Central limit theorem for sampled sums of dependent random variables

Clémentine PRIEUR

INSA Toulouse, **I**nstitut **M**athématique de **T**oulouse
Equipe de **S**tatistique et **P**robabilités

LIMIT THEOREMS and APPLICATIONS
JANUARY 14-16, 2008

Framework

Framework

$(E, \mathcal{E}, \mu)$ a probability space,
$T : E \longrightarrow E$ bijective, bimeasurable, μ-preserving.

We define $(\zeta_i)_{i \in \mathbb{Z}} = (T^i)_{i \in \mathbb{Z}}$ from (E, μ) to E.

Framework

$(E, \mathcal{E}, \mu)$ a probability space,
$T : E \longrightarrow E$ bijective, bimeasurable, μ-preserving.

We define $(\zeta_i)_{i\in\mathbb{Z}} = (T^i)_{i\in\mathbb{Z}}$ from (E, μ) to E.

$(\Omega, \mathcal{A}, \mathbb{P})$ a probability space,
$(X_i)_{i\geq 1} \in \mathbb{Z}^{\mathbb{N}^*}$ i.i.d. on $(\Omega, \mathcal{A}, \mathbb{P})$, .

We define $S_n = \sum_{i=1}^n X_i,\ n \geq 1, \quad S_0 \equiv 0.$

Framework

$(E, \mathcal{E}, \mu)$ a probability space,
$T : E \to E$ bijective, bimeasurable, μ-preserving.

We define $(\zeta_i)_{i\in\mathbb{Z}} = (T^i)_{i\in\mathbb{Z}}$ from (E, μ) to E.

$(\Omega, \mathcal{A}, \mathbb{P})$ a probability space,
$(X_i)_{i\geq 1} \in \mathbb{Z}^{\mathbb{N}^*}$ i.i.d. on $(\Omega, \mathcal{A}, \mathbb{P})$, .

We define $S_n = \sum_{i=1}^{n} X_i,\ n \geq 1, \quad S_0 \equiv 0.$

For $f \in \mathbb{L}^1(\mu)$ and $\omega \in \Omega$, we are interested in the sampled ergodic sums

$$\sum_{k=0}^{n-1} f \circ \zeta_{S_k(\omega)}.$$

$(S_k)_{k\geq 0}$ is universally representative for L^p, $p > 1$
if $\exists\ \Omega_0 \subset \Omega$, $\mathbb{P}(\Omega_0) = 1$ /

for every $\omega \in \Omega_0$,

for every dynamical system $(E, \mathcal{E}, \mu, T)$,

for every $f \in L^p$, $p > 1$,

$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^{S_k(\omega)}$ converges μ-almost surely.

$(S_k)_{k\geq 0}$ is universally representative for L^p, $p>1$
if $\exists\ \Omega_0 \subset \Omega$, $\mathbb{P}(\Omega_0)=1$ /

for every $\omega \in \Omega_0$,

for every dynamical system $(E, \mathcal{E}, \mu, T)$,

for every $f \in L^p$, $p>1$,

$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^{S_k(\omega)}$ converges μ-almost surely.

Proposition (Lacey et al.)

Assume $\mathbb{E}(X_1^2) < \infty$.
Then $(S_k)_{k\geq 0}$ is universally representative for L^p, $p>1$ iff the random walk is transient.

We assume now that the random walk $(S_k)_{k\geq 0}$ is transient.

Then, $\forall\ x \in \mathbb{Z}$, the Green function $G(0,x) = \sum_{k=0}^{+\infty} \mathbb{P}(S_k = x)$ is finite.

We assume now that the random walk $(S_k)_{k\geq 0}$ is transient.

Then, $\forall\ x \in \mathbb{Z}$, the Green function $G(0,x) = \sum_{k=0}^{+\infty} \mathbb{P}(S_k = x)$ is finite.

Examples :

1. $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 \neq 0$,
2. $\mathbb{E}X_1 = 0$ and $\forall\ x \in \mathbb{R}$, $\mathbb{P}(n^{-1/\alpha} S_n \leq x) \xrightarrow[n\to+\infty]{} F_\alpha(x)$, where F_α is the distribution function of a stable law with index $\alpha \in (0,1)$.

Weak dependence framework

Weak dependence framework

On the Euclidean space $\mathbb{R}^m$, define the metric $d_1(x, y) = \sum_{i=1}^m |x_i - y_i|$.

Let $\Lambda = \cup_{m \in \mathbb{N}^*} \Lambda_m$ where $\Lambda_m = \{f \ : \mathbb{R}^m \to \mathbb{R}, \text{ Lipschitz with respect to } d_1\}$.

If $f \in \Lambda_m$, define $\mathrm{Lip}\,(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_1(x, y)}$.

Define $\tilde{\Lambda} = \{f \in \Lambda \,/\, \mathrm{Lip}\,(f) \leq 1\}$.

Let $\xi \in \mathbb{R}^m$ be a square integrable r.v. on $(E, \mathcal{E}, \mu)$. For any sub σ-algebra $\mathcal{M} \subset \mathcal{E}$, define

$$\theta_2(\mathcal{M}, \xi) = \sup \left\{ \|\mathbb{E}(f(\xi)|\mathcal{M}) - \mathbb{E}(f(\xi))\|_2 \, , \; f \in \tilde{\Lambda} \right\} .$$

Let $\xi \in \mathbb{R}^m$ be a square integrable r.v. on $(E, \mathcal{E}, \mu)$. For any sub σ-algebra $\mathcal{M} \subset \mathcal{E}$, define

$$\theta_2(\mathcal{M}, \xi) = \sup\left\{ \|\mathbb{E}(f(\xi)|\mathcal{M}) - \mathbb{E}(f(\xi))\|_2 \,,\ f \in \tilde{\Lambda} \right\}.$$

Définition

$(\xi_i)_{i\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$: sequence of square integrable r.v.,

$\forall\ i \in \mathbb{Z}$, $\mathcal{M}_i = \sigma(\xi_j,\ j \le i)$.

Then, $\forall\ k \in \mathbb{N}^* \cup \{\infty\}$, $\forall\ n \in \mathbb{N}$, define $\theta_{k,2}(n)$ by

$$\max_{1\le l\le k} \frac{1}{l} \sup\{\theta_2(\mathcal{M}_p, (\xi_{j_1}, \ldots, \xi_{j_l})),\ p + n \le j_1 < \ldots < j_l\}.$$

Also define $\theta_2(n) = \theta_{\infty,2}(n) = \sup_{k\in\mathbb{N}^*} \theta_{k,2}(n)$.

We consider the stationary sequence $(\xi_x)_{x\in\mathbb{Z}} = (f \circ T^x)_{x\in\mathbb{Z}}$, where $f \in \mathbb{L}^2(\mu)$, $\mathbb{E}_\mu(f) = 0$ and $T\ : E \to E$ bijective, bimeasurable and μ-preserving.

We consider the stationary sequence $(\xi_x)_{x\in\mathbb{Z}} = (f \circ T^x)_{x\in\mathbb{Z}}$, where $f \in \mathbb{L}^2(\mu)$, $\mathbb{E}_\mu(f) = 0$ and $T\ : E \to E$ bijective, bimeasurable and μ-preserving.

Then $\mathcal{M}_i = \sigma\left(f \circ T^j,\ j \leq i\right)$, $i \in \mathbb{Z}$.

Assumption $\mathcal{DEP}$ **:** Assume that $\theta_2^\xi(\cdot)$ is bounded above by some non-negative function $g(\cdot)$ such that

$x \mapsto x^{3/2} g(x)$ is non-increasing,

$\exists\ 0 < \varepsilon < 1,\ \sum_{i=0}^{\infty} 2^{\frac{3i}{2}} g(2^{i\varepsilon}) < \infty.$

If $\theta_2^\xi(n) = \mathcal{O}\left(n^{-a}\right)$, $\mathcal{DEP}$ is satisfied as soon as $a > 3/2$.

Main result

For $f \in \mathbb{L}^2(\mu)$ such that $\mathbb{E}_\mu(f) = 0$, define

$$\sigma^2(f) = 2 \sum_{x \in \mathbb{Z}} G(0,x) \mathbb{E}_\mu(f \; f \circ T^x) - \mathbb{E}_\mu(f^2).$$

Theorem (Guillotin-Plantard & Prieur, 07)

Assume that $(f \circ T^x)_{x \in \mathbb{Z}}$ satisfies $\mathcal{DEP}$. Assume that $\sigma^2(f)$ is finite and positive.
Then, for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n} f \circ T^{S_k(\omega)} \xrightarrow[n \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(f)\right).$$

Remarks :

1. If $(f \circ T^x)_{x \in \mathbb{Z}}$ is a sequence of martingale differences, the clt holds with $0 < \sigma^2(f) = (2G(0,0) - 1)\,\mathbb{E}_\mu(f^2) < \infty$ (Guillotin-Plantard & Schneider, 03).

2. The stationarity assumption on $(f \circ T^x)_{x \in \mathbb{Z}}$ can be relaxed by a stationarity assumption of order 2.

Sketch of proof

Sketch of proof

$\forall\ n \in \mathbb{N}$, $\forall\ x \in \mathbb{Z}$, define the local time of the random walk

$$N_n(x) = \sum_{i=0}^{n} \mathbf{1}_{S_i = x}.$$

For every $\omega \in \Omega$,

$$\sum_{k=0}^{n} f \circ T^{S_k(\omega)} = \sum_{x \in \mathbb{Z}} N_n(x)(\omega) f \circ T^x.$$

Sketch of proof

$\forall\ n \in \mathbb{N}$, $\forall\ x \in \mathbb{Z}$, define the local time of the random walk

$$N_n(x) = \sum_{i=0}^{n} \mathbf{1}_{S_i = x}.$$

For every $\omega \in \Omega$,

$$\sum_{k=0}^{n} f \circ T^{S_k(\omega)} = \sum_{x \in \mathbb{Z}} N_n(x)(\omega) f \circ T^x.$$

We need a clt for triangular arrays of dependent random variables.

clt for triangular arrays of dependent r.v.

Define $X_{n,i} = a_{n,i}\xi_i, \ n = 0, 1, \ldots, \ i = -k_n, \ldots, k_n,$ where

- $(\xi_i)_{i\in\mathbb{Z}}$ is a sequence of centered, non essentially constant and square integrable real valued r.v.,
- $(k_n)_{n\geq 1}$ is a $\searrow$ sequence of positive integers s.t. $k_n \xrightarrow[n\to+\infty]{} +\infty$,
- $\{a_{n,i}, \ -k_n \leq i \leq k_n\}$ is a triangular array of real numbers s.t. $\forall \ n \in \mathbb{N}, \sum_{i=-k_n}^{k_n} a_{n,i}^2 > 0$.

We are interested in the asymptotic behaviour of

$$\Sigma_n = \sum_{i=-k_n}^{k_n} X_{n,i} = \sum_{i=-k_n}^{k_n} a_{n,i}\xi_i.$$

Let $\mathcal{M}_i = \sigma(\xi_j,\ j \le i)$ and $\sigma_n^2 = \mathrm{Var}(\Sigma_n)$.

Theorem (Guillotin-Plantard & Prieur, 07)

Assume that

(A_1)

(i) $\liminf_{n\to+\infty} \frac{\sigma_n^2}{\sum_{i=-k_n}^{k_n} a_{n,i}^2} > 0$,

(ii) $\lim_{n\to+\infty} \sigma_n^{-1} \max_{-k_n \le i \le k_n} |a_{n,i}| = 0$.

(A_2) $\left\{\xi_i^2\right\}_{i\in\mathbb{Z}}$ *is an uniformly integrable family.*

Assume moreover that $(\xi_i)_{i\in\mathbb{Z}}$ *satisfies* $\mathcal{DEP}$.

Then,

$$\frac{\Sigma_n}{\sigma_n} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Case of sampled sums

Define $\sigma_n^2(f) = \mathbb{E}_\mu\left(\left|\sum_{k=0}^{n} f \circ T^{S_k(\omega)}\right|^2\right)$.

Recall that $\sigma^2(f) = 2\sum_{x\in\mathbb{Z}} G(0,x)\mathbb{E}_\mu(f\ f\circ T^x) - \mathbb{E}_\mu(f^2)$.

Proposition (Guillotin-Plantard & Prieur, 07)

If $\sum_{x\in\mathbb{Z}} G(0,x)\mathbb{E}_\mu(f\ f\circ T^x) < +\infty$, *then*

$$\frac{\sigma_n^2(f)}{n} \xrightarrow[n\to+\infty]{\mathbb{P}\text{-}a.s.} \sigma^2(f).$$

Example :

Unsymmetric random walk on nearest neighbours with $p > q$.

$$\forall\ x \geq 0,\ G(0,x) = (p-q)^{-1},$$
$$\forall\ x \leq -1,\ G(0,x) = (p-q)^{-1}\left(\frac{p}{q}\right)^{x}.$$

If $f = h - h \circ T$, we get for $\sigma^2(f)$:

$$-2\frac{p-1}{p}\mathbb{E}_\mu(h^2) + 2\mathbb{E}_\mu(h\ h \circ T) - 2\frac{p-q}{pq}\sum_{x \geq 1}\left(\frac{p}{q}\right)^{x}\mathbb{E}_\mu(h\ h \circ T^x).$$

Define $M_n = \max_{0 \le k \le n} |S_k|$.

Then

$$\sum_{k=0}^{n} f \circ T^{S_k} = \sum_{|x| \le M_n} N_n(x) f \circ T^x .$$

We apply the clt to the triangular array

$$\left\{ X_{n,i} = \frac{N_n(i)}{\sqrt{n}} \, f \circ T^i, \; n \in \mathbb{N}, \; -M_n \le i \le M_n \right\}.$$

As $f \in \mathbb{L}^2(\mu)$, the stationary family $\left\{(f \circ T^i)^2\right\}_{i\in\mathbb{Z}}$ is uniformly integrable.

As $f \in \mathbb{L}^2(\mu)$, the stationary family $\left\{(f \circ T^i)^2\right\}_{i \in \mathbb{Z}}$ is uniformly integrable.

Define the self-intersection local time $\alpha(n,x) = \sum_{i,j=0}^{n} \mathbf{1}_{S_i - S_j = x}$, $x \in \mathbb{Z}$.

Proof of (A_1) (i) :
$\sum_{i=-M_n}^{M_n} a_{n,i}^2 = \frac{\alpha(n,0)}{n} \xrightarrow[n \to +\infty]{\mathbb{P}\text{-a.s.}} 2G(0,0) - 1 > 0$
(Guillotin-Plantard & Schneider, 03).

Moreover, $\mathrm{Var}\left(\sum_{i=1}^{n} X_{n,i}\right) = \frac{\sigma_n^2(f)}{n} \xrightarrow[n \to +\infty]{} \sigma^2(f) > 0$.

Proof of (A_1)(ii) :

$\forall\ \rho > 0$,

$$\max_{-M_n \le i \le M_n} |a_{n,i}| = \frac{1}{\sqrt{n}} \max_{i \in \mathbb{Z}} N_n(i) = o\left(n^{\rho - \frac{1}{2}}\right)\ \mathbb{P}\text{-a.s.}.$$

Hence $\left(\sqrt{\frac{\sigma_n^2(f)}{n}}\right)^{-1} \max_{-M_n \le i \le M_n} |a_{n,i}| \xrightarrow[n \to +\infty]{\mathbb{P}\text{-a.s.}} 0.$

Proof of (A_1)(ii) :

$\forall\ \rho > 0$,

$$\max_{-M_n \le i \le M_n} |a_{n,i}| = \frac{1}{\sqrt{n}} \max_{i \in \mathbb{Z}} N_n(i) = o\left(n^{\rho - \frac{1}{2}}\right) \ \mathbb{P}\text{-a.s.}.$$

Hence $\left(\sqrt{\frac{\sigma_n^2(f)}{n}}\right)^{-1} \max_{-M_n \le i \le M_n} |a_{n,i}| \xrightarrow[n \to +\infty]{\mathbb{P}\text{-a.s.}} 0.$

$\Rightarrow$ clt for the sampled sums under $\mathcal{DEP}$.

Example of sampled dynamical systems

Example of sampled dynamical systems

Let $I = [0, 1]$, $T\ : I \mapsto I$ s.t. $\exists$ a T-invariant probability μ on I.
Define $X_i = T^i$. Then $(X_i)_{i \geq 0}$ is strictly stationary from (I, μ) to I.

Example of sampled dynamical systems

Let $I = [0,1]$, $T\ : I \mapsto I$ s.t. $\exists$ a T-invariant probability μ on I.
Define $X_i = T^i$. Then $(X_i)_{i \geq 0}$ is strictly stationary from (I, μ) to I.

Perron-Frobenius operator :
Let $K : \mathbb{L}^1(I, \mu) \mapsto \mathbb{L}^1(I, \mu)$ defined by

$\forall\ h \in \mathbb{L}^1(I, \mu)$, $\forall\ k \in \mathbb{L}^\infty(I, \mu)$,

$$\int_0^1 K(h)(x)k(x)d\mu(x) = \int_0^1 h(x)(k \circ T)(x)d\mu(x).$$

Example of sampled dynamical systems

Let $I = [0,1]$, $T : I \mapsto I$ s.t. $\exists$ a T-invariant probability μ on I.
Define $X_i = T^i$. Then $(X_i)_{i\geq 0}$ is strictly stationary from (I, μ) to I.

Perron-Frobenius operator :
Let $K : \mathbb{L}^1(I,\mu) \mapsto \mathbb{L}^1(I,\mu)$ defined by

$\forall\ h \in \mathbb{L}^1(I,\mu)$, $\forall\ k \in \mathbb{L}^\infty(I,\mu)$,

$$\int_0^1 K(h)(x)k(x)d\mu(x) = \int_0^1 h(x)(k \circ T)(x)d\mu(x).$$

Define $(Y_i)_{i\geq 0}$ the stationary Markov chain with invariant distribution μ and transition kernel K. Then,
$\forall\ n \geq 0,\ (X_0, X_1, \ldots, X_n) \sim (Y_n, Y_{n-1}, \ldots, Y_0)$ (Gordin, 68).

We endow the space of BV-functions with the norm $\|h\|_v = \|dh\| + \|h\|_{1,\mu}$.

In the case where the spectral analysis of K on $(BV, \|\cdot\|_v)$ yields the existence of a spectral gap, $\exists\,!$ T-invariant $\mu << \lambda$ on I whose density f_μ is BV.

We endow the space of BV-functions with the norm $\|h\|_v = \|dh\| + \|h\|_{1,\mu}$.

In the case where the spectral analysis of K on $(BV, \|\cdot\|_v)$ yields the existence of a spectral gap, $\exists$! T-invariant $\mu << \lambda$ on I whose density f_μ is BV.

Assumptions : regularity, expansivity, topologically mixing (see Collet *et al*, 02).

Then $\exists\ C > 0$, $\exists\ 0 < \rho < 1$ s.t.
$\forall\ n \geq 0$, $\theta_2^Y(n) \leq C\ \rho^n$, Dedecker & Prieur (05).

For this example, the limit variance is

$$\sigma^2(f) = 2\sum_{x\in\mathbb{Z}} G(0,x)\mathrm{Cov}\left(f(X_0), f(X_{|x|})\right) - \mathrm{Var}\left(f(X_0)\right).$$

Parametric estimation by random sampling

Parametric estimation by random sampling

We observe a stationary process $(\zeta_i)_{i\in\mathbb{N}}$ at random times $S_n,\ n \geq 0$, with $(S_n)_{n\geq 0}$ a non negative, increasing and transient random walk.

Sampled empirical mean : $\hat{m}_n = \frac{1}{n}\sum_{i=1}^{n} \zeta_{S_i}$.

A quadratic criterion : $a(S) = \lim_{n\to+\infty} (n\mathrm{Var}\,\hat{m}_n)$.

If $(\mathrm{Cov}(\zeta_1, \zeta_{n+1}))_{n\in\mathbb{N}} \in l^1$,

$$a(S) = \sum_{k=-\infty}^{+\infty} \mathrm{Cov}\left(\zeta_{S_1}, \zeta_{S_{|k|+1}}\right) < \infty.$$

Corollaire (Guillotin-Plantard & Prieur, 07)

Assume that $(\zeta_i)_{i\in\mathbb{N}}$ *satisfies* $\mathcal{DEP}$. *Assume moreover that the random walk is transient,* $S_0 = 0$ *and* $(S_{n+1} - S_n)_{n\in\mathbb{N}}$ *takes its values in* $\mathbb{N}^*$.
Then, for $\mathbb{P}$*-a.e.* $\omega \in \Omega$,

$$\sqrt{n}\,(\hat{m}_n - m) \xrightarrow[n\to+\infty]{} \mathcal{N}\left(0, a(S)\right).$$

Proof : $f = Id - m$, $\sigma^2(f) = a(S)$.

Perspectives

- And using Lindeberg-Rio + blocks? $\theta_{k,2}(n)$?

- The case where the random walk is recurrent (Pène, 07).

Thank you for your attention

www.lsp.ups-tlse.fr/Fp/Prieur