Méthode(s) de Stein

Eddie Hamani
25/06/2015

Références :
- Ledoux, Nourdin, Peccati – Stein's method, logarithmic Sobolev and transport inequalities – 2015
- Nourdin, Peccati – Stein's method on Wiener chaos – 2008

- Chatterjee – 2007, 2010, 2015 (Survey)

But : Décrire les approches de "Stein" en concentration et en proximité gaussienne et potentiellement les liens qu'elles entretiennent.

Lemme de Stein : Soit X une v.a.r. Il y a équivalence entre

(i) $X \sim N(0,1)$

(ii) $\forall f \in \mathcal{C}^1$ telle que $\mathbb{E}\,|f'(X)| < \infty$, $\mathbb{E}\left[f'(X) - Xf(X)\right] = 0$

On appelle opérateur de Stein l'application T définie par

$$Tf(x) = f'(x) - xf(x)$$

1/ Proximité à la loi normale

But : Obtenir des vitesses de convergence dans des TCL (comme Berry-Esséen)

Idée : le lemme de Stein possède une forme de stabilité, au sens où l'on espère avoir des propriétés du type

$$\forall f \in \mathcal{F}, \ \mathbb{E}\left[f'(X) - Xf(X)\right] \simeq 0 \iff X \approx N(0,1)$$

classe fonctionnelle assez riche

Des distances entre mesures usuelles peuvent se mettre sous la forme

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

Ex : . $\mathcal{F} = \left\{ f : \|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|} \leq 1 \right\}$ $\longrightarrow$ Wasserstein

. $\mathcal{F} = \left\{ \mathbb{1}_B, \ B \in \mathcal{B}(\mathbb{R}) \right\}$ $\longrightarrow$ Variation totale

. $\mathcal{F} = \left\{ \mathbb{1}_{]-\infty, t]}, \ t \in \mathbb{R} \right\}$ $\longrightarrow$ Kolmogorov

Pour lier les deux formes on résout une E.D.O.. À $f \in \mathcal{F}$ fixé, on prend g solution de

$$f(x) - \mathbb{E}\left[f(N)\right] = g_f'(x) - x g_f(x) \qquad (S_f)$$

Comme S_f est une E.DO, on obtient des résultats du type

$$f \in \mathcal{F} \implies g_f \in \mathcal{G}$$

On obtient alors

$$d_{\tilde{\mathcal{F}}}(X,N) = \sup_{f\in\tilde{\mathcal{F}}} \left|\mathbb{E}\left[f(X) - f(N)\right]\right|$$

$$= \sup_{f\in\tilde{\mathcal{F}}} \left|\mathbb{E}\left[g_f'(X) - X g_f(X)\right]\right|$$

$$\leq \sup_{g\in\mathcal{G}} \left|\mathbb{E}\left[g'(X) - X g(X)\right]\right|$$

Selon le contexte, l'analyse du terme de droite est plus ou moins technique

Ex: Calcul de Malliavin chez Nourdin & Peccati

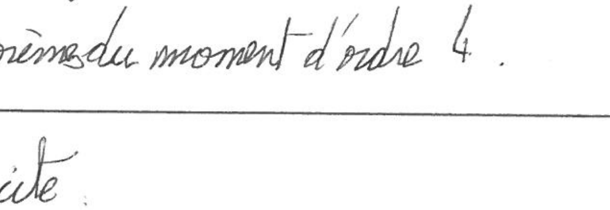
↳ Théorème du moment d'ordre 4.

Voici un exemple explicite :

Prop: Soit $\mathcal{G} = \mathcal{C}^2 \cap \{\|f\|_\infty \vee \|f'\|_\infty \vee \|f''\|_\infty \leq 1\}$. Alors

$$d_{Kol}(X,N) = \sup_{t\in\mathbb{R}} |\mathbb{P}(X\leq t) - \mathbb{P}(N\leq t)| \leq 2 \sup_{g\in\mathcal{G}} \left|\mathbb{E}\left[g'(X) - X g(X)\right]\right|^{\frac{1}{2}}$$

Dem: Soit $\varepsilon > 0$.



Soit g solution de (S_f). Des estimations standard (Chen, 2011) montrent

que $\frac{\varepsilon}{2} g \in \mathcal{G}$.

On écrit,

$$\left(g'(x) - x g(x) = f(x) - \mathbb{E} f(N) \right)$$

$$\begin{aligned}
\mathbb{P}(X \leq t) &\leq \mathbb{E}[f(X)] \\
&= \mathbb{E}[f(N)] + \mathbb{E}[g'(X) - X g(X)] \\
&\leq \underbrace{\mathbb{P}(N \leq t) + \frac{\varepsilon}{\sqrt{2\pi}}}_{\text{La densité de } N(0,1) \text{ est majorée par } \frac{1}{\sqrt{2\pi}}} + \frac{2}{\varepsilon} \mathbb{E}\left[\left(\tfrac{\varepsilon}{2} g\right)'(X) - X \left(\tfrac{\varepsilon}{2} g\right)(X) \right] \\
&\leq \mathbb{P}(N \leq t) + \frac{\varepsilon}{\sqrt{2\pi}} + \frac{2}{\varepsilon} \sup_{h \in \mathcal{G}} \left| \mathbb{E}[h'(X) - X h(X)] \right|
\end{aligned}$$

En regardant un lissage à gauche de $\mathbb{1}_{(-\infty, t]}$ on obtient

$$\mathbb{P}(X \leq t) \geq \mathbb{P}(N \leq t) - \frac{\varepsilon}{\sqrt{2\pi}} - \frac{2}{\varepsilon} \sup_{h \in \mathcal{G}} \left| \mathbb{E}[h'(X) - X h(X)] \right|$$

La preuve se termine par une optimisation de la borne portant sur $\varepsilon > 0$.

2/ Concentration

On dit que X et X' sont échangeables lorsque $(X,X') \overset{\mathcal{L}}{=} (X',X)$

Th (Chatterjee, 2007)

Soit $\mathcal{X}$ métrique séparable, et supposons que (X,X') est échangeable, $X \in \mathcal{X}$

$f : \mathcal{X} \to \mathbb{R}$ et $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ avec

$f, F \in L^2$
$$\begin{cases} F(X,X') = -F(X,X') \\ f(X) = \mathbb{E}\left[F(X,X') \mid X\right] \end{cases}$$ (antisymmétrie)

On pose $\Delta(X) = \frac{1}{2}\mathbb{E}\left[\left|(f(X)-f(X'))F(X,X')\right| \mid X\right]$.

Alors • $\mathbb{E}\, f(X) = 0$

• Si $\mathbb{E}\,\Delta(X) < \infty$, alors $\mathrm{Var}(f(X)) = \frac{1}{2}\mathbb{E}\left[(f(X)-f(X'))F(X,X')\right]$

• Si $\begin{cases} \forall \theta, & \mathbb{E}\left[e^{\theta f(X)}|F(X,X')|\right] < \infty \\ \exists B,C \geq 0, & \Delta(X) \leq B f(X) + C \end{cases}$, alors pour $t \geq 0$,

$$\mathbb{P}(f(X) \geq t) \leq \exp\left(-\frac{t^2}{2C+2Bt}\right) \quad \text{et} \quad \mathbb{P}(f(X) \leq -t) \leq \exp\left(-\frac{t^2}{2C}\right)$$

• $\forall k \geq 0$, $\mathbb{E}\left[f(X)^{2k}\right] \leq (2k-1)^k\, \mathbb{E}\left[\Delta(X)^k\right]$

Rq : * Toute la concentration revient à trouver les bons F et distribution jointe (X,X')

* $0 \leq \Delta(X) \leq B f(X) + C \Rightarrow f(X) \geq -\frac{B}{C}$ donc la déviation à gauche n'est pas étonnante.

Application 1 : Une inégalité de Hoeffding

$$X = \sum_{i=1}^{n} Y_i$$

$Y_i \in L^2$ indépendantes

$\mu_i = \mathbb{E} Y_i = 0$

$\boxed{|Y_i| \leq c_i \quad , \quad \sigma_i^2 = Var(Y_i)}$

Construction de X' : $I \sim \mathcal{U}_{\{1,\dots,n\}} \perp\!\!\!\perp (Y_i)$,

$(Y_i')_{1\leq i\leq n} \overset{\mathcal{L}}{=} (Y_i)_{1\leq i\leq n}$ indépendants du reste

$$X' = \left(\sum_{i\neq I} Y_i\right) + Y_I'$$

$$= X - Y_I + Y_I'$$

On pose $F(x,y) = n(x-y)$ et on vérifie que

$$f(X) = \mathbb{E}\left[F(X,X')|X\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[F(X,X')\,|Y_1,\dots,Y_n\right]|X\right]$$

$$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[n(Y_i - Y_i')\,|Y_1,\dots,Y_n\right]\,\Big|\, X\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} Y_i \,|X\right]$$

$$= X$$

De plus, $\Delta(X) = \frac{n}{2}\mathbb{E}\left[(X-X')^2 | X\right]$

$$= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[(Y_i - Y_i')^2 | X\right]$$ conditionnement par $(Y_1, \dots, Y_n)$

$$= \frac{1}{2}\sum_{i=1}^{m} \mathbb{E}\left[Y_i^2 | X\right] + \mathbb{E}\left[Y_i'^2\right]$$

$$\leq \underbrace{\frac{1}{2}\sum_{i=1}^{n} c_i^2 + \sigma_i^2}_{= C}$$

D'après le théorème, on obtient $\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq t\right) \leq \exp\left(-\frac{t^2}{\sum_{i=1}^{n} c_i^2 + \sigma_i^2}\right)$ □

Plus précis que le Hoeffding habituel

Démonstration du théorème :

- Un calcul préliminaire : Pour $h : \mathcal{X} \to \mathbb{R}$ telle que $\mathbb{E}\left[|h(X)F(X,X')|\right] < \infty$,

$$\mathbb{E}\left[h(X)f(X)\right] = \mathbb{E}\left[h(X)F(X,X')\right]$$ (X,X') échangeable

$$= \mathbb{E}\left[h(X')F(X',X)\right]$$ F antisymétrique

$$= -\mathbb{E}\left[h(X')F(X,X')\right]$$

D'où $\boxed{\mathbb{E}\left[h(X)f(X)\right] = \frac{1}{2}\mathbb{E}\left[(h(X)-h(X'))F(X,X')\right]}$

- $\underline{h=1}$: $\mathbb{E}\left[f(X)\right] = 0$
- $\underline{h=f}$: $\mathrm{Var}(f(X)) = \frac{1}{2}\mathbb{E}\left[(f(X)-f(X'))F(X,X')\right]$

- $\underline{h = e^{\theta f}}$, pour $\theta \in \mathbb{R}$.

On pose $m(\theta) = \mathbb{E}\left[e^{\theta f(X)}\right]$

$$m'(\theta) = \mathbb{E}\left[f(X) e^{\theta f(X)}\right]$$
$$= \frac{1}{2}\mathbb{E}\left[\left(e^{\theta f(X)} - e^{\theta f(X')}\right)F(X,X')\right]$$

Un peu de convexité montre que $|e^x - e^y| \leq \frac{1}{2}(e^x + e^y)|x-y|$, d'où

$$|m'(\theta)| \leqslant \frac{|\theta|}{4}\mathbb{E}\left[|f(X) - f(X')|\left(e^{\theta f(X)} + e^{\theta f(X')}\right)|F(X,X')|\right]$$

F antisymétrique

$$= \frac{|\theta|}{4}\left\{\mathbb{E}\left[2\Delta(X)e^{\theta f(X)}\right] + \mathbb{E}\left[2\Delta(X')e^{\theta f(X')}\right]\right\}$$
$$= |\theta|\ \mathbb{E}\left[\Delta(X)e^{\theta f(X)}\right]$$
$$\leqslant |\theta|\ \mathbb{E}\left[(Bf(X)+C)e^{\theta f(X)}\right]$$

D'où $\quad |m'(\theta)| \leqslant B|\theta|\, m'(\theta) + C|\theta|\, m(\theta)$.

En résolvant cette inéquation différentielle, on trouve

Déviation à droite

Déviation à gauche

$$\begin{cases} m(\theta) \leq \exp\left(\dfrac{C\theta^2}{2(1-\theta)}\right), & 0 \leq \theta < \dfrac{1}{B} \\ m(\theta) \leq \exp\left(\dfrac{C\theta^2}{2}\right), & \theta \leq 0 \end{cases}$$

D'où la concentration par un argument du type Laplace-Chernoff.

- $\underline{h = f^{2k-1}}$: On obtient les inégalités BDG sur les moments

Rq : Similitude avec l'argument de Herbst :

$$\frac{\mathrm{Ent}(e^{\sigma Z})}{\mathbb{E}[e^{\sigma Z}]} \leq \frac{\sigma^2 V}{2} \implies \mathbb{E}[e^{\sigma Z}] \leq e^{\frac{\sigma^2 V}{2}}$$

qui s'encode, pour $\Psi(\sigma) = \log m(\sigma)$, en

$$\sigma \Psi'(\sigma) - \Psi(\sigma) \leq \frac{\sigma^2 V}{2}$$

$$\boxed{\mathrm{Ent}(Y) = \mathbb{E}[Y \log Y] - (\mathbb{E} Y) \log(\mathbb{E} Y)}$$

Tentative de lien avec l'équation de Stein - Construction de F :

Si (X, X') est échangeable, on pose

$$Ph(x) = \mathbb{E}[h(X') \mid X = x]$$

Pour g solution de $\quad g - Pg = f \quad$, $\quad \approx$ "Équation de Stein"

on pose $\quad F(x,y) = g(x) - g(y)$

- F est antisymétrique
- $\mathbb{E}[F(X,X') \mid X] = g(X) - \mathbb{E}[g(X') \mid X]$

$$= g(X) - Pg(X)$$
$$= f(X)$$

Formellement, on prend $\quad g = \sum_{k \geq 0} P^k f \quad$ ce qui donne un choix canonique et explicite de F.

Application 2 : Modèle de Curie-Weiss de magnetisation

$\beta \geq 0$, $h \in \mathbb{R}$. $\mathbb{P}$ définie sur $\{-1,1\}^n$ par

$$\mathbb{P}(\{\sigma\}) = Z^{-1} \exp\left(\frac{\beta}{n} \sum_{i<j} \sigma_i \sigma_j + \beta h \sum_i \sigma_i \right)$$ (mesure de Gibbs)

pour tout $\sigma = (\sigma_1, \dots, \sigma_n)$.

σ_i : spin de la i^e particule
h : champ externe
β : inverse d'une température

On s'intéresse à la concentration de la magnetisation globale

$$m(\sigma) = \frac{1}{n} \sum_{i=1}^{n} \sigma_i .$$

- $\sigma \sim \mathbb{P}$
- σ' définie par $\sigma' = (\sigma_1, \dots, \sigma_{I-1}, \sigma_I', \sigma_{I+1}, \dots, \sigma_n)$

où $\begin{cases} I \sim \mathcal{U}_{\{1,\dots,n\}} \\ \mathcal{L}(\sigma_I') = \mathcal{L}(\sigma_I \mid \sigma_1, \dots, \sigma_{I-1}, \sigma_{I+1}, \dots, \sigma_n) \end{cases}$

- $F(\sigma, \sigma') = \sum_{i=1}^{n} (\sigma_i - \sigma_i')$
- $f(\sigma) = \mathbb{E}\left[F(\sigma, \sigma') \mid \sigma \right] = m(\sigma) - \frac{1}{n} \sum_{i=1}^{n} \tanh(\beta m_i(\sigma) + \beta h)$

où $m_i(\sigma) = \frac{1}{n} \sum_{j \neq i} \sigma_j$

On obtient finalement :

$$\mathbb{P}\left(|m(\sigma) - \tanh(\beta m(\sigma) + \beta h)| \geq \frac{\beta}{n} + \frac{t}{\sqrt{n}} \right) \leq 2 \exp\left(- \frac{t^2}{4(1+\beta)} \right) .$$

□

Rq : $m(\sigma)$ est concentrée autour des racines de l'équation

$$x = \tanh(\beta x + \beta h)$$