Une approche markovienne du théorème central limite

Claire Delplancke

Colloque Jeunes Probabilistes et Statisticiens
18 Avril 2016

Central limit theorem

$X_1, \ldots, X_n$ iid real rv with $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$.
γ the standard Gaussian measure.

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \qquad \underset{n\to\infty}{\rightharpoonup} \quad \gamma.$$

Central limit theorem

$X_1, \dots, X_n$ iid real rv with $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$.
γ the standard Gaussian measure.

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \qquad \underset{n\to\infty}{\rightharpoonup} \qquad \gamma.$$

$(Y_n)_{n\geq 1}$ is an **inhomogenous Markov chain**.

$$\begin{aligned} Y_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} X_i + \frac{1}{\sqrt{n}} X_n \\ &= \sqrt{\frac{n-1}{n}} Y_{n-1} + \frac{1}{\sqrt{n}} X_n. \end{aligned}$$

Motivations

Motivations:

- New proof of the CLT ;
- Quantification of the convergence in the CLT for a non-investigated quantity ;
- Application of the ideas to other stochastic algorithms.

Conjecture

Hypothesis and notations

Let (X_n) be iid real rv and $X_1 \sim \varphi\, d\gamma$ with $\boxed{\varphi \in L^2(\gamma)}$.

Then $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \sim f_n\, d\gamma$ with $f_n \in L^2(\gamma)$. Heuristic: '$f_n \to 1$.'

Conjecture

Hypothesis and notations

Let (X_n) be iid real rv and $X_1 \sim \varphi\, d\gamma$ with $\boxed{\varphi \in L^2(\gamma)}$.

Then $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \sim f_n\, d\gamma$ with $f_n \in L^2(\gamma)$. Heuristic: '$f_n \to 1$.'

Variance of f_n with respect to γ :

$$\mathrm{Var}_\gamma(f_n) = \int (f_n - 1)^2 d\gamma = \|f_n - 1\|^2_{L^2(\gamma)}.$$

Conjecture

Hypothesis and notations

Let (X_n) be iid real rv and $X_1 \sim \varphi \, d\gamma$ with $\boxed{\varphi \in L^2(\gamma)}$.

Then $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \sim f_n \, d\gamma$ with $f_n \in L^2(\gamma)$. Heuristic: '$f_n \to 1$.'

Variance of f_n with respect to γ :

$$\mathrm{Var}_\gamma(f_n) = \int (f_n - 1)^2 d\gamma = \|f_n - 1\|^2_{L^2(\gamma)}.$$

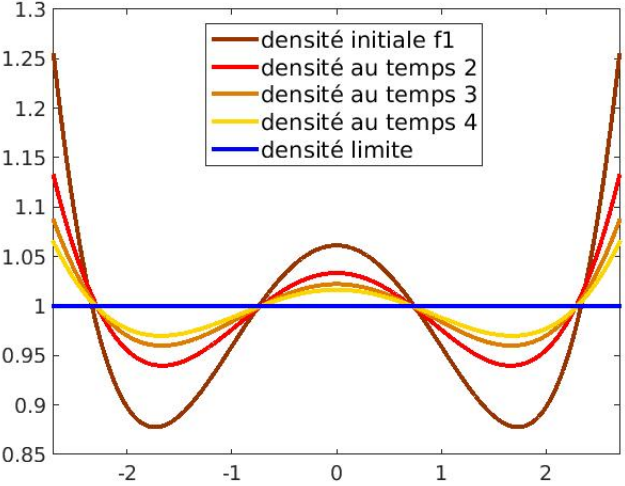
Conjecture

Let $(X_n)_{n\geq 1}$ be iid real rv such that $X_1 \sim \varphi \, d\gamma$ and $\varphi \in L^2(\gamma)$.
If $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$ then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}.$$

Toy model

Simple example: φ polynomial function of degree 4 such that φ is a density and $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$, $\mathbb{E}[X_1^3] = 0$.



Berry-Esseen-like theorems

$$d_{\mathcal{F}}(\mu,\gamma) = \sup_{f\in\mathcal{F}} \left| \int f d\mu - \int f d\gamma \right|.$$

Name	Definition	Rate of convergence
Kolmogorov	$\mathcal{F} = \mathbf{1}_{]-\infty,x]}$	$n^{-1/2}$ (Berry Esseen '48)
Total variation	$\mathcal{F} = \{\|f\|_\infty \leq 1\}$	$n^{-1/2}$ (Sirazhdinov Mamatov '62, Bally Caramellino '14)
Wasserstein	$\mathcal{F} = \mathrm{Lip}_1$	$n^{-1/2}$ (Ibragimov '66, Rio '11)
Relative entropy	$\mathrm{Ent}_\gamma(\nu) = \int \log \frac{d\nu}{d\gamma} d\gamma$	n^{-1} (Arstein et al. '04)
$\mathrm{Var}_\gamma^{1/2}$	$\mathcal{F} =$ unit ball of $L^2(\gamma)$	$n^{-1/2}$

Comparison with the state of the art I

$$d_{Kol}(\nu,\gamma) \leq d_{TV}(\nu,\gamma) \leq C\mathrm{Ent}_\gamma(\nu)^{1/2} \leq C\mathrm{Var}_\gamma^{1/2}(\frac{d\nu}{d\gamma}).$$

Assumptions for the convergence:

- Kolmogorov distance: $\mathbb{E}[|X_1|^3] < \infty$
- Total variation distance: $\mathcal{L}(X_1)$ a.c. w.r.t. γ + $\mathbb{E}[|X_1|^3] < \infty$
- Relative entropy : $\mathrm{Ent}_\gamma(\mathcal{L}(X_1)) < \infty$ + $\mathbb{E}[|X_1|^4] < \infty$
- Var_γ : $\varphi \in L^2(\gamma) \Rightarrow \forall k \in \mathbb{N}$, $\mathbb{E}[|X_1|^k] < \infty$.

Comparison with the state of the art II

We know

$$d_{TV}(m_n, \gamma) \leq \mathrm{Var}_\gamma(f_n)^{1/2}$$

and (Bally Caramellino 2014)

$$cn^{-1/2} \leq d_{TV}(m_n, \gamma) \leq Cn^{-1/2}.$$

Hence if

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}$$

the rate n^{-1} is **optimal**.

Result I

Let (X_n) be iid real rv and $X_1 \sim \varphi\, d\gamma$ with $\varphi \in L^2(\gamma)$.
Then $Y_n = (n^{-1/2}) \sum_{k=1}^{n} X_k \sim f_n\, d\gamma$ with $f_n \in L^2(\gamma)$.

Theorem

Let $(X_n)_{n\geq 1}$ be iid real rv such that $X_1 \sim \varphi\, d\gamma$ and $\varphi \in L^2(\gamma)$.
Suppose φ is polynomial + technical assumptions on its coefficients.
If $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$, then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}.$$

Result I

Let (X_n) be iid real rv and $X_1 \sim \varphi\, d\gamma$ with $\varphi \in L^2(\gamma)$.
Then $Y_n = (n^{-1/2}) \sum_{k=1}^n X_k \sim f_n\, d\gamma$ with $f_n \in L^2(\gamma)$.

Theorem

Let $(X_n)_{n\geq 1}$ be iid real rv such that $X_1 \sim \varphi\, d\gamma$ and $\varphi \in L^2(\gamma)$.
Suppose φ is polynomial + technical assumptions on its coefficients.
If $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$, then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}.$$

Technical assumptions: on the basis of Hermite polynomials,

- the leading coefficients are not too big
- the coefficients do not decay too fast.

Result II

Generalized theorem

Let $(X_n)_{n\geq 1}$ be iid real rv such that $X_1 \sim \varphi\, d\gamma$ and $\varphi \in L^2(\gamma)$.
Suppose φ is polynomial + technical assumptions on its coefficients.
For $r \geq 2$ natural number, if $\mathbb{E}[X_1^k] = \int x^k d\gamma$ for $k = 0, \dots, r$, then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n^{r-1}}.$$

Rough idea of the proof

Hilbert space $L^2(\gamma)$ with scalar product

$$(f,g) \in L^2(\gamma) \to \int fgd\gamma.$$

Variance with respect to γ $\leftrightsquigarrow$ **Hilbertian norm**

$$\mathrm{Var}_\gamma(f_n) = \int (f_n - 1)^2 d\gamma = \|f_n - 1\|^2_{L^2(\gamma)}.$$

Rough idea of the proof

Hilbert space $L^2(\gamma)$ with scalar product

$$(f, g) \in L^2(\gamma) \to \int fg d\gamma.$$

Variance with respect to $\gamma \leftrightsquigarrow$ **Hilbertian norm**

$$\mathrm{Var}_\gamma(f_n) = \int (f_n - 1)^2 d\gamma = \|f_n - 1\|^2_{L^2(\gamma)}.$$

Markov structure $\rightsquigarrow$ linear **operators on** $L^2(\gamma)$.

The Markov operator and the adjoint operator

The family of operators

$$Q_{p,q}[g](y) := \mathbb{E}\left[g(Y_q) \middle| Y_p = y\right] \quad (q \geq p)$$

is an inhomogenous Markov semi-group :

- $Q_{p,q}[\mathbf{1}] = \mathbf{1}$.
- $Q_{p,r} = Q_{p,q}Q_{q,r} \quad (r \geq q \geq p)$.

If $Q^*_{n-1,n}$ is the adjoint in $L^2(\gamma)$ of $Q_{n-1,n}$, then $\boxed{f_n = Q^*_{n-1,n}[f_{n-1}]}$.

Outline of the approach

Goal show that $\mathrm{Var}_\gamma(f_n) \lesssim \frac{1}{n}$.
We need to express

$$\begin{aligned}\mathrm{Var}_\gamma(f_n) &= \int (f_n - 1)^2 d\gamma \\ &= \int (Q^*_{n-1,n}[f_{n-1}] - 1)^2 d\gamma\end{aligned}$$

in terms of

$$\mathrm{Var}_\gamma(f_{n-1}) = \int (f_{n-1} - 1)^2 d\gamma.$$

How ? spectral analysis on the linear operator $Q^*_{n-1,n}$.

Outline of the approach

Goal show that $\mathrm{Var}_\gamma(f_n) \lesssim \frac{1}{n}$.
We need to express

$$\begin{aligned}\mathrm{Var}_\gamma(f_n) &= \int (f_n - 1)^2 d\gamma \\ &= \int (Q^*_{n-1,n}[f_{n-1}] - 1)^2 d\gamma\end{aligned}$$

in terms of

$$\mathrm{Var}_\gamma(f_{n-1}) = \int (f_{n-1} - 1)^2 d\gamma.$$

How ? spectral analysis on the linear operator $Q^*_{n-1,n}$.

(Write the (infinite) matrix corresponding to the operator $Q^*_{n-1,n}$ on the basis of Hermite polynomials.)

Perspectives I

We proved. . .

Theorem

Let $(X_n)_{n\geq 1}$ be iid real rv.
Suppose X_1 has a density φ wrt γ and $\varphi \in L^2(\gamma)$.
Suppose φ is polynomial + technical assumptions on its coefficients.
If $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$ then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}.$$

Perspectives II

We would like to prove...

Conjecture

Let $(X_n)_{n\geq 1}$ be iid real rv.
Suppose X_1 has a density φ wrt γ and $\varphi \in L^2(\gamma)$.
If $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$ then $\exists c > 0$ such that

$$\mathrm{Var}_\gamma(f_n) \leq \frac{c}{n}.$$

Perspectives III

More broadly . . .

- Extend to **CLTs** for objects that are not sums of variable.
- Extend techniques of proof to **other stochastic algorithms** that are also inhomogenous Markov chains.

Merci !

Upper bound of (2) I

Recall

$$(2) \quad = \quad \int (Q^*_{n-1,n} g)^2 d\gamma$$

Write the (infinite) matrix of $Q^*_{n-1,n}$ in the $(\bar{H}_k)$ basis : matrix $A = A_n^{\varphi}$ depending on

- Time n
- Coefficients φ_k of φ along the $(\bar{H}_k)$.

Upper bound of (2) II

$$\begin{aligned}
\int (Q^*_{n-1,n}g)^2 d\gamma &= \sum_{k\geq 0} \left(Q^*_{n-1,n}g\right)_k^2 \\
&= \sum_{k\geq 0} (Ag)_k^2 \\
&= \langle Ag, Ag\rangle_{L^2(\mathbb{R}^{\mathbb{N}})} \\
&= \langle g, A^T Ag\rangle_{L^2(\mathbb{R}^{\mathbb{N}})} \\
&\leq \sup\left\{|\lambda|, \lambda \text{ eigenvalue of } A^T A\right\} \|g\|_2^2.
\end{aligned}$$

Upper bound of (2) III

φ polynomial $\Rightarrow$ matrix $A^T A$ has a **finite** number of non-zero diagonals.

$+$ some additional hypothesis on the coefficients (φ_k) :

$$(2) \leq (1 - 6t_n + \mathcal{O}(t_n^2)) \int g^2 d\gamma.$$

Technical conditions I

Let g_i be the i-th coefficient of g in the basis of renormalized Hermite polynomials and denote

$$\begin{aligned} V_K &= \left\{ g \in L^2(\gamma) | \forall\, 0 \leq i \leq (K-1),\, g_i = 0 \right\} \\ V_K^N &= \left\{ g \in L^2(\gamma) | \forall\, 0 \leq i \leq (K-1),\, g_i = 0 \text{ and } \forall i > N,\, g_i = 0 \right\}. \end{aligned}$$

Set $C_k = \left(1 + \frac{N}{K}\right)^{k/2}$ and $\gamma_k = \frac{1}{\sqrt{k!}} C_k |\varphi_k|$.
(C1) If $K \leq N-2$, for all $K \leq k \leq N-2$,

$$(k+2)\gamma_{k+2} \leq \gamma_k.$$

Technical conditions II

The condition *(C1)* is equivalent to

$$\left|\frac{\varphi_{k+2}}{\varphi_k}\right| \leq \frac{1}{1+\frac{N}{K}}\left(1-\frac{1}{k+2}\right)^{1/2}$$

if $\varphi_k \neq 0$ and is implied by the simplest condition
(C1') If $K \leq N-2$, for all $K \leq k \leq N$,

$$\left|\frac{\varphi_{k+1}}{\varphi_k}\right| \leq r, \quad \text{with } r = \left(\left(1-\frac{1}{K+2}\right)^{1/2}\frac{1}{1+\frac{N}{K}}\right)^{1/2}.$$

Remark that the condition *(C1)* implies that $\gamma_k = 0 \Rightarrow \gamma_{k+2} = 0$. In the following, for the sake of simplicity, we will assume the stronger fact that $\forall k \in [\![K, N]\!]$, $\varphi_k \neq 0$.

Technical conditions III

The second condition *(C2)* has two parts :
(C2a) If $K \leq N-1$,

$$\left(\frac{2(K+1)\gamma_{K+1}}{N}\right)^{N}\left(\frac{K-1}{2\gamma_N}\right)^{K-1} \leq \frac{1}{(N-K+1)^{N-K+1}}$$

and

$$\left(\frac{2K\gamma_K}{N-1}\right)^{N-1}\left(\frac{K-2}{2\gamma_{N-1}}\right)^{K-2} \leq \frac{1}{(N-K+1)^{N-K+1}}.$$

(C2b) If $K = N$,

$$\gamma_N \leq \frac{1}{2(N-2)^{(N-2)/2}}.$$

Technical conditions IV

Loosely speaking, the condition *(C1)* (which has a content only if $K \leq N - 2$) amounts to ask a geometric decrease for the coefficients (φ_k).

If $K = N$ or $K = N - 1$, the condition *(C2)* requires the leading coefficients φ_N and φ_{N-1} to be not too big.

When $K < N - 1$, it can be interpreted as the requirement that the coefficients (φ_k) do not decay too fast.