Around Salem-Zygmund Central limit Theorem

Guillaume Poly (based on joint works with Jurgen Angst/Louis Gass/Thibault Pautrel)

23 October 2020

The standard CLT

The standard central limit Theorem:

1. $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space,
2. $(a_k)_{k\geq 1}$ independent and identically distributed random variables,
3. $\mathbb{E}(a_1^2) = 1$, $\mathbb{E}(a_1) = 0$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} a_k \xrightarrow[n\to\infty]{\text{distribution}} \mathcal{N}(0,1).$$

A general formulation of the problem

We add two new ingredients:

1. $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ where μ is a probability measure on $\mathbb{R}^d$,
2. $(\Phi_k)_{k\geq 1}$ a sequence of **bounded** functions from $\mathbb{R}^d$ to $\mathbb{R}$.

We are interested in proving (under some assumptions):

$$\mathbb{P}\text{-}\mathbf{a.s.},\ S_n(x) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} a_k \Phi_k(x) \xrightarrow[n\to\infty]{\text{Law under } \mu} \mathcal{N}(0, \sigma^2).$$

that is to say:

$$\mathbb{P}\left(\forall \phi \in \mathcal{C}_b(\mathbb{R}, \mathbb{R}),\ \int_{\mathbb{R}^d} \phi\left(S_n(x)\right) d\mu(x) \to \int_{\mathbb{R}} \phi(x) e^{-\frac{x^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} \right).$$

1. intuitive interpretation: if the coefficients $(a_k)_k$ are chosen generically as the realization of an i.i.d. sequence, we have a CLT under the sole randomness of the evaluation point x according to the measure μ.
2. if the coefficients are frozen, under the sole randomness of x, it is not anymore a sum of independent r.v,
3. Some conditions need to be imposed on the $(\phi_k)_{k\geq 1}$ since if we chose $\phi_1 = \phi_2 = \cdots = \phi_k = 1$ the conclusion fails! Indeed by the law of the iterated logarithm,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{2n\log(\log(n))}} \sum_{k=1}^{n} a_k\phi_k(x)$$
$$= \limsup_{n} \frac{1}{\sqrt{2n\log(\log(n))}} \sum_{k=1}^{n} a_k = 1,$$
$$\liminf_{n\to\infty} \frac{1}{\sqrt{2n\log(\log(n))}} \sum_{k=1}^{n} a_k\phi_k(x) = -1.$$

A general formulation of the problem

Let us remark that the problem is well posed:

$$A=\left(\forall\phi\in\mathcal{C}_b(\mathbb{R},\mathbb{R}),\ \int_{\mathbb{R}^d}\phi\left(S_n(x)\right)d\mu\to\int_{\mathbb{R}}\phi(x)e^{-\frac{x^2}{2\sigma^2}}\frac{dx}{\sqrt{2\pi\sigma^2}}\right)\in\mathcal{F}.$$

Indeed, let us recall that

$$X_n\xrightarrow[n\to\infty]{\text{Law}}X_\infty\Leftrightarrow\forall f\text{ s.t.}\|f\|_\infty+\|f'\|_\infty\leq 1:\ \mathbb{E}\left(f(X_n)\right)\to\mathbb{E}\left(f(X)\right).$$

Besides $K:=\{f\ \text{ s.t. }\ \|f\|_\infty+\|f'\|_\infty\leq 1\}$ is compact for $\|\cdot\|_\infty$, hence K is separable. Take f_n a dense sequence then:

$$A=\bigcap_{p\geq 1}\left(\int_{\mathbb{R}^d}f_p\left(S_n(x)\right)d\mu\to\int_{\mathbb{R}}f_p(x)e^{-\frac{x^2}{2\sigma^2}}\frac{dx}{\sqrt{2\pi\sigma^2}}\right).$$

Given that $\omega\to\int_{\mathbb{R}^d}f_p\left(S_n(x)\right)d\mu$ is a random variable, we obtain $A\in\mathcal{F}$.

Salem-Zygmund framework, Gaussian case

1. We take (a_k, b_k) two i.i.d. sequences of standard Gaussian r.v.,
2. we set $S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} a_k \cos(k\theta) + b_k \sin(k\theta)$,

we want to prove that:

$$\boxed{\mathbb{P}\left(\forall \phi \in \mathcal{C}_b(\mathbb{R},\mathbb{R}),\ \int_0^{2\pi} \phi\left(S_n(\theta)\right) \frac{d\theta}{2\pi} \to \int_{\mathbb{R}} \phi(x) e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}\right) = 1}$$

Proof of Gaussian case

1. $\mathbb{E}_{\mathbb{P}}$ stands for expectation w.r.t. $\mathbb{P}$,
2. $\mathbb{E}_{\theta}$ stands for expectation w.r.t. $\theta \sim \mathcal{U}_{[0,2\pi]}$.

$$\Delta_n := \mathbb{E}_{\mathbb{P}}\left(\left(\mathbb{E}_{\theta}\left(e^{itS_n(\theta)} - e^{-\frac{t^2}{2}}\right)\right)^2\right) = \mathbb{E}_{\mathbb{P}}\left(\mathbb{E}_{\theta}\left(e^{itS_n(\theta)} - e^{-\frac{t^2}{2}}\right) \times \mathbb{E}_{\theta'}\left(e^{itS_n(\theta')} - e^{-\frac{t^2}{2}}\right)\right)$$

where θ and θ' are two independent copies.

Proof of Gaussian case

Then we use **Fubini**:

$$
\begin{aligned}
&= \mathbb{E}_{\theta,\theta'}\left(\mathbb{E}_{\mathbb{P}}\left(e^{it(S_n(\theta)+S_n(\theta'))}\right) - 2e^{-\frac{t^2}{2}}\mathbb{E}_{\mathbb{P}}\left(e^{itS_n(\theta)}\right) + e^{-t^2}\right)\\
&= \mathbb{E}_{\theta,\theta'}\left(e^{-t^2(1+K_n(\theta,\theta'))} - e^{-t^2}\right)\\
&\quad \text{with } K_n(\theta,\theta') = \frac{1}{n}\sum_{k=1}^{n}\cos(k(\theta-\theta')).
\end{aligned}
$$

Indeed, with θ, θ' frozen, under $\mathbb{P}$ we have

1. $S_n(\theta) + S_n(\theta') \sim \mathcal{N}(0, 1 + K_n(\theta, \theta'))$,
2. $S_n(\theta) \sim \mathcal{N}(0, 1)$.

So, what seems to matters here is $K_n(\theta, \theta') \to 0$.

Proof of Gaussian case

$$\begin{aligned}
|\Delta_n| &\leq t^2 \mathbb{E}_{\theta,\theta'}\left(\left|K_n(\theta,\theta')\right|\right) \\
&\leq t^2 \mathbb{E}_{\theta}\left(\left|K_n(\theta)\right|\right) \quad (\theta-\theta' \sim \theta) \\
&\leq t^2 \sqrt{\mathbb{E}_{\theta}\left(K_n(\theta)^2\right)} = \frac{t^2}{\sqrt{2n}} = O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Now we take the subsequence n^3, $\Delta_{n^3} = O(1/n^{\frac{3}{2}})$ so by using Borel-Cantelli (for instance):

$$\mathbb{P} - a.s.,\ \mathbb{E}_{\theta}\left(e^{itS_{m^3}(\theta)} - e^{-\frac{t^2}{2}}\right) \to 0.$$

Proof of Gaussian case

On the other hand, for any integer $m \geq 1$ one may find $n \geq 1$ such that $n^3 \leq m \leq (n+1)^3$.

One has the decomposition

$$\begin{aligned} S_m(\theta) - S_n(\theta) &= \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \cdots - \frac{1}{\sqrt{m}} \sum_{k=1}^{n^3} \cdots \\ &+ \left(\sqrt{\frac{n^3}{m}} - 1 \right) S_{n^3} \end{aligned}$$

Setting $\Delta_{m,n^3} := \mathbb{E}_\theta\left[\left|S_m(\theta) - S_{n^3}(\theta)\right|^2\right]$, we deduce

$$\begin{aligned}
\Delta_{m,n^3} &\leq 2\left(\sqrt{\frac{n^3}{m}} - 1\right)^2 \underbrace{\mathbb{E}_\theta\left[S_{n^3}(\theta)^2\right]}_{=1} + \\
&\frac{2}{n^3}\mathbb{E}_\theta\left[\left|\sum_{k=n^3+1}^{m} a_k\cos(k\theta) + b_k\sin(k\theta)\right|^2\right] \\
&\leq \underbrace{2\left(\sqrt{\frac{n^3}{(n+1)^3}} - 1\right)^2 + 2\frac{(n+1)^3 - n^3}{n^3}}_{=O\left(\frac{1}{n}\right)} = O\left(\frac{1}{m^{\frac{1}{3}}}\right).
\end{aligned}$$

Then, almost surely w.r.t. $\mathbb{P}$, Δ_{m,n^3} tends to zero and this ends the proof.

Generalization1 a quantified estimate

Theorem

Suppose that $(a_k, b_k)_{k\geq 1}$ *is a sequence of independent and identically distributed random variables such that* $\mathbb{E}[a_1] = 0, \mathbb{E}\left[a_1^2\right] = 1$ *and* $\mathbb{E}\left[a_1^4\right] < \infty$*. Setting,*

$$C(a_1) := 81\sqrt{13 + |\mathbb{E}[a_1^3]|} + 8\sqrt{\mathbb{E}\left[a_1^4\right]} + \sqrt{2} + 8\mathbb{E}\left[|a_1|^3\right] + 24\mathbb{E}\left[|a_1|\right],$$

if $G \sim \mathcal{N}(0,1)$*, then one has*

$$\mathbb{E}\left[d_{\mathcal{C}^3}^{\theta}\left(f_n(\theta), G\right)\right] \leq \frac{C(a_1)}{\sqrt{n}}.$$

where

$$d_{\mathcal{C}^3}(X,Y) = \sup_{\|\phi\|_\infty + \|\phi'\|_\infty + \|\phi''\|_\infty + \|\phi^{(3)}\|_\infty} \mathbb{E}(\phi(X) - \phi(Y)).$$

Generalization2 a total variation CLT

Theorem

Let $(a_k, b_k)_{k\geq 1}$ *be independent and identically distributed random variables that are centered with unit variance and admit a third moment. Almost surely with respect to the probability* $\mathbb{P}$*, if* $G \sim \mathcal{N}(0,1)$ *under* $\mathbb{P}_\theta$*, then as n goes to infinity, we have*

$$\lim_{n\to+\infty} \mathrm{d}_{TV}^{\theta}\left(S_n(\theta), G\right) = 0.$$

Said otherwise, $S_n(\theta)$ admits a (random) density w.r.t to Lebesgue which (almost surely) converges in L^1 to the Gaussian density.

Generalization3: non uniforms distributions

Theorem

Let $(a_k, b_k)_{k\geq 1}$ be a sequence of independent and identically distributed random variables that are centered with unit variance and which admit a moment of order $\beta \geq 3$. Let X be an independent random variable on $[0, 2\pi]$ whose Fourier coefficients satisfy

$$\exists \alpha > 0, \, \forall k \in \mathbb{Z}/\{0\}, \, \left|\widehat{\mathbb{P}_X}(k)\right| \leq \frac{C}{|k|^\alpha}.$$

Then, provided that $\beta > \frac{2}{\min\left(\alpha, \frac{1}{2}\right)}$, $\mathbb{P}$ almost surely, under $\mathbb{P}_X$, one has

$$S_n(X) \xrightarrow[n\to\infty]{\text{law under } \mathbb{P}_X} \mathcal{N}(0,1).$$

Let us introduce the stochastic process $(g_n(t))_{t\in[0,2\pi]}$ defined by

$$g_n(t) := S_n\left(\theta + \frac{t}{n}\right).$$

Theorem

Suppose that $(a_k, b_k)_{k\geq 1}$ is a sequence of independent and identically distributed random variables that are centered with unit variance. Then $\mathbb{P}$ almost surely, as n goes to infinity, the process $(g_n(t))_{t\in[0,2\pi]}$ converges in distribution in the $\mathcal{C}^1$ topology, to a stationary Gaussian process $(g_\infty(t))_{t\in[0,2\pi]}$ with $\sin_c$ covariance function, i.e.

$$\mathbb{E}_\theta[g_\infty(t)g_\infty(s)] = \frac{\sin(t-s)}{t-s}.$$

Application to the number of roots of random trigonometric polynomials

Theorem

Let us consider a random trigonometric polynomial

$$f_n(t) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} a_k \cos(kt) + b_k \sin(kt), \quad t \in \mathbb{R},$$

where (a_k) and (b_k) are i.i.d., with unit variance and a moment of order four. Then, $\mathbb{P}$ almost surely, we have as n goes to infinity

$$\lim_{n\to+\infty} \frac{\mathcal{N}(f_n, [0, 2\pi])}{n} = \frac{2}{\sqrt{3}}.$$

and more generally for any interval $[a, b] \subset [0, 2\pi]$

$$\lim_{n\to+\infty} \frac{\mathcal{N}(f_n, [a, b])}{n} = \frac{b-a}{\pi\sqrt{3}}.$$

Key idea to relate Salem-Zygmund with roots of random polynomials

Lemma

If f is a $2\pi-$periodic function with a finite number of zeros, then for any $0 < h < 2\pi$, we have

$$\frac{h}{2\pi} \times \mathcal{N}(f, [0, 2\pi]) = \mathbb{E}_\theta \left[\mathcal{N}\left(f, [\theta, \theta + h] \right) \right],$$

where θ is a random variable, with uniform distribution in $[0, 2\pi]$.

Set $N = \mathcal{N}(f, [0, 2\pi])$ which is finite by hypothesis, and denote by $x_1, \ldots, x_N$ the zeros of f in $[0, 2\pi]$ and μ_f the associated empirical measure

$$\mu_f := \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}.$$

Naturally, we have for all $a < b$ such that $b - a \leq 2\pi$

$$\mathcal{N}(f,[a,b]) = N \times \int_0^{2\pi} \mathbf{1}_{[a,b] \bmod 2\pi}(t)\mu_f(dt).$$

If θ is uniform in $[0,2\pi]$, we have then applying Fubini inversion of sums

$$\begin{aligned}
\mathbb{E}_\theta\left[\mathcal{N}\left(f,[\theta,\theta+h]\right)\right] &= \frac{1}{2\pi}\int_0^{2\pi}\mathcal{N}\left(f,[x,x+h]\right)dx \\
&= \frac{N}{2\pi}\int_0^{2\pi}\left(\int_0^{2\pi}\mathbf{1}_{[x,x+h] \bmod 2\pi}(t)dx\right)\mu_f(dt) \\
&= \frac{N}{2\pi}\int_0^{2\pi}\left(\int_0^{2\pi}\mathbf{1}_{[t-h,t] \bmod 2\pi}(x)dx\right)\mu_f(dt) \\
&= \frac{N}{2\pi}\times h \times \int_0^{2\pi}\mu_f(dt) = \frac{N}{2\pi}\times h.
\end{aligned}$$

Applying it to $S_n(x) = \frac{1}{\sqrt{n}} \sum_{k=1}^n a_k \cos(kx) + b_k \sin(kx)$ leads to

$$\boxed{\frac{\mathcal{N}(S_n, [0, 2\pi])}{n} = \mathbb{E}_\theta \left(\mathcal{N}(g_n, [0, 2\pi]) \right)}$$

Heuristically the rest of the proof is (as $n \to \infty$)

$$\mathbb{E}_\theta \left(\mathcal{N}(g_n, [0, 2\pi]) \right) \approx \mathbb{E}_\mathbb{P} \left(\mathcal{N}((X_t)_t, t \in [0, 2\pi]) \right)$$

where $(X_t)_{t>0}$ is a stationnary Gaussian process with correlation $\frac{\sin(x)}{x}$.

Indeed, take $f \in \mathcal{C}^1([a, b])$ such that $|f| + |f'| > 0$ on $[a, b]$ (non-degeneracy assumption)

$$u_n \xrightarrow[n \to \infty]{C^1} f \quad \Rightarrow \quad \mathcal{N}(u_n, [a, b]) \to \mathcal{N}(f, [a, b]).$$

So one is left to compute $\mathbb{E}_{\mathbb{P}}\left(\mathcal{N}((X_t)_t, t \in [0,2\pi])\right)$. By using the Kac-Rice formula we have

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left(\mathcal{N}((X_t)_t, t \in [0,2\pi])\right) &= \mathbb{E}\left(\lim_{\delta\to 0}\int_0^{2\pi} |X_t'|\mathbf{1}_{|X_t|<\delta}\frac{dt}{2\delta}\right) \\
&= \lim_{\delta\to 0}\int_0^{2\pi} \mathbb{E}\left(|X_t'|\mathbf{1}_{|X_t|<\delta}\right)\frac{dt}{2\delta} \\
\text{(Independence of} X_t \ \text{ and } \ X_t') &= \int_0^{2\pi} \mathbb{E}\left(|X_t'|\right)\lim_{\delta\to 0}\mathbb{P}(|X_t|<\delta)\frac{dt}{2\delta} \\
&= 2\pi\mathbb{E}\left(|X_0'|\right)\frac{1}{\sqrt{2\pi}} \\
&= \sqrt{2\pi}\sigma(X_0')\sqrt{\frac{2}{\pi}} \\
&= 2\sqrt{-\operatorname{sin}_c{}''(0)} \\
&= \frac{2}{\sqrt{3}}
\end{aligned}$$

The case of dependent Gaussian coefficients

We consider here $(a_k)_{k\geq 1}$ and $(b_k)_{k\geq 1}$ two independent realizations of a stationary Gaussian field such that

- $\mathbb{E}(a_k a_l) = \rho(k-l)$ with ρ the correlation function,
- $\rho(k) = \hat{\mu}(k)$ with μ the spectral measure,
- μ possesses an absolutely continuous part ψ_μ such that $\int_0^{2\pi} |\log(|\psi_\mu(x)|)| dx < \infty$.

$$\mathbb{P} - a.s., \quad \frac{\mathcal{N}(f_n, [0, 2\pi])}{n} \to \frac{2}{\sqrt{3}}.$$

The case of dependent Gaussian coefficients

The conclusions fails if ψ_μ vanishes: (**if** ψ_μ **is continuous**):

$$\mathbb{P} - a.s., \; \frac{\mathcal{N}(f_n, [0, 2\pi])}{n} \to \frac{1}{\pi\sqrt{3}} \lambda\left(\psi_\mu \neq 0\right) + \frac{1}{\pi\sqrt{2}} \lambda\left(\psi_\mu = 0\right).$$

Conjecture1: The previous formula holds true whenever ψ_μ is measurable

Conjecture2: $\frac{2}{\sqrt{3}}$ is the least possible value of the asymptotic mean number of roots among all possible correlations of $(a_k)_{k\geq 1}$, $(b_k)_{k\geq 1}$.