Integration by parts and convergence in distribution norms in the CLT

Vlad Bally, L. Caramellino, G. Poly

University de Marne-la-Vallée, Tor Vergata and Rennes

Journés de Probas 2021

CLT $X_i \in R^d, i \in N, \quad i.i.d. \quad with \quad E(X_i) = 0, E(X_i^k X_i^p) = \delta_{k,p}$

$$E(\phi(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i)) \to \int \phi(y)\gamma(y)dy$$

$\phi \in C_b \quad \to \quad$ weak convergence

ϕ measurable bounded $\quad \to \quad$ convergence in total variation

THEOREM (Prohorov) Convergence in total variation is equivalent to : there exists m such that

$$Y_1 = X_1 + ... + X_m \sim f(x)dx + \nu(dx).$$

Remark

$$Z = Y_1 + Y_2 \sim g(x)dx + \mu(dx) \quad g \in C(R)$$

Consequence

$$g \geq \varepsilon \mathbf{1}_{B_r(x_0)} \geq \varphi_r(x - x_0) \quad \varphi_r \in C^\infty$$

Splitting

$$Z \sim \xi \times V + (1 - \xi)U$$

With : ξ, V, U independent with laws

$$P(V \in dx) = \frac{1}{c_r}\varphi_r(x - x_0)dx \quad \varphi_r \simeq \mathbf{1}_{B_r(0)}$$
$$P(\xi = 1) = \varepsilon c_r, \quad P(\xi = 0) = 1 - \varepsilon c_r$$
$$P(U \in dx) = \frac{1}{1 - \varepsilon c_r}(P(Z \in dx) - \varepsilon\varphi_r(x - x_0)dx)$$

Strategy : we use an "abstract Malliavin calculus" based on $V_1, V_2, ..., V_n, ...$ in order to prove Prohorov's Theorem.

Convergence in total variation.

$$\left|E(f(S_n) - \int f(x)\gamma(x)dx\right| \leq \frac{C}{\sqrt{n}} \|f\|_\infty$$

and moreover **Convergence in (almost) distribution norms :** let $|\alpha| = q$.

$$\left|E(\partial_\alpha f(S_n) - \int \partial_\alpha f(x)\gamma(x)dx\right| \leq \frac{C}{\sqrt{n}} \|f\|_\infty + e^{-cn} \|f\|_{q,\infty}$$

with

$$\|f\|_{q,\infty} = \sum_{|\alpha|\leq q} \|\partial_\alpha f\|_\infty .$$

Corollary

$$\gamma_\varepsilon(x) = \frac{1}{(2\pi\varepsilon)^{1/2}} e^{-\frac{|x|^2}{2\varepsilon}}$$

Then, with $\varepsilon_n = e^{-c'n}$

$$\left|E(\partial_\alpha f * \gamma_{\varepsilon_n}(S_n) - \int \partial_\alpha f(x)\gamma(x)dx\right| \leq \frac{C}{\sqrt{n}} \|f\|_\infty$$

Corollary If $Z_i \sim p(x)dx, \quad p \in W^{1,1}$ $(\nabla p \in L^1)$ Then $S_n \sim p_{S_n}(x)dx$ and

$$\left\| \partial^\alpha p_{S_n} - \partial^\alpha \gamma \right\|_\infty \leq \frac{C}{\sqrt{n}} \quad \forall\, |\alpha| \leq n$$

Extensions

A. $Z_i, i \in N$ are **not identically distributed**

B **Local theorems (Edgeworth expansions)**

$$\left| E(\partial_\alpha f(S_n) - \int \partial_\alpha f(x) (\sum_{i=0}^{k} \frac{1}{n^{i/2}} \theta_i(x)) \gamma(x) dx \right| \leq \frac{C}{n^{(k+1)/2}} \left\| f \right\|_\infty + e^{-cn} \left\| f \right\|_{q,\infty}$$

Expected number of roots for trigonometric polynomials

$$P_n(t,Y)=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(Y_k^1\cos\frac{kt}{n}+Y_k^2\sin\frac{kt}{n})$$
$$N_n(Y)=card\{t\in(0,n\pi):P_n(t,Y)=0\}.$$

Invarience principle

$$\lim_n\frac{1}{n}E(N_n(Y))=\lim_n\frac{1}{n}E(N_n(G))=\frac{1}{\sqrt{3}}$$

Kac-Rice lemma let $f:[0,a]\to R$ differentiable such that

$$\inf_{t\le a}(|f(t)|+\left|f'(t)\right|)>0.$$

Then

$$N_a(f)=\lim_{\delta\to 0}\int_0^a\left|f'(t)\right|1_{\{|f(t)|\le\delta\}}\frac{dt}{2\delta}.$$

We use the CLT for $S_n = (S_n^1, S_n^2) \in R^2$ with

$$S_n^1(t,Y) = P_n(t,Y) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (Y_k^1 \cos\frac{kt}{n} + Y_k^2 \sin\frac{kt}{n})$$

$$S_n^2(t,Y) = \partial_t P_n(t,Y) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{k}{n} (Y_k^2 \cos\frac{kt}{n} - Y_k^1 \sin\frac{kt}{n})$$

We take

$$\Phi_\delta(x^1, x^2) \simeq |x_2| \times 1_{(-\infty,0)}^{(\delta)}(x^1) \quad \to \quad \partial_1 \Phi_\delta(x_1, x_2) \approx x_2 \delta_0(x_1)$$

K-R lemma

$$N_n(Y) = \lim_{\delta \to 0} \int_0^{n\pi} \partial_1 \Phi_\delta(S_n^1(t,Y), S_n^2(t,Y))dt$$

CLT

$$\lim_n E(\partial_1 \Phi_\delta(S_n^1(t,Y), S_n^2(t,Y))) = \lim_n E(\partial_1 \Phi_\delta(S_n^1(t,G), S_n^2(t,G))$$

Varience

$$\lim_n \frac{1}{n} Var(N_n(G)) = c(G) \sim 0.56.$$

We prove (non-universality) :

$$\lim_n \frac{1}{n} Var(N_n(Y)) = c(G) + 30(E(Y_1^4) - 3).$$

Basic quantity

$$\frac{1}{n} \int_0^{\pi n} dt \int_0^{\pi n} ds \quad \partial_1 \Phi_\delta(S_n(t,Y)) \partial_1 \Phi(S_n(s,Y))$$

We use CLT for

$$\overline{S}_n(t,s,Y) = (S_n(t,Y), S_n(s,Y)) \in R^4$$

with the Edgorth developpement of order 4.

Integration by parts

$$F = f(\xi_i V_i + (1-\xi_i)U_i, i = 1, ..., n)$$
$$P(V \in dx) = \frac{1}{c_r}\varphi_r(x - x_0)dx \quad \varphi_r \simeq 1_{B_r(0)}$$

Derivatives

$$D_k F = \xi_k \frac{\partial}{V_k} f(\xi_i V_i + (1-\xi_i)U_i, i = 1, ..., n)$$

Duality

$$E(\langle DF, DG\rangle) = \sum_{k=1}^{n} E(D_k F D_k G) = E(FLG)$$

with

$$LG = -\sum_{k=1}^{N} D_k D_k G + D_k G \partial_v \ln \varphi_r(V_k - x_0)$$

Covarience matrix : $F=(F_1,...,F_d)$

$$\sigma_F^{i,j}=\left\langle DF_i, DF_j\right\rangle, \quad i,j=1,...,d.$$

Difficulty

$$P(\det\sigma_F=0)\geq P(\xi_1=...=\xi_n=0)>0.$$

Hoffding's inequality

$$P(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i<\frac{c_r}{2})\leq e^{-an}.$$

Regularization lemma $|\alpha|=m$

$$\begin{aligned}|E(\partial^\alpha f(F))-E(\partial^\alpha(f*\gamma_\varepsilon)(F)| &\leq C\left\|f\right\|_{m,\infty}P^{1/2}(\det\sigma_F\leq\eta)+\frac{\varepsilon}{\eta^{2m}}\left\|f\right\|_\infty \\ &\leq C\left\|f\right\|_{m,\infty}e^{-an}+\frac{\varepsilon}{\eta^{2m}}\left\|f\right\|_\infty\end{aligned}$$

Here γ_ε is a **super kernel**

1. Bally, V., Caramellino, L., Poly, G. (2018) Convergence in distribution norms in the CLT for non identical distributed random variables. Electronic Journal of Probability, 23, no. 45, 1-51.
2. Bally, V., Caramellino, L., Poly, G. (2017) On the non universality for the variance of the number of real roots of random trigonometric polynomials. Preprint `arXiv :1711.0336` 43 pages. November 2017 Probability Theory and Related Fields
3. Bally, V., Caramellino, L., Poly, G. (2019) : Regularization lemmas and convergence in total variation. ArXiv :1907.12328, 2019. Electronic Journal of Probability