Strong approximation for additive functionals of geometrically ergodic Markov chains

Florence Merlevède

Joint work with E. Rio

Université Paris-Est-Marne-La-Vallée (UPEM)

Luminy, 15-19 February 2016

Strong approximation in the iid setting (1)

- Assume that $(X_i)_{i\geq 1}$ is a sequence of iid centered real-valued random variables with a finite second moment σ^2 and define $S_n = X_1 + X_2 + \cdots + X_n$
- The ASIP says that a sequence $(Z_i)_{i\geq 1}$ of iid centered Gaussian variables may be constructed is such a way that

$$\sup_{1\leq k\leq n} \left|S_k - \sigma B_k\right| = o(b_n) \text{ almost surely,}$$

  where $b_n = (n \log\log n)^{1/2}$ (Strassen (1964)).
- When $(X_i)_{i\geq 1}$ is assumed to be in addition in $\mathbf{L}^p$ with $p > 2$, then we can obtain rates in the ASIP:

$$b_n = n^{1/p}$$

  (see Major (1976) for $p \in]2, 3]$ and Komlós, Major and Tusnády for $p > 3$).

Strong approximation in the iid setting (2)

- When $(X_i)_{i\geq 1}$ is assumed to have a finite moment generating function in a neighborhood of 0, then the famous Komlós-Major-Tusnády theorem (1975 and 1976) says that one can construct a standard Brownian motion $(B_t)_{t\geq 0}$ in such a way that

$$\mathbb{P}\Big(\sup_{k\leq n} |S_k - \sigma B_k| \geq x + c\log n\Big) \leq a\exp(-bx) \qquad (1)$$

  where a, b and c are positive constants depending only on the law of X_1.

- (1) implies in particular that

$$\sup_{1\leq k\leq n} \big|S_k - \sigma B_k\big| = O(\log n) \text{ almost surely}$$

- It comes from the Erdös-Rényi law of large numbers (1970) that this result is unimprovable.

Strong approximation in the multivariate iid setting

- Einmahl (1989) proved that we can obtain the rate $O((\log n)^2)$ in the almost sure approximation of the partial sums of iid random vectors with finite moment generating function in a neighborhood of 0 by Gaussian partial sums.
- Zaitsev (1998) removed the extra logarithmic factor and obtained the KMT inequality in the case of iid random vectors.
- What about KMT type results in the dependent setting?
- For functions of an iid sequence, see the recent paper Berkes, Liu and Wu (2014): the rate is $o(n^{1/p})$.
- What about strong approximation in the Markov setting?

What about strong approximation in the Markov setting?

- Let (ξ_n) be an irreducible and aperiodic Harris recurrent Markov chain on a countably generated measurable state space $(E, \mathcal{B})$. Let $P(x, .)$ be the transition probability.
- We assume that the chain is positive recurrent. Let π be its (unique) invariant probability measure.
- Then there exists some positive integer m, some measurable function h with values in $[0, 1]$ with $\pi(h) > 0$, and some probability measure ν on E, such that

$$P^m(x, A) \geq h(x)\nu(A) .$$

- We assume that $m = 1$
- The Nummelin splitting technique (1984) allows to extend the Markov chain in such a way that the extended Markov chain has a recurrent atom. This allows regeneration.

The Nummelin splitting technique

- Let $Q(x, \cdot)$ be the sub-stochastic kernel defined by $Q = P - h \otimes \nu$
- The minorization condition allows to define an extended chain $(\bar{\xi}_n, U_n)$ in $E \times [0,1]$ as follows.
- At time 0, U_0 is independent of $\bar{\xi}_0$ and has the uniform distribution over $[0,1]$; for any $n \in \mathbb{N}$,

$$\mathbb{P}(\bar{\xi}_{n+1} \in A \mid \bar{\xi}_n = x, U_n = y) = \mathbf{1}_{y \leq h(x)} \nu(A) + \mathbf{1}_{y > h(x)} \frac{Q(x, A)}{1 - h(x)}$$
$$:= \bar{P}((x, y), A)$$

  and U_{n+1} is independent of $(\bar{\xi}_{n+1}, \bar{\xi}_n, U_n)$ and has the uniform distribution over $[0,1]$.
- $\tilde{P} = \bar{P} \otimes \lambda$ (λ is the Lebesgue measure on $[0,1]$) and $(\bar{\xi}_n, U_n)$ is an irreducible and aperiodic Harris recurrent chain, with unique invariant probability measure $\pi \otimes \lambda$. Moreover $(\bar{\xi}_n)$ is an homogenous Markov chain with transition probability $P(x, .)$.

Regeneration

- Define now the set C in $E \times [0,1]$ by

$$C = \{(x,y) \in E \times [0,1] \ \text{ such that } y \le h(x)\}.$$

For any (x,y) in C, $\mathbb{P}(\bar{\bar{\xi}}_{n+1} \in A \mid \bar{\bar{\xi}}_n = x, U_n = y) = \nu(A)$. Since $\pi \otimes \lambda(C) = \pi(h) > 0$, the set C is an atom of the extended chain, and it can be proven that this atom is recurrent.

- Let

$$T_0 = \inf\{n \ge 1 : U_n \le h(\bar{\bar{\xi}}_n)\} \ \text{ and } \ T_k = \inf\{n > T_{k-1} : U_n \le h(\bar{\bar{\xi}}_n)\},$$

and the return times $(\tau_k)_{k>0}$ by $\tau_k = T_k - T_{k-1}$. Note that T_0 is a.s. finite and the return times τ_k are iid and integrable.

- Let $S_n(f) = \sum_{k=1}^n f(\bar{\bar{\xi}}_k)$.
- The random vectors $(\tau_k, S_{T_k}(f) - S_{T_{k-1}}(f))_{k>0}$ are iid and their common law is the law of $(T_0, S_{T_0}(f))$ under $\mathbb{P}_C$.

- Csáki and Csörgö (1995): If the r.v's $S_{T_k}(|f|) - S_{T_{k-1}}(|f|)$ have a finite moment of order p for some p in $]2,4]$ and if $\mathbf{E}(\tau_k^{p/2}) < \infty$, then one can construct a standard Wiener process $(W_t)_{t\geq 0}$ such that
$$S_n(f) - n\pi(f) - \sigma(f)\, W_n = O(a_n) \text{ a.s. } .$$
with $a_n = n^{1/p}(\log n)^{1/2}(\log\log n)^{\alpha}$ and $\sigma^2(f) = \lim_n \frac{1}{n}\mathrm{Var} S_n(f)$.
- The above result holds for any bounded function f only if the return times have a finite moment of order p.
- The proof is based on the regeneration properties of the chain and on an application of the results of KMT (1975) to the partial sums of the iid random variables $S_{T_{k+1}}(f) - S_{T_k}(f)$, $k > 0$.

On the proof of Csáki and Csörgö

- For any $i \geq 1$, let $X_i = \sum_{\ell=T_{i-1}+1}^{T_i} f(\bar{\xi}_\ell)$.
- Since the $(X_i)_{i>0}$ are iid, if $\mathbb{E}|X_1|^{2+\delta} < \infty$, there exists a standard Brownian motion $(W(t))_{t>0}$ such that

$$\sup_{k \leq n} \Big| \sum_{i=1}^{k} X_i - \sigma(f) W(k) \Big| = o(n^{1/(2+\delta)}) \quad a.s.$$

- Let $\rho(n) = \max\{k : T_k \leq n\}$. If $\mathbf{E}|\tau_1|^q < \infty$ for some $1 \leq q \leq 2$, then

$$\rho(n) = \frac{n}{\mathbf{E}(\tau_1)} + O(n^{1/q} (\log \log n)^\alpha) \quad a.s.$$

- $\sum_{i=1}^{\rho(n)} X_i - W\big(\rho(n)\big) = o(n^{\frac{1}{2+\delta}})$, $\sum_{i=1}^{\rho(n)} X_i - S_n(f) = o(n^{\frac{1}{2+\delta}})$ a.s.
- $W\big(\rho(n)\big) - W\big(\frac{n}{\mathbf{E}(\tau_1)}\big) = O\big(n^{1/(2q)} (\log n)^{1/2} (\log \log n)^\alpha\big)$ a.s.
- With this method, no way to do better than $O(n^{1/(2q)} (\log n)^{1/2})$ $(1 \leq q \leq 2)$ even if f is bounded and τ_1 has exponential moment.

Link between the moments of return times and the coefficients of absolute regularity (1)

- For positive measures μ and ν, let $\|\mu - \nu\|$ denote the total variation of $\mu - \nu$
- Set
$$\beta_n = \int_E \|P^n(x,.) - \pi\| d\pi(x) \,.$$
The coefficients β_n are called absolute regularity (or β-mixing) coefficients of the chain.
- Bolthausen (1980-1982): for any $p > 1$,
$$\mathbf{E}(\tau_1^p) = \mathbf{E}_C(T_0^p) < \infty \text{ if and only if } \sum_{n>0} n^{p-2}\beta_n < \infty \,.$$

Link between the moments of return times and the coefficients of absolute regularity (2)

- Requiring $\beta_n = O(\rho^n)$ for some real ρ with $0 < \rho < 1$ is equivalent to say that the Markov chain is geometrically ergodic (see Nummelin and Tuominen (1982)).
- If the Markov chain is GE then there exists a positive real δ such that
$$\mathbf{E}\big(e^{t\tau_1}\big) < \infty \text{ and } \mathbf{E}_\pi\big(e^{tT_0}\big) < \infty \text{ for any } |t| \le \delta\,.$$
- Heuristic: due to the decomposition in iid cycles, if f is bounded and the chain is geometrically ergodic, we can expect the same result as in the iid case when we have r.v.'s with exponential moments.

Main result: M. Rio (2015)

- Assume that

$$\beta_n = O(\rho^n) \text{ for some real } \rho \text{ with } 0 < \rho < 1,$$

- If f is bounded and such that $\pi(f) = 0$ then there exists a standard Wiener process $(W_t)_{t\geq 0}$ and positive constants a, b and c depending on f and on the transition probability $P(x, \cdot)$ such that, for any positive real x and any integer $n \geq 2$,

$$\mathbb{P}_\pi\Big(\sup_{k\leq n} \big|S_k(f) - \sigma(f) W_k\big| \geq c \log n + x \Big) \leq a \exp(-bx)\,.$$

where $\sigma^2(f) = \pi(f^2) + 2\sum_{n>0} \pi(fP^n f) > 0$.

- Therefore $\sup_{k\leq n} \big|S_k(g) - \sigma(f) W_k\big| = O(\log n)$ a.s.

A remark

- Let μ be any law on E such that

$$\int_E \|P^n(x,.) - \pi\| d\mu(x) = O(r^n) \text{ for some } r < 1.$$

  Then $\mathbb{P}_\mu(T_0 > n)$ decreases exponentially fast (see Nummelin and Tuominen (1982)).

- The result extends to the Markov chain (ξ_n) with transition probability P and initial law μ.

Some insights for the proof

- Let $S_n(f) = \sum_{\ell=1}^{n} f(\bar{\xi}_\ell)$ and $X_i = \sum_{\ell=T_{i-1}+1}^{T_i} f(\bar{\xi}_\ell)$. Recall that $(X_i, \tau_i)_{i>0}$ are iid.
- Let α be the unique real such that $\mathrm{Cov}(X_k - \alpha\tau_k, \tau_k) = 0$
- The random vectors $(X_i - \alpha\tau_i, \tau_i)_{i>0}$ of $\mathbb{R}^2$ are then iid and their marginals are non correlated.
- By the multidimensional strong approximation theorem of Zaitsev (1998), there exist two independent standard Brownian motions $(B_t)_t$ and $(\widetilde{B}_t)_t$ such that

$$S_{T_n}(f) - \alpha(T_n - n\mathbf{E}(\tau_1)) - vB_n = O(\log n) \text{ a.s.} \quad (1)$$

  and

$$T_n - n\mathbf{E}(\tau_1) - \tilde{v}\widetilde{B}_n = O(\log n) \text{ a.s.} \quad (2)$$

  where $v^2 = \mathrm{Var}(X_1 - \alpha\tau_1)$ and $\tilde{v}^2 = \mathrm{Var}(\tau_1)$.
- We associate to T_n a Poisson Process via (2).

- Let $\lambda = \frac{(\mathbb{E}(\tau_1))^2}{\mathrm{Var}(\tau_1)}$. Via KMT, one can construct a Poisson process N (depending on $\widetilde{B}$) with parameter λ in such a way that

$$\gamma N(n) - n\mathbb{E}(\tau_1) - \tilde{v}\widetilde{B}_n = O(\log n) \text{ a.s.}$$

- Therefore, via (2),

$$T_n - \gamma N(n) = O(\log n) \text{ a.s.}$$

and then, via (1),

$$S_{\gamma N(n)}(f) - \alpha\gamma N(n) + \alpha n\mathbb{E}(\tau_1) - vB_n = O(\log n) \text{ a.s.} \quad (3)$$

- The processes $(B_t)_t$ and $(N_t)_t$ appearing here are independent.
- Via (3), setting $N^{-1}(k) = \inf\{t > 0 : N(t) \geq k\} := \sum_{\ell=1}^{k} \mathcal{E}_\ell$,

$$S_n(f) = vB_{N^{-1}(n/\gamma)} + \alpha n - \alpha\mathbb{E}(\tau_1)N^{-1}(n/\gamma) + O(\log n) \text{ a.s.}$$

- If $v = 0$, the proof is finished. Indeed, by KMT, there exists a Brownian motion $\widetilde{W}_n$ (depending on N) such that

$$\alpha n - \alpha\mathbb{E}(\tau_1)N^{-1}(n/\gamma) = \widetilde{W}_n + O(\log n) \text{ a.s.}$$

- If $v \neq 0$ and $\alpha = 0$, we have

$$S_n(f) = vB_{N^{-1}(n/\gamma)} + O(\log n) \text{ a.s.}$$

- Using Csörgö, Deheuvels and Horváth (1987) (B and N are independent), one can construct a Brownian motion W (depending on N) such that

$$B_{N^{-1}(n/\gamma)} - W_n = O(\log n) \text{ a.s. } (*)\,,$$

  which leads to the expected result when $\alpha = 0$.

- However, in the case $\alpha \neq 0$ and $v \neq 0$, we still have

$$S_n(f) = vB_{N^{-1}(n/\gamma)} + \alpha n - \alpha \mathbb{E}(\tau_1) N^{-1}(n/\gamma) + O(\log n) \text{ a.s.}$$

  and then

$$S_n(f) = vW_n + \widetilde{W}_n + O(\log n) \text{ a.s.}$$

- Since W and $\widetilde{W}$ are not independent, we cannot conclude. Can we construct W_n independent of N (and then of $\widetilde{W}$) such that $(*)$ still holds?

The key lemma

- Let $(B_t)_{t\geq 0}$ be a standard Brownian motion on the line and $\{N(t) : t \geq 0\}$ be a Poisson process with parameter $\lambda > 0$, independent of $(B_t)_{t\geq 0}$.
- Then one can construct a standard Brownian process $(W_t)_{t\geq 0}$ independent of $N(\cdot)$ and such that, for any integer $n \geq 2$ and any positive real x,

$$\mathbb{P}\Big(\sup_{k\leq n}\big|B_k - \frac{1}{\sqrt{\lambda}}W_{N(k)}\big| \geq C\log n + x\Big) \leq A\exp(-Bx)\,,$$

  where A, B and C are positive constants depending only on λ.
- $(W_t)_{t\geq 0}$ may be constructed from the processes $(B_t)_{t\geq 0}$, $N(\cdot)$ and some auxiliary atomless random variable δ independent of the σ-field generated by the processes $(B_t)_{t\geq 0}$ and $N(\cdot)$.

Construction of W (1/3)

- It will be constructed from B by writing B on the Haar basis.
- For $j \in \mathbb{Z}$ and $k \in \mathbb{N}$, let

$$e_{j,k} = 2^{-j/2}\big(\mathbf{1}_{]k2^j,(k+\frac{1}{2})2^j]} - \mathbf{1}_{](k+\frac{1}{2})2^j,(k+1)2^j]}\big)\,,$$

and

$$Y_{j,k} = \int_0^\infty e_{j,k}(t)dB(t) = 2^{-j/2}\big(2B_{(k+\frac{1}{2})2^j} - B_{k2^j} - B_{(k+1)2^j}\big)\,.$$

Then, since $(e_{j,k})_{j\in\mathbb{Z},k\geq 0}$ is a total orthonormal system of $\ell^2(\mathbb{R})$, for any $t \in \mathbb{R}^+$,

$$B_t = \sum_{j\in\mathbb{Z}}\sum_{k\geq 0}\Big(\int_0^t e_{j,k}(t)dt\Big)Y_{j,k}\,.$$

- To construct W, we modify the $e_{j,k}$.

Construction of W (2/3)

- Let $E_j = \{k \in \mathbb{N} \,:\, N(k2^j) < N((k+\frac{1}{2})2^j) < N((k+1)2^j)\}$
- For $j \in \mathbb{Z}$ and $k \in E_j$, let

$$f_{j,k} = c_{j,k}^{-1/2}\big(b_{j,k}\mathbf{1}_{]N(k2^j),N((k+\frac{1}{2})2^j)]} - a_{j,k}\mathbf{1}_{]N((k+\frac{1}{2})2^j),N((k+1)2^j)]}\big)\,,$$

where

$$a_{j,k} = N((k+\frac{1}{2})2^j) - N(k2^j)\,,\ b_{j,k} = N((k+1)2^j) - N((k+\frac{1}{2})2^j)\,,$$

and $c_{j,k} = a_{j,k}b_{j,k}(a_{j,k}+b_{j,k})$

- $(f_{j,k})_{j\in\mathbb{Z},k\in E_j}$ is an orthonormal system whose closure contains the vectors $\mathbf{1}_{]0,N(t)]}$ for $t \in \mathbb{R}^+$ and then the vectors $\mathbf{1}_{]0,\ell]}$ for $\ell \in \mathbb{N}^*$.
- Setting $f_{j,k} = 0$ if $k \notin E_j$, we define

$$W_\ell = \sum_{j\in\mathbb{Z}}\sum_{k\geq 0}\Big(\int_0^\ell f_{j,k}(t)dt\Big)Y_{j,k} \text{ for any } \ell \in \mathbb{N}^* \text{ and } W_0 = 0$$

Construction of W (3/3)

$$W_\ell = \sum_{j\in\mathbb{Z}} \sum_{k\geq 0} \Big(\int_0^\ell f_{j,k}(t)dt \Big) Y_{j,k} \text{ for any } \ell \in \mathbb{N}^* \text{ and } W_0 = 0$$

- Conditionally to N, $(f_{j,k})_{j\in\mathbb{Z},k\in E_j}$ is an orthonormal system and $(Y_{j,k})$ is a sequence of iid $\mathcal{N}(0,1)$, independent of N.
- Hence, conditionally to N, $(W_\ell)_{\ell\geq 0}$ is a Gaussian sequence such that $\mathrm{Cov}(W_\ell, W_m) = \ell \wedge m$.
- Therefore this Gaussian sequence is independent of N
- By the Skorohod embedding theorem, we can extend it to a standard Wiener process $(W_t)_t$ still independent of N.

Thank you for your attention!