MACS203b

1 Convergence de variables aléatoires

Calcul sur les événements

Prop.
- *Si $(A_n)_n$ est croissante, $\mathbf{P}(\bigcup_n A_n) = \lim_{n\to\infty} \mathbf{P}(A_n)$.*
- *Si $(A_n)_n$ est décroissante, $\mathbf{P}(\bigcap_n A_n) = \lim_{n\to\infty} \mathbf{P}(A_n)$.*
- *Si $\forall n, \mathbf{P}(A_n) = 0$ alors $\mathbf{P}\left(\bigcup_n A_n\right) = 0$.*
- *Si $\forall n, \mathbf{P}(A_n) = A$ alors $\mathbf{P}\left(\bigcap_n A_n\right) = 1$.*

Def. $\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbf{N}} \bigcup_{k\geqslant n} A_k$, i.e. $\omega \in \limsup_n A_n \iff \forall n, \exists k \geqslant n, \omega \in A_k$.

Donc $\limsup_n A_n$ est réalisé ssi une infinité de A_n est réalisé.

Lem (de **Borel-Cantelli**). *Si $\sum_n \mathbf{P}(A_n) < \infty$, alors $\mathbf{P}(\limsup_n A_n) = 0$.*

Autrement dit, il y a une proba 1 pour que seulement un nombre fini de A_n soient réalisés.

Convergence p.s., en probabilité et dans L^p

Def. (i) On dit que $X_n \xrightarrow{\mathbf{P}} X$ (**converge en probabilité**) si $\forall \epsilon > 0, \mathbf{P}(\|X_n - X\| > \epsilon) \underset{n\to\infty}{\longrightarrow} 0$.

(ii) On dit que $X_n \xrightarrow{\text{p.s.}} X$ (**converge presque sûrement**), si $\forall \omega$ $\mathbf{P}$-p.p, $X_n(\omega) \to X(\omega)$. Autrement dit il existe $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 1$ et $\forall \omega \in A, \lim_n X_n(\omega) = X(\omega)$.

(iii) On dit que $X_n \xrightarrow{L^p} X$ (**converge vers X dans $L^p(\Omega, \mathbf{R}^d)$**) si $X_n, X \in L^p$ et $\mathbf{E}\left(\|X_n - X\|^p\right) \underset{n\to\infty}{\longrightarrow} 0$.

Prop. *On note $X_n = \left(X_n^{(1)}, \dots, X_n^{(d)}\right)$ sur $\mathcal{X} = \mathbf{R}^d$. Alors $X_n \xrightarrow{p.s.} X$ p.s. (resp. en probabilité, dans L^p) ssi $\forall k \in [\![1\,;d]\!], X_n^{(k)} \xrightarrow{p.s.} X^{(k)}$ (resp. en probabilité, dans L^p).*

Prop. *Si $X_n \xrightarrow{p.s.} X$ ou $X_n \xrightarrow{L^p} X$ alors $X_n \xrightarrow{\mathbf{P}} X$.*

Prop. *Si $\forall \epsilon > 0, \sum_n \mathbf{P}(\|X_n - X\| > \epsilon) < \infty$ alors $X_n \xrightarrow{p.s.} X$.*

Prop. *$X_n \xrightarrow{\mathbf{P}} X$ ssi de toute sous-suite $X_{\varphi(n)}$ on peut extraire une autre sous-suite $X_{\varphi\circ\psi(n)}$ telle que $X_{\varphi\circ\psi(n)} \xrightarrow{p.s.} X$.*

Th (de **continuité**). *X_n, X v.a. sur $\mathbf{R}^d$. Soit $h\colon \mathbf{R}^d \to \mathbf{R}^p$ mesurable et continue sur C tel que $\mathbf{P}(X \in C) = 1$, alors*

(i) Si $X_n \xrightarrow{p.s.} X$ alors $h(X_n) \xrightarrow{p.s.} h(X)$

(ii) Si $X_n \xrightarrow{\mathbf{P}} X$ alors $h(X_n) \xrightarrow{\mathbf{P}} h(X)$.

Th (Loi forte des grands nombres). *Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|) < \infty$. Alors $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p.s.} \mathbf{E}(X_1)$.*

Th (Loi faible des grands nombres). *Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|^2) < \infty$. On a $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E}(X_1)$.*

Convergence en loi

Rappels : une mesure de proba μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est caractérisée par sa fonction de répartition F_μ. $F_\mu(x_1, \dots, x_d) = \mu(\prod_i]-\infty\,; x_i])$.

On a :
- F_μ croissante.
- $F_\mu(-\infty) = 0$, $F_\mu(+\infty) = 1$
- F_μ est continue à droite et $\mu(x_0) = F_\mu(x_0) - F_\mu(x_0^-)$

Soit $X\colon \Omega \to \mathbf{R}^d$ une v.a. On note $P_X = \mathbf{P} \circ X^{-1}$ la loi de X. P_X est une mesure de proba sur $\mathbf{R}^d$. On note F_X sa fonction de répartition. Pour $d = 1$, $F_X(x) = \mathbf{P}(X \leqslant x)$.

Def. Soit $(\mu_n)_n, \mu$ des mesures de proba sur $\mathbf{R}^d$. On dit que μ_n converge faiblement (ou étroitement) vers μ si $F_{\mu_n}(x) \longrightarrow F_\mu(x)$ en tout x point de continuité de F_μ. On note $\mu_n \Rightarrow \mu$.

Def. $(X_n)_n, X$ v.a. sur $\mathbf{R}^d$. On dit que X_n converge en loi vers X (noté $X_n \xrightarrow{\mathcal{L}} X$) si $P_{X_n} \implies P_X$.

Prop. $\left.\begin{array}{r}\textit{cv ps}\\ \textit{ou}\\ \textit{cv } L^p\end{array}\right\} \implies$ *cv proba* $\implies$ *cv loi*

Th (de représentation de Skorohod). *Soit $(\mu_n)_n, \mu$ des mesures de proba sur $\mathbf{R}^d$ telles que $\mu_n \implies \mu$. Il existe un espace de proba et des v.a. $(Y_n), Y$ sur cet espace telles que :*
- *$Y \sim \mu$, $\forall n, Y_n \sim \mu_n$*
- *$\forall \omega, Y_n(\omega) \longrightarrow Y(\omega)$*

Th (de continuité). *Soit $X_n \xrightarrow{\mathcal{L}} X$ définie sur $(\Omega, \mathcal{F}, \mathbf{P})$. $h\colon \mathbf{R}^d \to \mathbf{R}^p$ continue sur C telle que $\mathbf{P}(X \in C) = 1$. Alors $h(X_n) \xrightarrow{\mathcal{L}} h(X)$.*

Th (de Portmanteau). *On a équivalence entre :*

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

(ii) $\forall f\colon \mathbf{R}^d \to \mathbf{R}$ *continue bornée,* $\mathbf{E}(f(X_n)) \longrightarrow \mathbf{E}(f(X))$,

(iii) $\forall A \subset \mathbf{R}^d$ *tel que* $\mathbf{P}(X \in \delta A) = 0$, *on a* $\mathbf{P}(X_n \in A) \longrightarrow P(X \in A)$ *où* $\delta A = \bar{A} \setminus \mathring{A}$

Lem (d'Helly). *Soit* $(F_n)_n$ *une suite de fonctions de répartition. Il existe une sous-suite* φ_n *et* $F\colon \mathbf{R} \to [0\,;1]$ *croissante, continue à droite, telle que* $F_{\varphi_n}(x) \longrightarrow_n F(x)$ *en tout* x *point de continuité de* F.

On ajoute une condition pour que la limite vérifie $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$.

Def. $(\mu_n)_n$ est dite **tendue** si $\forall\varepsilon > 0, \exists\mathcal{K}$ compact, $\forall n, \mu_n(\mathcal{K}) \geqslant 1-\varepsilon$.

Dans le cas $d = 1$ on peut prendre $\mathcal{K} = [-K\,;K]$.

Def. $(X_n)_n$ est tendue si $\forall\varepsilon > 0, \exists\mathcal{K}$ compact, $\forall n, \mathbf{P}(X_n \in \mathcal{K}) \geqslant 1-\varepsilon$.

Th (de Prokhorov). *Soit* $(\mu_n)_n$ *tendue. Il existe une mesure de probabilité* μ *sur* $\mathbf{R}^d$ *et une suite* $(\varphi_n)_n$ *telle que* $\mu_{\varphi_n} \implies \mu$.

Prop. *Si toute sous-suite faiblement convergente de* $(\mu_n)_n$ *tendue converge vers* μ^*, *alors* $\mu_n \implies \mu^*$.

Fonction caractéristique, TCL

La fonction caractéristique d'une mesure de proba μ sur $\mathbf{R}^d$ est

$$\varphi_\mu\colon \begin{array}{ccc} \mathbf{R}^d & \to & \mathbf{C} \\ t & \mapsto & \int e^{i\langle t|x\rangle}\,\mathrm{d}\mu(x) \end{array}$$

...

Rappel : $\varphi_\mu = \varphi_\nu \implies \mu = \nu$.

Ex. $\varphi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.

Pour $Y = AX + b$ on a $\varphi_Y(t) = e^{i\langle t|b\rangle}\varphi_X(A^\mathsf{T} t)$.

Prop. φ_μ *est continue en zéro.*

Th (de Lévy). *Soit* $(\mu_n)_n$, μ *des mesures de probabilité sur* $\mathbf{R}^d$. $\mu_n \implies \mu$ *ssi* $\forall t \in \mathbf{R}^d, \varphi_{\mu_n}(t) \longrightarrow \varphi_\mu(t)$.

Th (Procédé de Cramer-Wold). *Soit* X_n, X *des v.a. sur* $\mathbf{R}^d$. *On a* $X_n \xrightarrow{\mathcal{L}} X \iff \forall t, \langle t \mid X_n\rangle \xrightarrow{\mathcal{L}} \langle t \mid X\rangle$.

Théorème centrale limite

Not. $\mathcal{N}(m,\sigma^2)$ désigne la loi de densite $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$, et si $\sigma^2 = 0$ c'est la loi δ_m.

Pour X un vecteur gaussien, sa fonction caractéristique vérifie $\phi_X(t) = e^{i\langle t|m\rangle}e^{-\frac{t^\mathsf{T}\Sigma t}{2}}$ où ...

Th (central limite). ...

Th (de Linderbergh). *Soit un tableau de v.a.* $(X_{i,n})_{1\leqslant i\leqslant n}$ *tel que*

- $\forall n$, *les v.a* $X_{1,n},\dots,X_{n,n}$ *sont indépendantes,*
- $\forall n, \forall i \leqslant n, \mathbf{E}(X_{i,n}) = 0$,
- $\lim_n \sum_{i=1}^n \mathrm{Cov}(X_{i,n}) = \Sigma$
- $\forall\varepsilon > 0, \lim_n \sum_{i=1}^n \mathbf{E}(\|X_{i,1}\|^2 \mathbf{1}_{\|X_{i,n}\|} > \varepsilon) = 0$.

Alors $\sum_{i=1}^n X_{i,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)$.

2 Manipulation des convergences

Rappel de règles

Prop. *(i)* $X_n \xrightarrow{\mathcal{L}} c \iff X_n \xrightarrow{\mathbf{P}} c$

(ii) $\left.\begin{array}{r} X_n \xrightarrow{\mathcal{L}} X \\ Y_n - X_n \xrightarrow{\mathcal{L}} 0 \end{array}\right\} \implies Y_n \xrightarrow{\mathcal{L}} X$

(iii) $\left.\begin{array}{r} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} c \end{array}\right\} \implies (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$

(iv) $\left.\begin{array}{r} X_n \xrightarrow{\mathbf{P}} X \\ Y_n \xrightarrow{\mathbf{P}} c \end{array}\right\} \implies (X_n, Y_n) \xrightarrow{\mathbf{P}} (X, c)$

Notation o_P, O_P

Soit (X_n) des v.a. $\Omega \to \mathbf{R}^d$ et (Y_n) v.a.r.

Def. La notation $X_n = o_P(1)$ signifie $X_n \xrightarrow{\mathbf{P}} 0$. $X_n = o_P(Y_n)$ signifie $\exists(Z_n) \xrightarrow{\mathbf{P}} 0, X_n = Z_nY_n$. $X_n = O_p(1)$ signifie que (X_n) est tendue. On dit que X_n est "bornée en probabilité" lorsqu'elle est tendue. $X_n = O_P(Y_n)$ signifie $\exists(Z_n) = O_p(1), X_n = Z_nY_n$.

Prop. *Si* $X_n \xrightarrow{\mathcal{L}} X$ *alors* $X_n = O_P(1)$.

Prop. *(i)* $o_P(1) + O_P(1) = O_P(1)$,

(ii) $o_P(1) \cdot O_P(1) = o_P(1)$,

(iii) $o_P(1) + o_P(1) = o_P(1)$,

(iv) $\frac{1}{1+o_P(1)} = O_P(1)$.

Lemme de Slutsky et applications

Lem (de Slutsky). *Si* $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{\mathcal{L}} c$ *alors* $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$, $X_n Y_n \longrightarrow cX$ *et* $\frac{X_n}{Y_n} \longrightarrow \frac{X}{c}$ *(si* $c \neq 0$*).*

Delta-méthode

Soit $g \colon \mathbf{R}^d \to \mathbf{R}^m$ dérivable en un point $\nu \in \mathbf{R}^d$ de matrice jacobienne $\nabla g(\nu)$.

Rappel : $\lim_{h\to 0} \frac{\|g(\nu+g)-g(\nu)-\nabla g(\nu)\cdot h\|}{\|h\|} = 0$, $\nabla g(\nu) = \left(\frac{\partial g_i(\nu)}{\partial \nu_j}\right)_{i\in[\![1;d]\!], j\in[\![1;m]\!]}$.

Th. *Soit* $g \colon \mathbf{R}^d \to \mathbf{R}^m$ *dérivable en* ν. *Soient* T_n, T *des v.a. sur* $\mathbf{R}^d$ *et* (r_n) *une suite réelle telle que* $r_n \longrightarrow +\infty$, $r_n(T_n - \nu) \xrightarrow{\mathcal{L}} T$. *Alors* $r_n(g(T_n) - g(\nu)) \xrightarrow{\mathcal{L}} \nabla g(\nu) \cdot T$.

3 Statistique asymptotique

Not.
- $(\Omega, \mathcal{F}, \mathbf{P})$ espace de proba,
- $(X_i)_{i\in\mathbf{N}}$ v.a. iid
- $\forall i \in \mathbf{N}, X_i = (X_i^{(1)}, \dots, X_i^{(d)})^\mathsf{T}$,
- $\|\cdot\|$ norme euclidienne.

Introduction

Def. Estimateur $\hat{\theta}_n$ à valeurs dans $\Theta \subset \mathbf{R}^q$: transformation mesurable de $(X_1, \dots, X_n)$. $\hat{\theta}_n$ est faiblement consistant si $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$. $\hat{\theta}_n$ est fortement consistant si $\hat{\theta}_n \xrightarrow{\text{p.s.}} \theta_0$. $\hat{\theta}_n$ est asymptotiquement normal si $\sqrt{n}(\hat{\theta}_n - \theta_0) \Longrightarrow \mathcal{N}(0, \sigma_0^2)$.

Rem. La consistance est différente du biais. En effet, soit $\bar{X}^n = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{\theta}_n = \bar{X}^n + \frac{1}{n}$ est fortement consistant (si $\mathbf{E}(X_1) < \infty$) et biaisé car $\mathbf{E}(\hat{\theta}_n) - \mathbf{E}(X_1) = \frac{1}{n}$.

À l'inverse $\hat{\theta}_n = X_1$ est sans biais mais non consistant.

Def. $\hat{\theta}_n$ est un M-estimateur si $\hat{\theta}_n \in \arg\min_{\theta\in\Theta} M(\theta)$. $\hat{\theta}_n$ est un Z-estimateur si $\Psi_n(\hat{\theta}_n) = 0$.

Ex.
- Moindres carrés : $\hat{\beta}_n$ est défini par $\hat{\beta}_n = \arg\min_{\beta\in\mathbf{R}^d} \sum_{i=1}^n (Y_i - X_i^\mathsf{T}\beta)^2$.
- Maximum de vraisemblance : soit la famille paramétrique $\mathcal{P} = \{f_\theta \mid \theta \in \Theta\}$ selon lauelle est distribuée les données $(X_1, \dots, X_n)$.

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n} \log(f_\theta(X_i))$$

- Estimateur des moments et estimateur des moments généralisés : $\hat{\theta}_n \in \arg\min_{\theta\in\Theta} \left\|\frac{1}{n}\sum_{i=1}^n g(X_i) - \int g \, \mathrm{d}\mathbf{P}_\theta\right\|$.

Rem. Un Z-estimateur est toujours un M-estimateur car $\forall\theta \in \Theta, 0 = \left\|\Psi_n(\hat{\theta}_n)\right\| \leqslant \|\Psi_n(\theta)\|$. Un M-estimateur est un Z-estimateur si M_n est continuement dérivableh sur Θ et $\hat{\theta}_n$ est un point intérieur à Θ. Alors $\nabla M_n(\hat{\theta}_n) = 0$.

Prop (Consistance).

Prop (Consistance Z-estimateur). *Si* $\hat{\theta}_n$ *est un Z-estimateur et*

- $\sup_{\theta\in\Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{\mathbf{P} \text{ resp. p.s.}} 0$,
- $\forall\varepsilon > 0, \inf_{\theta\in\Theta\setminus B(\theta_0,\varepsilon)} \|\Psi(\theta)\| > \|\Psi(\theta_0)\|$,

alors $\hat{\theta}_n \xrightarrow{\mathbf{P} \text{ resp. p.s.}} \theta_0$.

Lem. *Supposons*

(i) Θ *compact,*

(ii) $\forall\theta \in \Theta, \mathbf{E}(|\rho(X_1, \theta)|) < \infty$,

(iii) $\exists r \colon \mathcal{X} \to \mathbf{R}_+, \mathbf{E}(r(X_1)) < \infty$ *où ...*