Le Théorème de la Limite Centrale pour des variables dépendantes.

Travail en collaboration avec Paul Doukhan

Olivier Wintenberger, CEREMADE, Université Paris Dauphine
owintenb@ceremade.dauphine.fr

Lille, 05 septembre 2008

Théorème de la limite centrale (TLC)

On considère une suite stationnaire et ergodique de variables aléatoires centrées $(X_k)_{k\in\mathbb{N}}$. Quelles sont les conditions pour qu'il existe $\sigma^2 \geq 0$ tel que

$$\frac{S_n}{\sqrt{n}} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2). \qquad \text{(TLC)}$$

Dans le cas indépendant, CNS de Lindeberg $\mathbb{E}X_0^2 < +\infty$.

Plan

1. Exemples
2. Approximation par différence de martingale
3. Autres méthodes
4. TLC pour des suites faiblement dépendantes

Que peut-il arriver sans l'hypothèse d'indépendance?

- $\sigma^2 = 0$ si on considère la suite 2-dépendante $X_i = Y_i - Y_{i-1}$, Y_i iid telle que $\mathrm{E}|Y_0| < \infty$.
- Si $(X_i)_{i\in\mathbb{N}}$ est un processus gaussien, CNS : $\sum_{i=0}^{\infty} \mathrm{Cov}(X_0, X_r) < \infty$.
- Il existe un processus borné indépendant 2 à 2 ($\sum_{i\geq 1} \mathrm{Cov}(X_0, X_i) = 0$) ne vérifiant pas (TLC), Bradley, 1989.
- Pour n'importe quel type de mélange, il existe une chaîne de Markov $(X_i)_{i\in\mathbb{N}}$ telle que $\mathrm{E}X_0^2 < \infty$ mais ne vérifiant pas (TLC) car $S_n \approx \sum_{i=1}^{p} Y_i$ avec Y_i iid et $\mathrm{E}Y_i^2 = \infty$ (Doukhan *et al.*, 1994, Bradley, 1997).

Quelles conditions de dépendance et de moment?

Plan

1. Exemples
2. Approximation par différence de martingale
3. Autres méthodes
4. TLC pour des suites faiblement dépendantes

Différence de martingale

Definition (Différence de martingale)

$(M_k)_{k\in\mathbb{N}}$ *est telle que* M_k *est mesurable par rapport à* $\mathcal{F}_k$ *:*

$$E(M_k|\mathcal{F}_{k-1}) = 0 \textit{ presque sûrement,}$$

et $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$ *pour tout* $k > 0$*.*

Condition pour (TLC) dans le cas stationnaire et ergodique (Billingsley, Ibragimov, 1961) : $\mathrm{E}M_0^2 < \infty$.

Remarque : ici $\sigma^2 = \mathrm{E}M_0^2$.

Approximation par différence de martingale

Le processus $(X_k)_{k\in\mathbb{N}}$ est stationnaire et ergodique.

Idée de Gordin (1969) : si

$$X_k = M_k + Z_k - Z_{k-1}, \qquad k > 0,$$

avec $(M_k)_{k\in\mathbb{N}}$ une différence de martingale avec $\sigma^2 = \mathrm{E}M_0^2 < \infty$ et $(Z_k)_{k\in\mathbb{N}}$ stationnaire telle que $\mathrm{E}|Z_0| < \infty$, alors (TLC) est vérifié.

Plus généralement, si $T_n = \sum_{k=1}^n M_k$ et

$$n^{-1}\mathrm{E}(S_n - T_n)^2 \to 0 \text{ quand } n \to \infty,$$

alors (TLC) est vérifié (Hall et Heyde, 1980).

Conditions d'approximation par différence de martingale

Gordin (et Heyde) (1975)

$\sum_{n=0}^{\infty} \mathrm{E}(X_n|\mathcal{F}_0)$ et $\sum_{n=0}^{\infty}(X_{-n} - \mathrm{E}(X_{-n}|\mathcal{F}_0))$ convergent dans $\mathbf{L}^1$ et

$$\lim_{n\to\infty} \sup n^{-1/2}\mathrm{E}|S_n| < \infty.$$

Maxwell et Woodroofe (2000)

$$\sum_{n\geq 1} \frac{\|\mathrm{E}(S_n|\mathcal{F}_0)\|_2}{n^{3/2}} < \infty.$$

Heyde (et Hannan) (1973)

$\sum_{n>0}(\mathrm{E}(X_n|\mathcal{F}_1) - \mathrm{E}(X_n|\mathcal{F}_0))$ converge dans $\mathbf{L}^2$ et

$$n^{-1}\mathrm{E}S_n^2 \to \sigma^2 \text{ quand } n \to \infty.$$

Plan

1. Exemples
2. Approximation par différence de martingale
3. Autres méthodes
4. TLC pour des suites faiblement dépendantes

Méthode de Lindeberg

Dedecker (1998)

Si $\sum_{n=0}^{\infty} X_0 \mathrm{E}(X_n|\mathcal{F}_0)$ converge dans $\mathbf{L}^1$, alors $\exists\, \sigma \geq 0$ tel que

$$\lim_{n\to\infty} \mathrm{E}(h(n^{-1/2}S_n)) = \mathrm{E}(h(\sigma N))$$

pour tout h 3-fois continument dérivables et N distribuée comme $\mathcal{N}(0,1)$.

En particulier, (TLC) est vérifié.

Méthode de Lindeberg

- Blocs $U_i = n^{-1/2}(X_{q(i-1)+1} + \cdots + X_{qi})$, $\Delta_i = \sigma n^{-1/2}(\varepsilon_{q(i-1)+1} + \cdots + \varepsilon_{qi})$ où ε_k sont des gaussienne centrées réduites iid, $V_k = \sum_{i=1}^k U_i$ et $\Gamma_k = \sum_{i=1}^k \Delta_i$.
- $V_p \approx S_n$ et $\Gamma_p \approx \sigma N$ avec $p \approx n/q$.
- Avec $\Gamma_{p+1} = V_0 = 0$:

$$\begin{aligned} \mathrm{E}(h(S_n) - h(\sigma N)) &= \sum_{k=1}^{p} \mathrm{E}(h(V_k + \Gamma_{k+1}) - h(V_{k-1} + \Gamma_{k+1})) \\ &\quad + \sum_{k=1}^{p} \mathrm{E}(h(V_{k-1} + \Gamma_{k+1}) - h(V_{k-1} + \Gamma_k)). \end{aligned}$$

- Développement de Taylor à l'ordre 2 et 3 sur h pour faire apparaître les restes $p\,\mathrm{E}(U_k^2(1 \vee |U_k|))$ et $p\,\mathrm{E}(\Delta_k^2(1 \vee |\Delta_k|))$.
- Les termes croisés $p\,\mathrm{E}U_k h'$ et $p\,\mathrm{E}((U_k^2 - \Delta_k^2)h'')/2$ sont contrôlés par $\sum_{n=0}^\infty X_0 \mathrm{E}(X_n|\mathcal{F}_0)$.

Le mélange fort

Definition (Rosenblatt, 1956)

$$\alpha(r) = \sup_{\|f\|_\infty, \|g\|_\infty \leq 1} |Cov(f(X_{s_1}, \ldots, X_{s_u}), g(X_{t_1}, \ldots, X_{t_v}))| \text{ où } t_1 - s_u \geq r.$$

Doukhan *et al.*, 1994, donne une condition nécessaire et suffisante pour (TLC) :

$$\int_0^1 \alpha^{-1}(t)[Q(t)]^2 dt < \infty, \qquad \text{(DMR)}$$

où $\alpha^{-1}(t)$ est l'inverse de $\alpha(t) = \alpha([t])$ et $Q(t)$ est la fonction quantile.

Remarque : (DMR) est nécessaire et suffisante pour $\sum_{i \geq 0} |\mathrm{Cov}(X_0, X_i)| < \infty$, Rio (1993).

Idée de preuve : On utilise critère d'approximation par martingale (Heyde) ou critère de Dedecker.

Cas de suite stationnaire associée

Definition (Newman, 1980)

$$Cov(f(X_1,\dots,X_u), g(X_{u+r},\dots,X_v)) \geq 0 \; f,g : \boldsymbol{R}^n \to \boldsymbol{R} \textit{ fonctions } \uparrow .$$

Theorem (Newman, 1980)

(TLC) *est vérifié dès que* $\sum_{i\geq 1} Cov(X_0, X_i) < \infty$.

Extension possible aux suites quasi-associées, Bulinski et Shashkin (2007), i.e. telles que $BL(r) \to 0$ quand $r \to \infty$ avec

$$|\mathrm{Cov}(f(X_{s_1},\dots,X_{s_u}), g(X_{t_1},\dots,X_{t_v})| \leq (u \wedge v)\, \mathrm{Lip}\, f \,\mathrm{Lip}\, g\; BL(r),$$

pour tout $s_1 \leq \cdots \leq s_u \leq s_u + r \leq t_1 \leq \cdots \leq t_v$ et toutes fonctions Lipschitziennes f et g.

Utilisation des fonctions caractéristiques, Newman (1980)

Soit $\varphi_t(X) = \mathrm{E}(\exp(itX))$, fonctions complexes, bornées et Lipschitziennes et les blocs de Bernstein $U_{i,n} = n^{-1/2}(X_{q(i-1)+1} + \cdots + X_{qi-r})$ et $\widetilde{U}_{i,n} = n^{-1/2}(X_{qi-r} + \cdots + X_{qi})$ avec $r \ll q \ll n$.

- $\varphi_t(n^{-1/2}S_n) \approx \varphi_t(n^{-1/2}\sum_{i=1}^p U_{i,n})$ car $\mathrm{Var}\,(\widetilde{U}_{i,n}) \leq r(\mathrm{E}X_0^2 + BL(1))$.
- $\varphi_t(n^{-1/2}\sum_{i=1}^p U_{i,n}) \approx \varphi_t(p^{-1/2}U_{0,n}|U_{0,n}|^{-1/2})^p$ car la différence s'écrit comme une somme de covariances et $BL(r) \to 0$ quand $r \to \infty$.
- $\varphi_t(p^{-1/2}U_{0,n}|U_{0,n}|^{-1/2})^p \approx \varphi_t(\sigma N)$ par le TLC pour les tableaux triangulaires iid car $\mathrm{Var}\,(U_{0,n})/U_{0,n} \to \sigma$.

Plan

1. Exemples
2. Approximation par différence de martingale
3. Autres méthodes
4. TLC pour des suites faiblement dépendantes

La dépendance faible

Definition (Doukhan et Louhichi, 1997)

$(X_n)_{n\in\mathbb{N}}$ est dit ϵ-faiblement dépendant s'il existe une suite $\epsilon(r)\downarrow 0$ ($r\uparrow\infty$) telle que :

$$|Cov(f(X_{s_1},\dots,X_{s_u}),g(X_{t_1},\dots,X_{t_v}))|\le\Psi(u,v,f,g)\epsilon(r),$$

pour $s_1\le\cdots\le s_u\le s_u+r\le t_1\le\cdots\le t_v$ et toutes fonctions bornées Lipschitziennes f et g.

- $\Psi(u,v,f,g)=v\|f\|_\infty\mathrm{Lip}\,g$ alors $\epsilon(r)=\theta(r)$, Dedecker et Doukhan, 2003,
- $\Psi(u,v,f,g)=u\|g\|_\infty\mathrm{Lip}\,f+v\|f\|_\infty\mathrm{Lip}\,g+uv\mathrm{Lip}\,f\mathrm{Lip}\,g$ alors $\epsilon(r)=\lambda(r)$, Doukhan et W., 2007.

(TLC) sous λ-faible dépendance

Theorem (Doukhan et W., 2007)

Si $E|X|^m < \infty$ *avec* $m > 2$, (TLC) *est vérifiée dès que* $\lambda_r \leq C(r^{-\lambda})$ *avec* $\lambda > 4 + 2/(m-2)$.

Remarque : Dans le cas θ-faible dépendance, la condition s'écrit $\theta > 2(1 + 1/(m-1))$ alors que $\theta > 1 + 1/(m-1)$ est optimale (Doukhan & Dedecker, 2003).

Idée de la preuve : Par récurrence, on contrôle $\mathrm{E}|S_n|^{m'} \leq Cn^{m'/2}$ pour un certain $2 \leq m' < m$ puis on applique la méthode de Lindeberg sur les fonctions caractéristiques des blocs de Bernstein.

Méthode de Lindeberg sur les fonctions caractéristiques, Bardet *et al.*, 2008

Un tableau triangulaire $U_{i,n}$ vérifie $\sum_{i=1}^{p} U_{i,n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2)$ si

- $\sum_{i=1}^{p} \mathrm{Cov}(U_{0,n}, U_{i,n})$ converge,
- $\sum_{i=1}^{p} |\mathrm{Cov}(\varphi_t(U_{1,n} + \ldots + U_{i,n}), \varphi_t(U_{i+1,n})| \to 0$,
- $\sum_{i=1}^{p} \mathrm{E}|U_{i,n}|^{m'} \to 0$ pour un certain $m' > 2$.

On applique la méthode de Lindeberg à $\varphi_t(n^{-1/2}\sum_{i=1}^{p} U_{i,n}) - \varphi_t(\sigma N)$ qui fait apparaître des termes croisés en somme de covariances et un reste contrôler par $\sum_{i=1}^{p} \mathrm{E}|U_{i,n}|^{m'}$.

Conclusions

Inconvénients

- Conditions sous-optimales et pas de moments généralisés possible.

Avantages

- Outils permettant de généraliser les résultats asymptotiques du cas iid comme le (TLC),
- Généralise la quasi-association et le mélange fort,
- Traite le cas des systèmes dynamiques $X_i = T(X_{i-1})$ avec T déterministe (θ-faible dépendance),
- Conditions vérifiables sur de nombreux modèles.

Schémas de Bernoulli à innovations dépendantes

Soit $X_t = H(\cdots, \xi_{j-1}, \xi_j, \xi_{j+1}, \cdots)$ avec ξ_j tels que $\lambda_\xi(r) \le Cr^{-\lambda_\xi}$ et

$$|H(\cdots, x_{j-1}, x_j, x_{j+1}, \cdots) - H(\cdots, x_{j-1}, y_j, x_{j+1}, \cdots)| \le b_j(\sup_{t \neq j} |x_t| \vee 1)|x_j - y_j|,$$

avec $b_j \le C(1 + |j|)^{-b}$.

Theorem (Doukhan et W., 2007)

(TLC) *est vérifié sur* X_i *si* $E|\xi_0|^{m_\xi} < \infty$ *avec* $m_\xi > 2(l+1)$ *et*

$$\lambda_\xi > \frac{b+1}{b-1}\left(4 + \frac{2}{m_\xi - 2}\right) \qquad \text{si } l = 0, b > 1,$$

$$\lambda_\xi > \frac{b(m_\xi - 1 + l)}{(b-2)(m_\xi - 1 - l)}\left(4 + \frac{2}{m_\xi(l+1) - 2}\right) \qquad \text{si } l > 0, b > 2.$$