

Lois Limites Fonctionnelles pour le Processus Empirique et Applications

Sarah Ouadah

LSTA, UPMC, Paris VI

CIRM, 19 avril 2012

- 1 Histoire de Lois Limites Fonctionnelles
- 2 Estimation non paramétrique de la densité
- 3 Contributions
- 4 Bibliographie

Lévy (1948)

Soit $\{W(t) : t \geq 0\}$ un processus de Wiener standard

Loi du logarithme itéré

$$\limsup_{T \rightarrow \infty} \frac{|W(T)|}{\sqrt{2T \log \log T}} = 1 \text{ p.s.}$$

Strassen (1964)

Pour tout $x \in [0, 1]$,

$$\eta_n(x) := \frac{W(nx)}{\sqrt{2n \log \log n}} \in C(0, 1)$$

Ensemble de Strassen

$$\mathbb{S} := \left\{ f \in \mathcal{AC}[0, 1], f(0) = 0, \left\{ \int_0^1 \dot{f}(u)^2 du \right\}^{1/2} \leq 1 \right\}$$

Loi limite fonctionnelle

$$\eta_n(\cdot) \rightsquigarrow \mathbb{S}$$

$$\eta_n(\cdot) \rightsquigarrow \mathbb{S} :$$

- $\forall n_k, k \geq 1, \exists n_{k_j}, \exists g \in \mathbb{S}$ tel que

$$\eta_{n_{k_j}}(x) \rightarrow g(x) \text{ uniformément en } x \in [0, 1]$$

- $\forall g \in \mathbb{S}, \exists n_k = n_k(g)$ tel que

$$\eta_{n_k}(x) \rightarrow g(x) \text{ uniformément en } x \in [0, 1]$$

Finkelstein (1971)

Soient U_1, U_2, \dots des variables aléatoires iid uniformes sur $(0, 1)$.

Fonction de répartition empirique basée sur U_1, \dots, U_n :

$$\mathbb{U}_n(u) := n^{-1} \#\{U_i \leq u : 1 \leq i \leq n\}, u \in \mathbb{R}$$

Processus empirique uniforme

$$\alpha_n(u) := n^{1/2} (\mathbb{U}_n(u) - u), u \in \mathbb{R}$$

Pour tout $x \in [0, 1]$,

$$v_n(x) := \frac{\alpha_n(x)}{\sqrt{2n \log \log n}} \in \mathcal{F}(0, 1)$$

Ensemble de Finkelstein

$$\mathbb{F} := \{f \in \mathbb{S}, f(1) = 0\}$$

Loi limite fonctionnelle

$$v_n(\cdot) \rightsquigarrow \mathbb{F}$$

Deheuvels et Mason (1992)

Fonction d'incrément : pour tout $h > 0$ et $t \in [0, 1]$,

$$\xi_n(h, t; u) := \alpha_n(t + hu) - \alpha_n(t), u \in \mathbb{R}$$

Loi limite fonctionnelle

$$\frac{\xi_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} \rightsquigarrow \mathbb{S}$$

Estimateurs à noyau de la densité

Soit X_1, X_2, \dots une suite de variables aléatoires réelles i.i.d. selon $f(\cdot)$ continue sur $J \subseteq \mathbb{R}$

Estimateur à noyau de la densité de Parzen-Rosenblatt (1964)

$$f_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), x \in \mathbb{R}$$

avec

- K la fonction noyau
- $h > 0$ la fenêtre

On étudie

$$f_{n,h}(x) - f(x) = \underbrace{f_{n,h}(x) - \mathbb{E}f_{n,h}(x)}_{(1)} + \underbrace{\mathbb{E}f_{n,h}(x) - f(x)}_{(2)}$$

(1) : terme aléatoire

(2) : terme déterministe (biais)

- L'ordre du biais dépend des propriétés de $f(\cdot)$ et du noyau $K(\cdot)$
- L'ordre du terme aléatoire est évalué dans le Théorème 2

Le lien

$$n^{1/2}h(f_{n,h}(x) - \mathbb{E}f_{n,h}(x)) = - \int_0^1 \xi_n(hf(x_0); F(x); \tau(u)) d\tilde{K}(u),$$

où

$$\tau(u) = \frac{F(x + hu) - F(x)}{hf(x_0)} \text{ et } \tilde{K}(u) := K(-u), 0 \leq u \leq 1$$

Deheuvels et Mason (1992)

Soit

$$0 < h_n < 1, \quad h_n \downarrow 0, \quad nh_n \uparrow \infty, \quad \frac{nh_n}{\log n} \rightarrow \infty \text{ et } \frac{\log(1/h_n)}{\log \log n} \rightarrow \infty$$

Théorème limite

Lorsque $n \rightarrow \infty$,

$$\left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{f_{n,h_n}(x) - \mathbb{E}f_{n,h_n}(x)\} = \sigma(f, K) \text{ p.s.}$$

Choix de la fenêtre

Problème : Compromis biais-variance

Solution : “Uniformité” en la fenêtre h

Einmahl et Mason (2005)

$$\limsup_{n \rightarrow \infty} \left\{ \frac{nh}{\log(1/h) \vee \log \log n} \right\}^{1/2} \sup_{\frac{r \log n}{n} \leq h \leq 1} \left(\sup_{x \in I} |f_{n,h}(x) - \mathbb{E}f_{n,h}(x)| \right) < \infty$$

Soient $0 < a_n \leq b_n \leq 1$ telles que,

$$b_n \rightarrow 0 \quad \text{et} \quad \frac{na_n}{\log n} \rightarrow \infty \quad (1)$$

$\mathcal{H}_n = [a_n, b_n]$ et $\mathcal{I} := [u, v] \subseteq [0, 1]$, $u < v$

Pour tout $h \in \mathcal{H}_n$,

$$\mathcal{F}_{n;\mathcal{I}}(h) := \left\{ \frac{\xi_n(h; t; \cdot)}{\sqrt{2h \log_+(1/h)}} : t \in [0, 1 - h] \cap \mathcal{I} \right\},$$

où $\log_+ s = \log(s \vee e)$, $s \in \mathbb{R}$

Distance de Hausdorff

Pour tout $A \subseteq \mathcal{B}[0, 1]$

$$A^\varepsilon = \{g \in \mathcal{B}[0, 1] : \exists f \in A, \|f - g\| < \varepsilon\},$$

où $\|f\| = \sup_{0 \leq t \leq 1} |f(t)|$

Distance de Hausdorff entre deux ensembles $A, B \subseteq \mathcal{B}[0, 1]$

$$\Delta(A, B) = \begin{cases} \inf \{\varepsilon > 0 : A \subseteq B^\varepsilon \text{ et } B \subseteq A^\varepsilon\} & \text{lorsqu'un tel } \varepsilon \text{ existe} \\ \infty & \text{sinon} \end{cases}$$

Loi limite fonctionnelle uniforme

Théorème 1 (*Deheuvels et Ouadah (2011)*)

Soit $\mathcal{H}_n = [a_n, b_n]$ vérifiant la condition (1). Alors, pour $\mathcal{H}_n = [a_n, b_n]$ et tout $\mathcal{I} = [u, v] \subseteq [0, 1]$, lorsque $n \rightarrow \infty$, on a

$$\sup_{h \in \mathcal{H}_n} \Delta(\mathcal{F}_{n;\mathcal{I}}(h), \mathbb{S}) = o_{\mathbb{P}}(1)$$

Théorème limite

Théorème 2 (*Deheuvels et Ouadah (2011)*)

Soit $\mathcal{H}_n = [a_n, b_n]$ vérifiant la condition (1). Alors, lorsque $n \rightarrow \infty$,

$$\sup_{h \in \mathcal{H}_n} \left| \left\{ \frac{nh}{2 \log_+(1/h)} \right\}^{1/2} \sup_{x \in I} \pm \{f_{n,h}(x) - \mathbb{E}f_{n,h}(x)\} - \sigma(f, K) \right| = o_{\mathbb{P}}(1)$$

où $\sigma(f, K) = \sup_{x \in I} \left\{ f(x) \int_{\mathbb{R}} K^2(t) dt \right\}^{1/2}$






Éléments de preuve du Théorème 1

1°) Poissonisation

$$\mathbb{P} \left(\frac{\xi_n(h, t; \cdot)}{\sqrt{2h \log_+(1/h)}} \notin S^\epsilon \right) \leq 2\mathbb{P} \left(\frac{\xi_n^*(h, t; \cdot)}{\sqrt{2h \log(1/h)}} \notin S^\epsilon \right)$$

2°) Schilder+KMT

$$\begin{aligned} \mathbb{P} \left(\frac{\xi_n^*(h, t; \cdot)}{\sqrt{2h \log(1/h)}} \notin S^\epsilon \right) &\leq \mathbb{P} \left(\frac{W(nh \cdot)}{\sqrt{2nh \log(1/h)}} \notin S^{\epsilon_1} \right) \\ &+ \mathbb{P} \left(\sup_{0 \leq s \leq 1} \left| \frac{W(nhs) - \{\Pi(nhs) - nhs\}}{\sqrt{2nh \log(1/h)}} \right| \geq \epsilon_2 \right) \end{aligned}$$

-  Deheuvels, P. et Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.* **20**, 1248–1287.
-  Deheuvels, P. et Ouadah, S. (2011). Uniform in Bandwidth Functional Limit Laws. *Journal of theo. proba.*, à paraître.
-  Einmahl, U. et Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380–1403.
-  Finkelstein, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42**, 607–615.
-  Lévy, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris. 365 pp.