_d cX$
(iii) $X_n / Y_n \rightarrow_d X/c$ si $c \neq 0$.

1.2.5 Méthode delta

La méthode delta nous permet d'étudier les fluctuations de $g(X_n)$ autour de $g(c)$.

Soit $X_1, X_2, \ldots, Y$ des vecteurs aléatoires de dimension k tels que :

$$a_n(X_n - c) \rightarrow_{d.} Y$$

pour c appartenant à $\mathbb{R}^k$ et (a_n) une suite de nombres positifs tendant vers $+\infty$.

Alors pour toute fonction $g : \mathbb{R}^k \rightarrow \mathbb{R}$ différentiable en c, on a :

$$a_n[g(X_n) - g(c)] \rightarrow_{d.} [\nabla g(c)]^T Y$$

2 Optimisation

2.1 Analyse

2.1.1 Définition de Limite

Concept central :
$\lim_{x \to a} f(x) = l$
$\forall \epsilon > 0$, $\exists \delta > 0$ tel que
$||x - a||_2 \leq \delta \implies ||f(x) - l|| < \epsilon$

2.1.2 Notions de base de topologie

Dans $\mathbb{R}^n$, La boule fermée $B_f(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbb{R}^n$ tels que $\|x - y\|_2 \leq \epsilon$.

Dans $\mathbb{R}^n$, La boule ouverte $B_o(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbb{R}^n$ tels que $\|x - y\|_2 < \epsilon$.

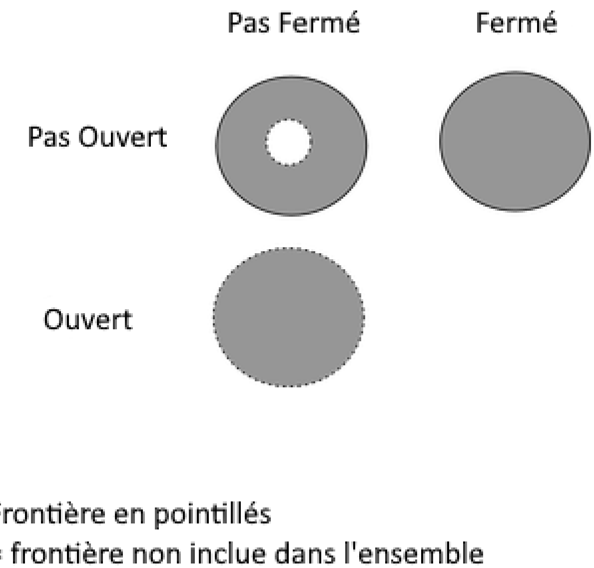
U est un ouvert de $\mathbb{R}^n$ ssi $U \subset \mathbb{R}^n$ et pour tout x dans U, il existe $\epsilon > 0$, tel que $B_f(x, \epsilon) \subset U$.

On peut dire qu'un ouvert est un ensemble qui n'inclut aucune frontière.

U est un fermé de $\mathbb{R}^n$ ssi son complément dans $\mathbb{R}^n$ est un ouvert de $\mathbb{R}^n$.

On peut dire qu'un fermé est un ensemble qui inclut toutes ses frontières.

Figure 1 – Schéma représentatif d'ensembles ouvert, fermé, et ni ouvert ni fermé.



Exemple : [0 , 5[est un sous-ensemble des Réels
Il n'est ni ouvert ni fermé car on inclut une frontière mais pas l'autre frontière.

2.2 Jacobienne, Gradient

$f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, de classe $\mathcal{C}^1$ ($\mathcal{C}^1$, qui ont une dérivée continue)
Dérivée partielle $f : x \mapsto (f_1(x), f_2(x), \ldots, f_m(x))$

$$\frac{\partial f_i}{\partial x_j} : (a_1, \ldots, a_n) \mapsto \lim_{h \to 0} \frac{f_i\left(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots a_n\right) - f_i\left(a_1, \ldots, a_j, \ldots, a_n\right)}{h}$$

2.2.1 Gradient

Gradient de $f_1 : U \subset \mathbb{R}^n \to \mathbb{R}$

$$\nabla f_1 : x \mapsto \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f_1(x)}{\partial x_n} \end{bmatrix}$$

2.2.2 Jacobienne

Matrice Jacobienne (transposée du gradient si $\mathbf{m} = \mathbf{1}$)

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

"Chain rule"

$$J_{f \circ g}(\mathbf{a}) = J_f(g(\mathbf{a})) J_g(\mathbf{a})$$

Exemple de gradient / Jacobienne :

$$f : \left| \begin{array}{c} \mathbb{R}^2 \to \mathbb{R}^2 \\ x, y \mapsto x+y, x-y \end{array} \right.$$

$$g : \left| \begin{array}{c} \mathbb{R}^2 \to \mathbb{R}^3 \\ x, y \mapsto x^2, y^2, x^2+y^2 \end{array} \right.$$

Que vaut $J_{g \circ f}$?

$$g \circ f : \left| \begin{array}{c} \mathbb{R}^2 \to \mathbb{R}^3 \\ x, y \mapsto (x+y)^2, (x-y)^2, (x+y)^2 + (x-y)^2 \end{array} \right.$$

$$J_f(x,y) = \begin{bmatrix} \frac{\partial(x+y)}{\partial x} & \frac{\partial(x+y)}{\partial y} \\ \\ \frac{\partial(x-y)}{\partial x} & \frac{\partial(x-y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$J_g(x,y) = \begin{bmatrix} \frac{\partial x^2}{\partial x} & \frac{\partial x^2}{\partial y} \\ \\ \frac{\partial y^2}{\partial x} & \frac{\partial y^2}{\partial y} \\ \\ \frac{\partial x^2+y^2}{\partial x} & \frac{\partial x^2+y^2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & 0 \\ 0 & 2y \\ 2x & 2y \end{bmatrix}$$

$$J_{g \circ f}(x,y) = J_g(f(x,y)) J_f(x,y) = \begin{bmatrix} 2(x+y) & 0 \\ \\ 0 & 2(x-y) \\ \\ 2(x+y) & 2(x-y) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \\ 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2(x+y) & 2(x+y) \\ \\ 2(x-y) & -2(x-y) \\ \\ 4x & 4y \end{bmatrix}$$

2.3 Optimisation sans contraintes

$$f : \mathbb{R}^n \mapsto \mathbb{R}$$
$$\min_x f(x)?$$

x^* est un minimum global de f ssi pour tout $x \in \mathbb{R}^n, f(x^*) \leq f(x)$ x^* est un minimum local de f ssi il existe un ensemble ouvert $U \subset \mathbb{R}^n$ contenant x^* tel que pour tout $x \in U, f(x^*) \leq f(x)$.

2.4 Conditions d'optimalité

Conditions nécessaires, premier ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbb{R}^n$ contenant x, alors $\nabla f(x^*) = 0$.