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 $\forall\omega$ **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**). *Soit $h\colon \mathbf{R}^d \to \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$:*

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

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

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 $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E}(X_1)$.*

2 Convergence en loi

Définitions et propriétés

Def. Une mesure de proba μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est caractérisée par sa **fonction de répartition**

$$F_\mu\colon (x_1, \dots, x_d) \mapsto \mu\left(\prod_i]-\infty\,; x_i]\right).$$

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

Th. *Deux mesures distinctes ne peuvent pas avoir la même fonction de répartition.*
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 $\lim_n F_{\mu_n}(x) = F_\mu(x)$ 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 $P_{X_n} \Rightarrow P_X$.

Prop. *Si $X_n \xrightarrow{\mathbf{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.*

Th (de représentation de **Skorohod**). *Soit $(\mu_n)_n, \mu$ des mesures de proba sur $\mathbf{R}^d$ telles que $\mu_n \Rightarrow \mu$. Il existe un espace de proba et des v.a. $(Y_n), Y$ sur cet espace à valeurs dans $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ telles que $Y \sim \mu$, $\forall n, Y_n \sim \mu_n$ et Y est limite simple des Y_n, i.e. $\forall \omega, Y_n(\omega) \longrightarrow Y(\omega)$.*

Th (de continuité). *Soit $h\colon \mathbf{R}^d \to \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$. Si $X_n \xrightarrow{\mathcal{L}} X$ alors $h \circ X_n \xrightarrow{\mathcal{L}} h \circ X$.*

Th. *Les affirmations suivantes sont équivalentes :*

(i) $\mu_n \Rightarrow \mu$,

(ii) pour toute fonction f continue, $\mu_n(f) \longrightarrow \mu(f)$,

(iii) pour toute fonction f lipschitzienne, $\mu_n(f) \longrightarrow \mu(f)$,

(iv) pour tout $A \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mu(\delta A) = 0$ (frontière de A), $\mu_n(A) \longrightarrow \mu(A)$.

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 $\mathbf{P}(X_n \in A) \longrightarrow \mathbf{P}(X \in A)$ où $\delta A = \bar{A} \setminus \mathring{A}$,

(iv) $\forall t \in \mathbf{R}^d, \lim_n \phi_{X_n}(t) = \phi_X(t)$.

Th. *Soit $m \in \mathbf{N}^*$ et $h\colon \mathbf{R}^d \to \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$. Si $\mu_n \Rightarrow \mu$ alors $\mu_n h^{-1} \Rightarrow \mu h^{-1}$.*

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

Mesures tendues

Lem (d'Helly). *Soit $(F_n)_n$ une suite de fonctions de répartition. Il existe une sous-suite $(\varphi(n))_n$ et $F\colon \mathbf{R} \to [0\,;1]$ croissante, continue à droite, telle que $\lim_n F_{\varphi(n)}(x) = F(x)$ 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, $\forall n, \mathbf{P}(X_n \in \mathcal{K}) \geqslant 1 - \varepsilon$.

Prop. *(i) Toute famille finie de mesures de probabilité est tendue.*

(ii) Une suite de mesures qui converge faiblement est tendue.

Th (de Prokhorov). *Soit une famille $\mathcal{M}$ de mesures de probabilité. Alors $\mathcal{M}$ est tendue si et seulement si elle est relativement séquentiellement compacte pour la topologie de la convergence faible, i.e. de toute suite $(\mu_n)_n$ de $\mathcal{M}$ on peut extraire une sous suite $(\mu_{\varphi(n)})_n$ telle que $\mu_{\varphi(n)} \Rightarrow \mu$ avec μ une mesure de probabilité.*

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

Fonction caractéristique

Def. La **fonction caractéristique** d'une mesure de proba μ sur $\mathbf{R}^d$ est $\phi_\mu\colon \begin{array}{ccc} \mathbf{R}^d & \to & \mathbf{C} \\ t & \mapsto & \int e^{i\langle t|x\rangle}\,\mathrm{d}\mu(x) \end{array}$.

Th. $\varphi_\mu = \varphi_\nu \implies \mu = \nu$.

Prop. *Soit μ une mesure sur $\mathbf{R}$ et $p \in \mathbf{N}$ tel que $\int |x|^p\,\mathrm{d}\mu(x) < \infty$. Alors ϕ_μ est p fois continument dérivable et $\forall t \in \mathbf{R}$, sa pe dérivées satisfait $\forall t \in \mathbf{R}, \phi_\mu^{(p)} = \int i^p x^p e^{itx}\,\mathrm{d}\mu(x)$.*

Ex. $\forall t \in \mathbf{R}, \phi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.

Prop. *Pour $Y = AX + b$ on a $\forall t, \phi_Y(t) = e^{i\langle t|b\rangle}\phi_X(A^* t)$.*

Prop. *ϕ_μ est continue en zéro.*

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

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 gaussien, sa fonction caractéristique vérifie $\phi_X(t) = e^{i\langle t|m\rangle} e^{-\frac{1}{2}\langle t|\Sigma t\rangle}$ où $m = \mathbf{E}(X)$ et $\Sigma = \mathrm{Cov}(X)$.

Th (central limite). *Soit $(X_n)_n$ iid tel que $\mathbf{E}(\|X_1\|^2) < \infty$. Alors*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbf{E}(X_1)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Cov}(X_1))\,.$$

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,*
- $\forall n, \forall i \leqslant n, \mathbf{E}(X_{i,n}) = 0$,
- $\exists \Sigma, \lim_n \sum_{i=1}^n \mathrm{Cov}(X_{i,n}) = \Sigma$
- *$\forall \varepsilon > 0, \lim_n \sum_{i=1}^n \mathbf{E}(\|X_{i,1}\|^2\, \mathbf{1}_{\|X_{i,n}\|} > \varepsilon) = 0$ (condition de Lindebergh).*

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

Lem. *Soit* $\mathcal{U} := \{z \in \mathbf{C} \mid |z| \leqslant 1\}$, $n \in \mathbf{N}^*$ *et deux familles* $(z_i)_{1\leqslant i\leqslant n}$ *et* $(w_i)_{1\leqslant i\leqslant n}$ *de* $\mathcal{U}$. *Alors* $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \leqslant \sum_{i=1}^n |z_i - w_i|$.

3 Manipulation des convergences

Quelques règles

Prop. (i) $\forall a \in \mathbf{R}^d, X_n \xrightarrow{\mathcal{L}} a \iff X_n \xrightarrow{\mathbf{P}} a$

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

(iii) $\forall a \in \mathcal{Y}, \left.\begin{array}{r} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} a \end{array}\right\} \Longrightarrow (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, a)$

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

Notation o_P, O_P

Not.
- On dit que $X_n = o_P(1)$ si $X_n \xrightarrow{\mathbf{P}} 0$.
- On dit que $X_n = o_P(Y_n)$ si $\exists (Z_n) \xrightarrow{\mathbf{P}} 0, X_n = Z_n Y_n$.
- On dit que $X_n = O_p(1)$ si (X_n) est borné en probabilité, i.e. (X_n) est tendue.
- On dit que $X_n = O_P(Y_n)$ si $\exists (Z_n) = O_p(1), X_n = Z_n Y_n$.

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

Lem (de **Slutsky**). *Soit* $(X_n)_n$ *et* $(Y_n)_n$ *tels que* $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{\mathcal{L}} a$ *:*

(i) *Pour* $a \in \mathcal{X}$ *quelconque,* $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$.

(ii) *Dans le cas réel* $a \in \mathbf{R}$, $X_n Y_n \longrightarrow cX$ *et* $\frac{X_n}{Y_n} \longrightarrow \frac{X}{c}$ *si* $a \neq 0$.

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* $\lim_n r_n = +\infty$ *et* $r_n(T_n - \nu) \xrightarrow{\mathcal{L}} T$. *Alors* $r_n(g(T_n) - g(\nu)) \xrightarrow{\mathcal{L}} \nabla g(\nu) \cdot T$.

4 Statistique asymptotique

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

Introduction

Def. Un **estimateur** $\hat\theta_n$ à valeurs dans $\Theta \subset \mathbf{R}^q$ est une transformation mesurable de $(X_1, \ldots, X_n)$.
- $\hat\theta_n$ 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 $\sqrt{n}(\hat\theta_n - \theta_0) \Longrightarrow \mathcal{N}(0, \sigma_0^2)$ avec $\sigma_0^2 > 0$.

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. Soit $\Theta \subset \mathbf{R}^q$.
- $\hat\theta_n$ 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 laquelle sont distribuées les données $(X_1, \ldots, X_n)$.

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

- Estimateur des moments par rapport à $\mathcal{P} = \{\mathbf{P}_\theta \mid \theta \in \Theta\}$ et $g\colon \mathcal{X} \to \mathbf{R}^p$: $\hat{\theta}_n$ est pris tel que

$$\frac{1}{n}\sum_{i=1}^n g(X_i) = \int g \,\mathrm{d}\mathbf{P}_{\hat{\theta}_n} = \mathbf{E}_{\hat{\theta}_n}(g(X_1)) \,.$$

- Estimateur des moments généralisés GMM (si pas de solution avec le précédent) :

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

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 dérivable sur Θ et $\hat{\theta}_n$ est un point intérieur à Θ. Alors $\nabla M_n(\hat{\theta}_n) = 0$.

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

- $\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \overset{\mathbf{P}\ \textit{resp. p.s.}}{\longrightarrow} 0$ *(convergence uniforme)*
- $\forall\varepsilon > 0, \inf_{\theta\in\Theta\setminus B(\theta_0,\varepsilon)} M(\theta) > M(\theta_0)$

alors $\hat{\theta}_n \overset{\mathbf{P}\ \textit{resp. p.s.}}{\longrightarrow} \theta_0$.

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

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

alors $\hat{\theta}_n \overset{\mathbf{P}\ \textit{resp. p.s.}}{\longrightarrow} \theta_0$.

En pratique on doit vérifier les 2 hypothèses des résultats précédents. Souvent $M_n(\theta) = \frac{1}{n}\sum_{i=1}^n \rho(X_i,\theta)$ et $\Psi_n(\theta) = \frac{1}{n}\sum_{i=1}^n \Psi(X_i,\theta)$.

Lem. *Soit Θ compact. Supposons*

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

(ii) $\exists r\colon \mathcal{X} \to \mathbf{R}_+$ *tel que* $\mathbf{E}(r(X_1)) < \infty$ *et* $\forall(\theta,\theta')\in\Theta^2, \forall x\in\mathcal{X}, |\rho(x,\theta) - \rho(x,\theta')| \leqslant r(x)\,\|\theta - \theta'\|$,

alors $\sup_{\theta\in\Theta} \left|\frac{1}{n}\sum_{i=1}^n \rho(X_i,\theta) - \mathbf{E}[\rho(X_1,\theta)]\right| \overset{p.s.}{\longrightarrow} 0$.

Lem (Vérification de la condition d'identifiabilité). *Soit $\Theta \subset \mathbf{R}^q$ compact et $M \in \mathcal{C}^0(\Theta)$ telle que θ_0 en est l'unique maximum. Alors $\forall\varepsilon > 0, \inf_{\theta\in\Theta\setminus B(\theta_0,\varepsilon)} M(\theta) > M(\theta_0)$.*

On peut ainsi facilement vérifier les conditions de la propriété de consistance.

Prop. *Soit Θ compact. Supposons $\mathbf{E}_{\theta_0}(\|g\|) < +\infty$ et $\hat{\theta}_n \in \arg\min_{\theta\in\Theta} \left\|\frac{1}{n}\sum_i g(X_i) - \mathbf{E}_\theta g\right\|$. Si, de plus, $\theta \mapsto \mathbf{E}_\theta g$ est injective et continue alors $\hat{\theta}_n \overset{p.s.}{\longrightarrow} \theta_0$.*

Normalité asymptotique

On considère ici uniquement les Z-estimateurs : $\frac{1}{n}\sum_{i=1}^n \psi(X_i,\hat{\theta}_n) = \Psi_n(\hat{\theta}_n) = 0$.

Th. *Supposons que :*

(i) $\hat{\theta}_n \overset{\mathbf{P}}{\longrightarrow} \theta$,

(ii) *il existe un voisinage $v(\theta_0)$ tel que $\forall x\in\mathcal{X}, \theta\mapsto\psi(x,\theta)$ est $\mathcal{C}^2$ sur $v(\theta_0)$ et $\forall k\in[\![1\,;q]\!]$,* $\mathbf{E}\left[\sum_{\theta\in v(\theta_0)} \|\psi_k(X_1,\theta)\|\right] < +\infty$,

(iii) *soit* $\Phi\colon (x,\theta) \mapsto \begin{pmatrix} \nabla_\theta\psi_1(x,\theta) \\ \vdots \\ \nabla_\theta\psi_q(x,\theta) \end{pmatrix} \in \mathbf{R}^{q\times q}$ *tel que* $\mathbf{E}\left[\|\Phi(X_1,\theta_0)\|\right] < +\infty$ *et* $\mathbf{E}[\Phi(X_1,\theta_0)] = Q(\theta_0)$ *est inversible,*

(iv) $\mathbf{E}[\|\psi(X_1,\theta_0)\|^2] < +\infty$.

Alors $\sqrt{n}(\hat{\theta}_n - \theta_0) = -Q(\theta_0)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \psi(X_i,\theta_0)\right) + o_P(1)$. *En particulier*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(0, Q(\theta_0)^{-1}\operatorname{Var}(\psi(X_1,\theta_0))(Q(\theta_0)^{-1})^\mathsf{T}\right) \,.$$

5 Le bootstrap

Cache de travail et principe général du Bootstrap

Les estimateurs statistiques les plus utilisés sont de la forme $\hat{\theta}_n = \theta(\mathbf{P}_n)$ où $\mathbf{P}_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}(x)$.

On a $\theta_0 = \theta(\mathbf{P}_0)$ et $\hat{\theta}_n = \theta(\mathbf{P}_n)$. On sait qu'en général $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(O,\sigma)$, avec $\mathbf{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $X_i \sim \mathbf{P}_0$.

Def. Le **bootstrap** a pour but premier de reproduire le comportement de $\sqrt{n}(\hat{\theta}_n - \theta_0)$ pour décrire la précision de $\hat{\theta}_n$, avec $\sqrt{n}\left(\theta(\mathbf{P}_n^*) - \theta(P_n)\right)$ où $\mathbf{P}_n^* = \frac{1}{n}\sum_{i=1}^n \delta_{X_{i,n}^*}$ et $X_{i,n}^* \sim \mathbf{P}_n$.

Le plug-in principle : même procédure que la procédure d'estimation classique mais en remplaçant $\mathbf{P}_0$ par $\mathbf{P}_n$. Comme $\mathbf{P}_n$ est connue, on peut reproduire cette procédure autant de fois que l'on souhaite et calculer ainsi $\theta_1^* = \theta(\mathbf{P}_{n,1}^*), \theta_2^* = \theta(\mathbf{P}_{n,2}^*), \ldots, \theta_B^* = \theta(\mathbf{P}_{n,B}^*)$.

Deux étape :

- définition : $\theta(\mathbf{P}_n^*)$,
- simulation : $\theta_1^*, \ldots, \theta_B^*$.

Algorithme : on a $X_1, \ldots, X_n$ et θ.

1) calcul de $\mathbf{P}_n$,
2) calcul de $\theta(\mathbf{P}_n) = \hat{\theta}_n$,
3) bootstrap :
   - tirage de $X_{1,n}^*, X_{2,n}^*, \ldots, X_{n,n}^*$ iid selon $\mathbf{P}_n$,
   - $\mathbf{P}_n^* = \frac{1}{n}\sum_{i=1}^n \delta_{X_{i,n}^*}$
   - $\theta_n^* = \theta(\mathbf{P}_n^*)$,
4) simulation : calcul de $\theta_{n,1}^*, \ldots, \theta_{n,B}^*$ (itération du point 3). Heuristique :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\theta(\mathbf{P}_n) - \theta(\mathbf{P}_0)) \simeq \sqrt{n}(\theta(\mathbf{P}_n^*) - \theta(\mathbf{P}_n)) = \sqrt{n}(\theta_n^* - \hat{\theta}_n)$$

5) On possède B versions de $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$ utilisées pour approcher $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Ex (Moyenne empirique). $\theta_0 = \int x \, \mathrm{d}\mathbf{P}_0(x)$, $\quad \hat{\theta}_n = \int x \, \mathrm{d}\mathbf{P}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i$, $\quad \theta_n^* = \int x \, \mathrm{d}\mathbf{P}_n^*(x) = \frac{1}{n}\sum_{i=1}^n X_{i,n}^*$.

Rem. Tirer selon $\mathbf{P}_n$ revient à tirer uniformément dans $(X_1, \ldots, X_n)$. Donc $\theta_n^* = \sum_{i=1}^n N_{i,n} X_i$ avec $N_{i,n}$ une v.a. à valeurs dans $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}$.

Utilisation des $\theta_{n,1}^*, \ldots, \theta_{n,B}^*$: comme $\sqrt{n}(\theta_{n,b}^* - \hat{\theta}_n) = R_{n,b}^*$ ont une loi similaire à $\sqrt{n}(\hat{\theta}_n - \theta_0)$ on les utilise pour calculer les quantiles.

On définit $F_B(t) = \frac{1}{B}\sum_{b=1}^B \mathbf{1}_{\{R_{n,b}^* \leqslant t\}}$.

Intervalle de confiance : $IC_{\text{bootstrap}} = \left[\hat{\theta}_n - \frac{1}{\sqrt{n}} F_B^{-1}\left(1 - \frac{\alpha}{2}\right); \hat{\theta}_n - \frac{1}{\sqrt{n}} F_B^{-1}\left(\frac{\alpha}{2}\right)\right]$ comme l'IC asymptotique sauf que le quantile n'est pas calculé de la même façon. En asymptotique c'est $\sqrt{\frac{n}{\sigma}}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. Avec bootstrap $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \sqrt{n}(\theta_b^* - \hat{\theta}_n)$.

On peut prendre B suffisamment grand tel que F_B est la fonction de répartition de $R_{n,1}^*$ (les $R_{n,1}^*, \ldots, R_{n,B}^*$ sont iid dans l'espace de proba conditionnel aux $X_1, \ldots, X_n$). On néglige l'erreur de simulation car on peut prendre B très grand.

La question qui reste est : les quantiles de $R_{n,1}^*$ sont-ils des quantiles asymptotiquement consistants ? C'est-à-dire, en définissant $\xi_n(\alpha)$ comme étant ce quantile de niveau α, a-t-on $\lim_n \mathbf{P}\left(\xi_n\left(\frac{\alpha}{2}\right) \leqslant \sqrt{n}(\hat{\theta}_n - \theta_0) \leqslant \xi_n\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$?

Lem. *Soit F_n la fonction de répartition de $R_{n,1}^*$. Supposons*

- *$\forall x \in \mathbf{R}, F_n(x) \xrightarrow{p.s.} \Phi_\sigma(x)$ où $\Phi_\sigma(x) = \frac{\Phi(x/\sigma)}{\sigma}$ avec Φ la loi normale standard, donc Φ_σ est la loi normale centrée de variance σ^2,*
- *$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$*

alors $\lim_n \mathbf{P}\left(\xi_n\left(\frac{\alpha}{2}\right) \leqslant \sqrt{n}(\hat{\theta}_n - \theta_0) \leqslant \xi_n\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$ est vérifiée.

Th (TCL bootstrap). *Soit $(X_n)_n$ iid réelles tel que $\mathbf{E}(X_1^2) < \infty$ alors $\sqrt{n}(\hat{\theta}_n - \theta_0)$ a la même loi asymptotique que $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$ avec $\theta_0 = \mathbf{E}(X_1)$, $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\theta_n^* = \frac{1}{n}\sum_{i=1}^n X_{i,n}^*$.*