Nouvelle approche pour les processus empiriques de variables dépendantes

Olivier Durieu*

en collaboration avec Herold Dehling** et Dalibor Volný*

*Université de Rouen - Laboratoire de Mathématiques Raphaël Salem

**Fakultät für Mathematik - Ruhr-Universität Bochum

Journées de Probabilités, Lille
1-5 septembre 2008

- Soit $(X_n)_{n\geq 0}$ un processus aléatoire stationnaire à valeurs dans $[0,1]$.
- $F(t) := P(X_0 \leq t)$.
- La fonction de répartition empirique $(F_n(t))_{0\leq t\leq 1}$ et le processus empirique $(U_n(t))_{0\leq t\leq 1}$ sont définis par

$$
\begin{aligned}
F_n(t) &:= \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(X_i),\ 0\leq t\leq 1,\\
U_n(t) &:= \sqrt{n}(F_n(t)-F(t)),\ 0\leq t\leq 1.
\end{aligned}
$$

cas i.i.d.

Théorème (Donsker, 1952)

Si $(X_n)_{n\geq 0}$ est i.i.d. et si F est continu, alors

$$(U_n(t))_{0\leq t\leq 1} \xrightarrow{\mathcal{L}} (B(t))_{0\leq t\leq 1}.$$

cas i.i.d.

Théorème (Donsker, 1952)

Si $(X_n)_{n\geq 0}$ est i.i.d. et si F est continu, alors

$$(U_n(t))_{0\leq t\leq 1} \xrightarrow{\mathcal{L}} (B(t))_{0\leq t\leq 1}.$$

Démo :

TLC fini-dimensionnel + Tension.

avec l'inégalité de 4$^{\text{ème}}$ moment : pour tout $s < t$,

$$\mathbb{E}\left(U_n(t) - U_n(s)\right)^4 \leq C\left(\frac{F(t) - F(s)}{n} + \left(F(t) - F(s)\right)^2\right).$$

Variables dépendantes

- processus φ-mélangeants : Billingsley (1968)
- processus α-mélangeants : Berkes, Phillip (1977)
- processus absolument réguliers : Douhkan, Massart, Rio (1995)
- autres conditions de dépendence faible : Douhkan, Louichi, Prieur, Dedecker, Wu ...
- applications dilatantes de l'intervalle : Collet, Martinez et Schmitt (2004)

Hypothèse 1

Pour toute fonction lipschitzienne f, le TLC a lieu, *i.e.*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - \mathbb{E}f(X_i)\} \xrightarrow{\mathcal{L}} N(0, \sigma^2), \tag{1}$$

où $N(0, \sigma^2)$ est la loi normale centrée de variance

$$\sigma^2 = \mathbb{E}(f(X_0) - \mathbb{E}f(X_0))^2 + 2 \sum_{i=1}^{\infty} \mathrm{Cov}(f(X_0), f(X_i)).$$

Hypothèse 2

Une inégalité de 4ème moment :
Pour toute fonction lipschitzienne bornée f telle que $\mathbb{E}(f(X_0)) = 0$,

$$\mathbb{E}\left\{\sum_{i=1}^{n} f(X_i)\right\}^4 \leq Cm_f^3\left(n\|f(X_0)\|_1 \log^{\alpha}(1+\|f\|) + n^2\|f(X_0)\|_1^2 \log^{\beta}(1+\|f\|)\right) \tag{2}$$

où C est une constante, α et β sont des entiers positifs,

$$\|f\| = \sup_x \mid f(x) \mid + \sup_{x\neq y} \frac{\mid f(x)-f(y) \mid}{\mid x-y \mid}$$

et

$$m_f = \max\{1, \sup_x \mid f(x) \mid\}.$$

Théorème (Dehling, D., Volný, 2008)

Soit $(X_i)_{i\geq 0}$ *un processus stationnaire à valeurs dans* $[0,1]$ *tel que les conditions (1) et (2) aient lieu. On suppose que*

$$\omega_F(\delta) \leq D|\log(\delta)|^{-\gamma}$$

pour un certain $D > 0$ *et* $\gamma > \max\left\{\frac{\alpha}{2}, \beta\right\}$. *Alors*

$$(U_n(t))_{0\leq t\leq 1} \xrightarrow{\mathcal{L}} (W(t))_{0\leq t\leq 1},$$

où $W(t)$ *est un processus gaussien centré de covariances*

$$\begin{aligned}
\mathbb{E}W(s)\cdot W(t) \quad = \quad & \mathrm{Cov}(1_{[0,s]}(X_0), 1_{[0,t]}(X_0)) \\
& + \sum_{k\geq 1} \mathrm{Cov}(1_{[0,s]}(X_0), 1_{[0,t]}(X_k)) \\
& + \sum_{k\geq 1} \mathrm{Cov}(1_{[0,s]}(X_k), 1_{[0,t]}(X_0)).
\end{aligned}$$

Démonstration

Théorème

Soit (S, ρ) un espace métrique complet et séparable . Soit $Y_n, Y_n^{(m)}$ et $Y^{(m)}$, $n, m \geq 1$ des variables aléatoires à valeurs dans S vérifiant

$$Y_n^{(m)} \xrightarrow{\mathcal{L}} Y^{(m)} \text{ quand } n \to \infty, \forall m$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(\rho(Y_n, Y_n^{(m)}) \geq \varepsilon) = 0, \forall \varepsilon > 0.$$

Alors il existe une variable aléatoire Y à valeurs dans S telle que

$$Y_n \xrightarrow{\mathcal{L}} Y \text{ quand } n \to \infty.$$

De plus, $Y^{(m)} \xrightarrow{\mathcal{L}} Y$ quand $m \to \infty$.

Démonstration

Soit une subdivision régulière

$$0 = t'_0 < \ldots < t'_m = 1$$

on définit

$$t_j = F^{-1}(t'_j)$$

où F^{-1} est donné par

$$F^{-1}(t) = \sup\{s \in [0,1] : F(s) \le t\}.$$

On a la subdivision

$$0 \le t_0 < \ldots < t_m = 1.$$

Démonstration

On considère les fonctions $\varphi_j : [0,1] \longrightarrow \mathbb{R}$ définies par

$$\varphi_j(x) = \varphi\left(\frac{x - t_{j-1}}{t_{j-1} - t_{j-2}}\right), \qquad \text{pour } j = 2, \ldots, m$$

où

$$\varphi(x) = 1_{(-\infty,-1]}(x) - x1_{(-1,0]}(x)$$

et $\varphi_1 \equiv 0$.

La fonction φ_j sert d'approximation lipschitzienne de la fonction indicatrice $1_{(-\infty,t_{j-1}]}(x)$.

Démonstration

On définit le processus

$$
\begin{aligned}
F_n^{(m)}(t) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} 1_{[t_{j-1},t_j)}(t)\varphi_j(X_i) \\
&= \sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i)\right) 1_{[t_{j-1},t_j)}(t).
\end{aligned}
$$

Pour $t \in [t_{j-1}, t_j]$, on a

$$F_n(t_{j-2}) \leq F_n^{(m)}(t) \leq F_n(t_{j-1}).$$

On définit le processus centré et normalisé

$$U_n^{(m)}(t) = \sqrt{n}\left(F_n^{(m)}(t) - \mathbb{E}F_n^{(m)}(t)\right).$$

Démonstration

Proposition

Pour toute subdivision $0 = t'_0 < \ldots < t'_m = 1$, *il existe un processus gaussien constant par morceaux* $\left(W^{(m)}(t)\right)_{0 \le t \le 1}$ *tel que*

$$\left(U_n^{(m)}(t)\right)_{0\le t\le 1} \xrightarrow{\mathcal{D}} \left(W^{(m)}(t)\right)_{0\le t\le 1}.$$

Démonstration

Proposition

Pour toute subdivision $0 = t'_0 < \ldots < t'_m = 1$, *il existe un processus gaussien constant par morceaux* $\left(W^{(m)}(t)\right)_{0\leq t\leq 1}$ *tel que*

$$\left(U_n^{(m)}(t)\right)_{0\leq t\leq 1} \xrightarrow{\mathcal{D}} \left(W^{(m)}(t)\right)_{0\leq t\leq 1}.$$

Proposition

Pour tout $\varepsilon, \eta > 0$ *il existe une subdivision* $0 = t'_0 < \ldots < t'_m = 1$ *telle que*

$$\limsup_{n\to\infty} P\left(\sup_{0\leq t\leq 1}\left|U_n(t) - U_n^{(m)}(t)\right| > \varepsilon\right) \leq \eta.$$

Démonstration de la Proposition

Sur l'intervalle $[t_{j-1}, t_j]$ on introduit une suite de subdivisions

$$t_{j-1} = s_0^{(k)} < s_1^{(k)} < \ldots < s_{2^k}^{(k)} = t_j$$

où

$$s_\ell^{(k)} = F^{-1}(t'_{j-1} + \ell \cdot \frac{h}{2^k}) \quad , \quad 0 \leq \ell \leq 2^k, k \geq 0.$$

Démonstration de la Proposition

Sur l'intervalle $[t_{j-1}, t_j]$ on introduit une suite de subdivisions

$$t_{j-1} = s_0^{(k)} < s_1^{(k)} < \ldots < s_{2^k}^{(k)} = t_j$$

où

$$s_\ell^{(k)} = F^{-1}(t'_{j-1} + \ell \cdot \frac{h}{2^k}) \quad , \quad 0 \leq \ell \leq 2^k, k \geq 0.$$

Pour tout $t \in [t_{j-1}, t_j)$ et $k \geq 0$ on définit

$$\ell(k,t) = \max\left\{\ell : s_\ell^{(k)} \leq t\right\}.$$

On obtient une chaîne

$$t_{j-1} = s_{\ell(0,t)}^{(0)} \leq s_{\ell(1,t)}^{(1)} \leq \ldots \leq s_{\ell(k,t)}^{(k)} \leq t \leq s_{\ell(k,t)+1}^{(k)},$$

On définit les fonctions $\psi_\ell^{(k)}$, $k \geq 0$, $0 \leq \ell \leq 2^k$, par

$$\psi_\ell^{(k)}(x) = \varphi\left(\frac{x - s_\ell^{(k)}}{s_\ell^{(k)} - s_{\ell-1}^{(k)}}\right).$$

On définit les fonctions $\psi_\ell^{(k)}$, $k \geq 0$, $0 \leq \ell \leq 2^k$, par

$$\psi_\ell^{(k)}(x) = \varphi\left(\frac{x - s_\ell^{(k)}}{s_\ell^{(k)} - s_{\ell-1}^{(k)}}\right).$$

Alors

$$\begin{aligned}\varphi_j(x) = \psi_{\ell(0,t)}^{(0)}(x) &\leq \psi_{\ell(1,t)}^{(1)}(x) \\ &\leq \dots \\ &\leq \psi_{\ell(K,t)}^{(K)}(x) \leq 1_{(-\infty,t]}(x) \leq \psi_{\ell(K,t)+2}^{(K)}(x).\end{aligned}$$

On étudie $|U_n(t) - U_n^{(m)}(t)|$ à travers la décomposition

$$\begin{aligned}
F_n(t) - F_n^{(m)}(t) &= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left(\psi_{\ell(k,t)}^{(k)}(X_i) - \psi_{\ell(k-1,t)}^{(k-1)}(X_i) \right) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \left(1_{(-\infty,t]}(X_i) - \psi_{\ell(K,t)}^{(K)}(X_i) \right).
\end{aligned}$$

et on utilise l'inégalité de 4ème moment.

Application

Soit $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité et $\mathcal{L}$ l'espace de fonctions lipschitziennes bornées de Ω dans $\mathbb{R}$ muni de la norme lipschitzienne.

Application

Soit $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité et $\mathcal{L}$ l'espace de fonctions lipschitziennes bornées de Ω dans $\mathbb{R}$ muni de la norme lipschitzienne.

Une chaîne de Markov $(X_k)_{k\geq 0}$ d'opérateur Q est dite $\mathcal{L}$-géométriquement ergodique s'il existe $C > 0$ et $0 < \theta < 1$ tels que pour tout $f \in \mathcal{L}$,

$$\|Q^k f - \pi f\| \leq C\theta^k \|f\|,$$

où $\pi f = \mathbb{E}(f(X_0))\mathbf{1}$.

Proposition (D.)

Si $(X_n)_{n\geq 0}$ *est une chaîne de Markov* $\mathcal{L}$*-géométriquement ergodique, alors (2) a lieu avec* $\alpha = 3$ *et* $\beta = 2$*, pour tout* $f \in \mathcal{L}$ *tel que* $\mathbb{E}\left(f(X_0)\right) = 0$*.*

Exemple : processus linéaires

Corollary

Soit $(X_k)_{k\geq 0}$ *un processus linéaire réel défini par une suite de formes linéaires* $(a_i)_{i\geq 0}$ *et une suite de variables aléatoires bornéés* $(e_i)_{i\in\mathbb{Z}}$ *définies sur un espace de Banach A. Supposons qu'il existe* $C > 0$ *et* $0 < \theta < 1$ *tels que*

$$|a_i| \leq C\theta^i,$$

et la fonction de répartition F de X_0 *vérifie*

$$\omega_F(\delta) \leq D|\log(\delta)|^{-\gamma} \text{ pour un } D > 0 \text{ et un } \gamma > 2.$$

Alors $(U_n(t))_{0\leq t\leq 1}$ *converge en loi vers un processus gaussien.*