$$\boxed{\text{Convergence mod-}\phi\text{ et estimées précises de déviation}}$$

(joint work w.
P.-L. Méliot, Orsay
A. Nikeghbali, U. Zurich)

I Introduction

Sommes de variables i.i.d. $S_n = \sum_{i=1}^{n} X_i$

Enormément de résultats :
- TCL
- vitesse de convergence (Berry-Esseen)
- théorème limite local (estimée pour $\mathbb{P}(S_n = k)$) $k = O(\sqrt{n})$
- (*) grande déviation (estimée pour $\mathbb{P}(S_n \geq \varepsilon n)$)
  (ou déviation modérée pour $\mathbb{P}(S_n \geq x_n)$) $\sqrt{n} \ll x_n \ll n$

cet exposé

Outil : fonction caractéristique

Fait : Beaucoup de modèles ont une fonction caractéristique qui "ressemble" à celle de S_n. $\rightsquigarrow$ on va formaliser ça avec la notion de "cv mod-φ"

But :
- obtenir automatiquement (*) ;
- voir influence du facteur correctif sur (*).

II Convergence mod-ϕ

Notation : $(X_n)_{n\geq 1}$ v.a. réelles

$\varphi_n(z) := \mathbb{E}(e^{zX_n})$ MGF

ϕ distribution de proba sur $\mathbb{R}$ supposée infiniment divisible (pour être représentée Σ somme de n i.i.d pour tout n.) $\to \int_{\mathbb{R}} e^{zx}\phi(dx) = \exp(\eta(z))$ pour un certain η CGF

$S_{(c,d)} := \{ z : c < \operatorname{Re} z < d \}$. On supposera $c < 0 < d$.

$\psi : S_{(c,d)} \to \mathbb{C}$ analytique, qui satisfait $\psi(z) \neq 0$ pour z réel (dans $]c;d[$)

Def: On dit que X_n converge mod-Φ sur $S_{(c,d)}$ avec paramètres $(t_n)_{n\geq 1}$ et fonction limite Ψ si

(Nikeghbali - Kowalski) ~ similar to Hwang quasi powers '96 (but stronger)

$$\Psi_n(t) := \exp\left(-t_n \eta(z)\right) \mathbb{E}\left(e^{zX_n}\right) \longrightarrow \Psi(z) \quad \text{pour } z \in S_{(c,d)}$$

en particulier, on suppose que ça existe pr $z \in S_{(c,d)}$

uniforme cv on comp. subsets of $S_{(c,d)}$

bon cadre pour estimées de déviations

Many other examples:

dans cet exposé
- $\log(X_n)$ (polynôme charactéristique) of random matrix ensembles,
- subgraph counts in $G(n,p)$,
- patterns in random permutations
- prime divisors in random integers
- magnetization in Ising / Curie-Weiss models.

~~First~~ Premier example: X_n = nb cycles in σ_n ; $\sigma_n \in_u S_n$

Fait: $X_n \overset{(d)}{=} \sum_{i=1}^{n} B_i$, avec $B_i \sim \mathrm{Bern}(1/i)$ (B_i) indépendantes

$$\mathbb{E}(e^{zX_n}) = \prod_{i=1}^{n}\left(1 + \frac{1}{i}(e^z - 1)\right) = \frac{\prod_{i=1}^{n}(e^z + i + 1)}{n!}$$

$$= \frac{\Gamma(n + e^z)}{\Gamma(n+1)\,\Gamma(e^z)} \underset{\text{Stirling}}{\sim} \frac{1}{\Gamma(e^z)} \exp\left[\log(n)\cdot(e^z - 1)\right]$$

i.e. X_n converge mod-Poisson vers $\frac{1}{\Gamma(e^z)}$ avec paramètres $t_n = \log(n)$. sur $\mathbb{C}$

III Normalité asymptotique et grandes déviations

Prop Supposons que X_n conv mod ϕ
Alors $\tilde{X}_n := \dfrac{X_n - t_n\eta'(0)}{\sqrt{t_n\eta''(0)}} \xrightarrow{d} \mathcal{N}(0,1)$.

Preuve : $\mathbb{E}\left(e^{it\tilde{X}_n}\right) = \mathbb{E}\left[e^{it/\sqrt{t_n\eta''(0)}\cdot X_n}\right] e^{-it\sqrt{t_n}\frac{\eta'(0)}{\sqrt{\eta''(0)}}}$

$$= e^{-t\sqrt{t_n}\frac{\eta'(0)}{\sqrt{\eta''(0)}}} \underbrace{e^{t_n\eta\left(\frac{it}{\sqrt{t_n\eta''(0)}}\right)}}_{e^{t_n\left(\underset{0}{\underbrace{\eta(0)}} + \eta'(0)\frac{it}{\sqrt{t_n\eta''(0)}} - \frac{\eta''(0)}{2}\frac{t^2}{t_n\eta''(0)}\right)}} \cdot \underbrace{\psi_n\left(\frac{it}{\sqrt{t_n\eta''(0)}}\right)}_{\to\, \psi(0) = 1}$$

$$\longrightarrow e^{-t^2/2} \quad \blacksquare$$

Def (transformée de Legendre-Fenchel) :

$$F(x) := \sup_{h\in\mathbb{R}} hx - \eta(h)$$

Note : pour nous, η convex (CGF are convex by Hölder ineq.)
et h optimal vérifie $\eta'(h) = x$.

Thm : Fixons $x \in \left]\eta'(0), \eta'(d)\right[$. Supposons X_n cv mod-ϕ.
et déf h par éq. $\eta'(h) = x$

Alors $\mathbb{P}[X_n \geq t_n x] \sim \begin{cases} \dfrac{\exp(-t_n F(x))}{h\sqrt{2\pi t_n \eta''(h)}}\psi(h) & \text{si } \phi \text{ ne vit pas sur un réseau } (\mu + \lambda\mathbb{Z}) \\ \dfrac{\exp(-t_n F(x))}{(1-e^{-h})\sqrt{2\pi t_n\eta''(h)}}\psi(h) & \text{si } \mathbb{Z} \text{ est le réseau minimal de } \phi. \text{ et } t_n x \in \mathbb{Z}. \end{cases}$

∃ résultat similaire pr $\mathbb{P}[X_n \leq t_n x]$ avec $x \in \left]\eta'(c); \eta'(0)\right[$

Note : on obtient facilement $\forall h$, $\mathbb{P}[X_n \geq t_n x] \leq \dfrac{\mathbb{E}(e^{hX_n})}{e^{hx\cdot t_n}} \leq \exp\left[-t_n(hx - \eta(h))\right](\psi(h) + o(1))$
(Chernoff bound)

Méthode (classique en grande déviation) :

changement de mesure : on définit Y_n de la manière suivante

$$\mathbb{E}[f(Y_n)] = \frac{1}{\varphi_{X_n}(h)} \mathbb{E}\left[e^{hX_n} f(X_n)\right] \quad \forall f \text{ continue bornée}$$

$$\frac{\text{on déf. } Y_n \text{ par}}{\varphi_{Y_n}(z) = \mathbb{E}(e^{zY_n})} = \frac{1}{\varphi_{X_n}(h)} \mathbb{E}\left[e^{(h+z)X_n}\right] = \varphi_{X_n}(h+z) / \varphi_{X_n}(h)$$

Fait $\Rightarrow Y_n$ cv mod $\tilde{\phi}$ avec paramètre t_n et fonction limite $\tilde{\psi}$ sur $S_{(c-h, d-h)}$

où $\int e^{zy} \tilde{\phi}(dy) = \exp(\eta(z+h) - \eta(h))$, $\tilde{\psi}(z) = \dfrac{\psi(z+h)}{\psi(h)}$.

"Preuve" du thm ds le cas "non-réseau" :

$$\mathbb{P}[X_n \geq t_n \eta'(h)]$$
$$= \varphi_{X_n}(h) \cdot \mathbb{E}\left[e^{-hY_n} \mathbb{1}[Y_n \geq t_n \eta'(h)]\right]$$

on pose $\tilde{Y}_n = \dfrac{Y_n - t_n \eta'(h)}{\sqrt{t_n \eta''(h)}}$

Comme Y_n cv mod $\tilde{\phi}$, $\tilde{Y}_n$ tend vers $\mathcal{N}(0,1)$.

$$= \underbrace{\varphi_{X_n}(h)}_{\substack{\text{tend vers} \\ \psi(h)}} \cdot \underbrace{e^{t_n \eta(h)} \cdot e^{-h \cdot t_n \eta'(h)}}_{e^{-t_n F(x)}} \underbrace{\mathbb{E}\left[e^{-h\sqrt{t_n \eta''(h)} \cdot \tilde{Y}_n} \mathbb{1}[\tilde{Y}_n \geq 0]\right]}_{\text{appelons-le } A}$$

$$A \underset{\substack{\uparrow \\ \text{comme } \tilde{Y}_n \to \mathcal{N}(0,1)}}{\sim} \frac{1}{\sqrt{2\pi}} \int_{u=0}^{\infty} e^{-h\sqrt{t_n \eta''(h)} \cdot u} e^{-u^2/2} du \underset{\substack{\uparrow \\ \text{exercice} \\ \text{(IPP)}}}{\sim} \frac{1}{h\sqrt{t_n \eta''(h)}}$$

⚠ on triche ici car $e^{-h\sqrt{t_n \eta''(h)} \cdot u}$ dépend de n

→ il faut une estimée de l'approximation $\tilde{Y}_n \to \mathcal{N}(0,1)$.
(type Berry-Essein)

" ■ "

Plus généralement, on peut regarder

$$B := \mathbb{P}\left[X_n \geq t_n\, \eta'(0) + \sqrt{t_n\, \eta''(0)}\, \tilde{x}_n\right] = \mathbb{P}\left[X_n \geq t_n x_n\right]$$

$\tilde{x}_n = \Theta(\sqrt{t_n})$ gde déviation : équivalent dépend de ϕ et de Ψ.

correction à la normalité
ne dépend que de ϕ

$\sim \dfrac{\exp(-t_n F(x_n))}{\tilde{x}_n \sqrt{2\pi}}$ $\quad x_n = \eta'(0) + o(1)$

(plus on s'approche de $\sqrt{t_n}$, plus il faut prendre en compte de termes ds dvt de F pr avoir un éq explicite)

$\tilde{x}_n = \Theta(t_n^{1/6})$

$\tilde{x}_n = O(1)$ normalité asymptotique

$B \sim \mathbb{P}[Z \geq \tilde{x}_n]$

IV - Applications

A - Nombre de cycles dans une perm. aléatoire

$\phi = \text{Poisson}(1)$ $\qquad F(x) = x \log x - x + 1$ $\qquad \Psi(h) = \Gamma(e^h)^{-1} = \Gamma(x)^{-1}$

$\eta = e^z - 1$ $\qquad h = \log(x)$

Rappel $t_n = \log(n)$, $\Psi(h) = \Psi(\log x) = \dfrac{1}{\Gamma(x)}$

Pr $x > \eta'(0) = 1$

On obtient $\mathbb{P}\left[X_n \geq x(\log n)\right] \sim \dfrac{n^{-(x\log x - x + 1)}}{\sqrt{2\pi x \log n}}\ \dfrac{1}{1 - 1/x}\ \dfrac{1}{\Gamma(x)}$

B - $\mathrm{Re}[\log(X_M)]$ de matrice aléatoire

Soit $M_n \in U(n)$ distr selon la mesure de Haar

$X_n = \mathrm{Re}\left[\log(\det(1 - M_n))\right]$ could be replaced by any z with $|z|=1$

Motivations : · analogue continu nb de cycles ← mult. de 1 ds matr. de permutations

· "modèle" pour $Y_n = \zeta(\frac{1}{2} + i\tau)$ pr $\tau \in_u [T; 2T]$.
avec $n = \log(T/2\pi)$

Prop (based on Keating-Snaith '00) :

X_n cv mod-Gaussien vers $\Psi(z) = \dfrac{G(1+\frac{z}{2})^2}{G(1+z)}$, paramètre $t_n = \dfrac{\log n}{2}$ sur $S_{(-1;+\infty)}$

G : Barnes' G-function
déf par $G(z+1) = \Gamma(z)G(z)$ $\quad G(1) = 1$

$$\text{Prop: } \mathbb{P}\left(X_n \geq x \cdot \frac{\log n}{2}\right) \underset{x>1}{\sim} \frac{G\left(1+\frac{x}{2}\right)^2}{G(1+x)} \frac{1}{x\, n^{x^2/4} \sqrt{\pi \log n}}$$

Mod-Gaussien $\eta(z) = \frac{z^2}{2}$ $h = x$ $F(x) = \frac{x^2}{2}$

→ extension to CβE ensemble
(Dal Borgo-Hovhannisyan-Rouault '19)

→ extension to the range $\mathbb{P}(X_n \geq a_n)$ avec $\log n \ll a_n \lesssim n$
(Méliot - Nikeghbali '22)

→ links with number theory prediction
(Amzallag - Arguin - Bailey - Hui - Rao '21)

C - Comptes de sous-graphes dans $G(n,p)$

X_n = nb de triangles dans $G(n,p)$

(on suppose ici p constant)

Connu: $\tilde{X}_n = \frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathrm{Var}(X_n)}} \xrightarrow{d} \mathcal{N}(0,1)$. ($\mathbb{E}[X_n] = \binom{n}{3}p^3$; $\mathrm{Var}(X_n)$ ordre n^4)

(Rucinski '88, avant?)

Obtenu par méthode de moments, on ne connait pas $\mathbb{E}(e^{zX_n})$ explicitement

→ mais on peut borner les cumulants

Rappel: $\log(\mathbb{E}(e^{zX_n})) = \sum_{r\geq 1} \frac{\kappa^{(r)}(X_n)}{r!} z^r$ ← cumulants

Lemme: $|\kappa^{(r)}(X_n)| \leq 2^{r-1}\, r^{r-2}\, n^{r+2}\, 81^{r-1}$

après le corollaire: Note: une borne $|\kappa^{(r)}(X_n)| \leq C_r\, n^{r+2}$ pr tt r fixe implique
$\frac{X_n - \mathbb{E}[X_n]}{n^2} \to$ loi normale. (cf Janson '88)

• pour cv mod-Gaussienne, on a besoin de la borne uniforme en r ci-dessus

Corollaire : $Y_n = \dfrac{X_n - \mathbb{E}(X_n)}{n^{5/3}}$ cv mod-Gaussien vers

$$\psi(z) = \exp\left(\frac{L}{6} z^3\right) \text{ avec paramètres } t_n = \sigma n^{2/3}$$

$$L = 108 p^7 (1-p)(7-8p) \qquad \sigma = 18 p^5 (1-p)$$

Estimée pour proba de déviation

$$\mathbb{P}\left[X_n \geq \binom{n}{3} p^3 + \frac{v}{6} n^2\right] \quad \text{pr } 1 \ll v \ll n^{1/2}$$

$$\simeq \sqrt{\frac{9p^5(1-p)}{\pi v^2}} \exp\left(-\frac{v^2}{36p^5(1-p)} + \frac{(7-8p)v^3}{324\, n\, p^8 (1-p)^2}\right)$$

(mentionner Saulis-Statulevicius)

→ voir aussi Goldschmidt – Griffiths – Scott '20
for extensions through martingales...