Méthodes de Stein et inégalités de concentration

Eddie Aamari
25/10/2018

(1)

Références :
- C. Stein (1972) – A bound for the error in the normal approximation to the distribution of a sum of dependent random variables
- C. Stein (1986) – Approximate computation of expectations
- * Chen, Goldstein & Shao (2011) – Normal approximation by Stein's method
- S. Chatterjee (2014) – A short survey of Stein's method
- * S. Chatterjee – Stein's method for concentration inequalities

---

La méthode de Stein est une technique basée sur la notion de paires échangeables initiallement développée pour obtenir des vitesses de convergence dans des résultats de type TCL, comme Berry-Esseen [1972]. Plus récemment [2007], cette technique a permis d'obtenir des résultats de concentration de la mesure dans des cadres très variés (modèles de spin, matrices aléatoires).
Le but de ces notes est de donner un aperçu de la méthode de Stein en en présentant deux variantes accompagnées d'exemples. On suivra les grandes lignes des articles de Sourav Chatterjee traitant du sujet [2007, 2014].

Note : Charles Stein, 1920 – 2016
Papier fondateur : 1972

1/ Proximité à la loi normale

<u>But</u> : Obtenir des vitesses de convergence dans des TCL (comme Berry-Esseen)

Stein part du constat suivant :

<u>Lemme de Stein</u> : Soit Z une variable aléatoire réelle. Il y a équivalence entre :

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

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

<u>Dem</u> : Cf Section 2.1 de [Chen, Goldstein & Shao, 2011]

<u>Idée</u> : le lemme de Stein possède une forme de stabilité, au sens où l'on espère obtenir des propriétés du type

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

classe fonctionnelle assez riche

Les mesures de proximité entre lois de probabilités se mettent souvent sous la forme

$$d_{\mathcal{F}}(P,Q) = \sup_{f \in \mathcal{F}} \left| \int f \, dP - \int f \, dQ \right|$$

<u>Ex</u> : • $\mathcal{F} = \left\{ f \mid \|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|} \leq 1 \right\}$ $\longrightarrow$ Wasserstein 1

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

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

Pour formaliser l'idée de stabilité du lemme de Stein, on résout, à $f \in \mathcal{F}$ fixée l'EDO

Équation de Stein

$$\boxed{g'(x) - x\,g(x) = f(x) - \mathbb{E}\left[f(Z)\right]}, \qquad (S_f)$$

où $Z \sim N(0,1)$. Si $\mathbb{E}\,|f(Z)| < \infty$, une solution de (S_f) est

$$g_f(x) = e^{\frac{x^2}{2}} \int_{-\infty}^{x} e^{-y^2/2} \left(f(y) - \mathbb{E}\left[f(Z)\right] \right) dy,$$

et si de plus f est Lipschitz,

$$g_f(x) = - \int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbb{E}\left[Z f\left(\sqrt{t}\,x + \sqrt{1-t}\,Z\right)\right] dt.$$

En analysant plus précisément ces formules, on obtient des résultats du type

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

On obtient alors

$$\boxed{\begin{aligned} d_{\mathcal{F}}(X,Z) &= \sup_{f \in \mathcal{F}} \left| \mathbb{E}\left[f(X) - f(Z)\right] \right| \\ &= \sup_{f \in \mathcal{F}} \left| \mathbb{E}\left[g_f'(X) - X g_f(X) \right] \right| \\ &\leq \sup_{g \in \mathcal{G}} \left| \mathbb{E}\left[g'(X) - X g(X) \right] \right|. \end{aligned}}$$

Rq: Selon le contexte, le terme de droite est plus ou moins technique à borner.

↳ Calcul de Malliavin (Voir Nourdin - Pecatti)

↳ À la main, si X a une forme explicite exploitable

↳ Via des paires échangeables → Exemple juste après !

Prenons le cas de la distance de Wasserstein 1 :

$$d_W(X,Z) = \sup_{\|f\|_{Lip}\leq 1} \left| \mathbb{E}\left[f(X) - f(Z) \right] \right| .$$

Lemme : Si f est Lipschitz, alors $g = g_f$ vérifie

$$\|g\|_\infty \leq \|f\|_{Lip} \;,\; \|g'\|_\infty \leq \sqrt{\frac{2}{\pi}} \|f\|_{Lip} \;,\; \|g''\|_\infty \leq 2\|f\|_{Lip}$$

Dem : Cf lemma 2.4 de [Chen, Goldstein & Shao, 2011].

En particulier on obtient $d_W(X,Z) \leq \sup_{\substack{\|g\|_\infty\leq 1 \\ \|g'\|_\infty \leq \sqrt{\frac{2}{\pi}} \\ \|g''\|_\infty \leq 2}} \left| \mathbb{E}\left[g'(X) - Xg(X) \right] \right| .$

Méthode des paires échangeables

Def : On dit que (X,X') est echangeable lorsque $(X,X') \overset{\mathcal{L}}{=} (X',X)$

Prop : Soit (X,X') échangeable telle qu'il existe $\lambda \in (0,1)$ tel que

$$\mathbb{E}(X'-X \mid X) = -\lambda X \quad \text{p.s}$$

Compromis entre
- $|X'-X|$ petit $\Rightarrow \mathbb{E}|X'-X|^3$ petit
- $\frac{1}{\lambda}$ pas trop grand

Si $\mathbb{E}X^2 = 1$, alors

$$d_W(X,Z) \leq \sqrt{\frac{2}{\pi} \operatorname{Var}\left(\mathbb{E}\left(\frac{1}{2\lambda}(X'-X)^2 \mid X \right) \right)} + \frac{1}{3\lambda} \mathbb{E}|X'-X|^3$$

Rq :
- λ doit être vu comme un paramètre petit, typiquement de l'ordre de $\frac{1}{n}$
- X' est obtenue par une petite perturbation de X
- En fait, le résultat reste vrai si $X \overset{\mathcal{L}}{=} X'$ (Pas forcément échangeable)

Dem : On va borner $|\mathbb{E}\, g'(X) - Xg(X)|$ pour $\|g\|_\infty \leq 1$, $\|g'\|_\infty \leq \sqrt{\frac{2}{\pi}}$ et $\|g''\|_\infty \leq 2$, mais d'abord faisons quelques remarques :

- $\mathbb{E}\, X = \mathbb{E}\, X'$ et $\mathbb{E}\, X^2 = \mathbb{E}\, X'^2$ car $X \overset{\mathcal{L}}{=} X'$
- $\boxed{\mathbb{E}\, X = 0}$, car $\mathbb{E}[-\lambda X] = \mathbb{E}[\mathbb{E}(X'-X|X)] = \mathbb{E}[X'-X] = 0$
  et que $\lambda \neq 0$.
- $\boxed{\mathbb{E}(X'-X)^2 = 2\lambda}$, car

$$\begin{aligned}
\mathbb{E}(X'-X)^2 &= \mathbb{E}[X'^2 + X^2 - 2X'X] \\
&= \mathbb{E}[2X^2 - 2X'X] \\
&= \mathbb{E}[2X(X-X')] \\
&= \mathbb{E}[2X\,\mathbb{E}(X-X'|X)] = \mathbb{E}[2\lambda X^2] = 2\lambda .
\end{aligned}$$

Fixons maintenant g telle que $\|g\|_\infty \leq 1$, $\|g'\|_\infty \leq \sqrt{\frac{2}{\pi}}$ et $\|g''\|_\infty \leq 2$.
On note $G(x) = \int_0^x g(y)\,dy$. Clairement, G est trois fois différentiable.
Par un développement de Taylor à l'ordre 3 on a

$$\begin{aligned}
0 &= \mathbb{E}[G(X') - G(X)] \\
&= \mathbb{E}\left[(X'-X)\,g(X) + \frac{1}{2}(X'-X)^2 g'(X) + \text{Reste}\right],
\end{aligned}$$

où $|\text{Reste}| \leq \frac{1}{6}|X'-X|^3 \|g''\|_\infty \leq \frac{1}{3}|X'-X|^3$. Ensuite, on écrit

$$\begin{aligned}
-\lambda \underbrace{\mathbb{E}[X g(X)]}_{\text{Un bout du terme qu'on cherche à borner}} &= \mathbb{E}[\mathbb{E}(X'-X|X)\,g(X)] \\
&= \mathbb{E}[(X'-X)\,g(X)],
\end{aligned}$$

et donc d'après ce qui précède,

$$-\lambda \mathbb{E}[Xg(X)] = -\mathbb{E}\left[\frac{1}{2}(X'-X)^2 g'(X) + \text{Reste}\right]$$

$$= -\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left((X'-X)^2 | X\right) g(X)\right] - \mathbb{E}\left[\text{Reste}\right]$$

En divisant par λ des deux cotés et en ajoutant $\mathbb{E}\, g'(X)$ on a

$$\left|\mathbb{E}[g'(X) - Xg(X)]\right| \leq \left|\mathbb{E}\left[g'(X)\cdot\left(\mathbb{E}\left(\frac{1}{2\lambda}(X'-X)^2|X\right) - 1\right)\right]\right| + \frac{1}{3\lambda}\mathbb{E}|X'-X|^3.$$

Enfin, comme $\|g'\|_\infty \leq 1$ et que $\mathbb{E}\, \frac{1}{2\lambda}(X'-X)^2 = 1$, il vient

$$\left|\mathbb{E}[g'(X) - Xg(X)]\right| \leq \sqrt{\frac{2}{\pi}}\,\mathbb{E}\left|\mathbb{E}\left(\frac{1}{2\lambda}(X'-X)^2|X\right) - 1\right| + \frac{1}{3\lambda}\mathbb{E}|X'-X|^3$$

$$\leq \sqrt{\frac{2}{\pi}\operatorname{Var}\left(\mathbb{E}\left(\frac{1}{2\lambda}(X'-X)^2|X\right)\right)} + \frac{1}{3\lambda}\mathbb{E}|X'-X|^3 \quad \square$$

<u>Application</u> : Somme de variables (indépendantes).

Soient $Y_1, \dots, Y_n$ des variables indépendantes telles que $\begin{cases}\mathbb{E}\, Y_i = 0 \\ \mathbb{E}\, Y_i^2 = 1\end{cases}$.

On considère $X = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$.

<u>Construction de X'</u> : Soient $Y_1', \dots, Y_n'$ des copies indépendantes des $Y_1, \dots, Y_n$, et $I \sim \text{Uniforme}\{1, \dots, n\}$ indépendante de tout le reste.

On construit X' en remplaçant Y_I par Y_I' (par rapport à X). Autrement dit,

$$X' = \frac{1}{\sqrt{n}}\sum_{i \neq I} Y_i + \frac{Y_I'}{\sqrt{n}} = X + \frac{Y_I' - Y_I}{\sqrt{n}}.$$

On vérifie aisément les propriétés suivantes :

- (X, X') est <u>échangeable</u>

- $$\mathbb{E}(X'-X \mid X) = \frac{1}{\sqrt{n}}\, \mathbb{E}(Y'_I - Y_I \mid X)$$

conditionnement par I

$$= \frac{1}{\sqrt{n}} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(Y'_i - Y_i \mid X)$$

$\begin{cases} (Y'_1, \dots, Y'_n) \perp\!\!\!\perp X \\ \mathbb{E}\, Y_i = 0 \end{cases}$

$$= -\frac{1}{n}\, \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \,\Big|\, X\right)$$

$$= -\frac{1}{n} X \qquad \boxed{\lambda = \frac{1}{n}}$$

- $$\frac{1}{3\lambda}\, \mathbb{E}\, |X'-X|^3 = \frac{n}{3 n^{3/2}}\, \mathbb{E}\, |Y_I - Y'_I|^3$$

$$= \frac{1}{3 n^{3/2}} \sum_{i=1}^{n} \mathbb{E}\, |Y_i - Y'_i|^3$$

$$\leq \frac{8}{3 n^{3/2}} \sum_{i=1}^{n} \mathbb{E}\, |Y_i|^3$$

- $$\mathbb{E}\left(\frac{1}{2\lambda}(X'-X)^2 \,\Big|\, X\right) = \frac{1}{2n} \sum_{i=1}^{n} \underbrace{\mathbb{E}\left((Y'_i - Y_i)^2 \mid X\right)}$$

$$= \mathbb{E}\left(Y_i'^2 + Y_i^2 - 2 Y_i Y_i' \mid X\right)$$

$$= 1 + \mathbb{E}\left(Y_i^2 \mid X\right)$$

$$= \frac{1}{2} + \frac{1}{2n} \sum_{i=1}^{n} \mathbb{E}(Y_i^2 \mid X)$$

donc

$$\mathrm{Var}(\downarrow) = \mathrm{Var}\left(\mathbb{E}\left(\frac{1}{2n} \sum Y_i^2 \,\Big|\, X\right)\right)$$

$$\leq \mathrm{Var}\left(\frac{1}{2n} \sum Y_i^2\right) \leq \frac{1}{4 n^2} \sum_{i=1}^{n} \mathbb{E}\, Y_i^4$$

En mettant tout bout à bout, on a montré que

$$\boxed{d_W\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i, Z\right) \leq \frac{1}{\sqrt{n}} \times \left(\sqrt{\frac{1}{2\pi} \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, Y_i^4} + \frac{8}{3} \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}|Y_i|^3\right)}$$

Rq : - le résultat final fait apparaître les moments d'ordre 4 des variables. Cette version de Berry-Esseen n'est donc pas du tout optimale.

- En attaquant directement $|\mathbb{E}\, g'(X) - X g(X)|$ analytiquement (sans paire échangeable) on peut toutefois s'en passer. Voir [Stein 1972] et [Section 2.3.1, Chen, Goldstein, Shao 2011].

Dans ce qui suit, on oublie la proximité avec la loi normale, et on illustre l'utilisation des paires échangeables pour obtenir des résultats de concentration.

2/ Concentration

Énnonçons le résultat principal dans toute sa généralité.

Théorème (Chatterjee, 2007).

Soit $\mathcal{X}$ métrique séparable et $X, X' \in \mathcal{X}$ des variables aléatoires telles que (X, X') est échangeable.
Soient $f : \mathcal{X} \to \mathbb{R}$ et $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ de carrés intégrables telles que

$$\begin{cases} F(X,X') = -F(X',X) & \text{(antisymétrie)} \\ f(X) = \mathbb{E}\left[F(X,X') \mid X\right] \end{cases}$$

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

Alors,

(Espérance) • $\mathbb{E} f(X) = 0$

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

(Déviation) • Si $\begin{cases} \forall \theta \in \mathbb{R}, \ \mathbb{E}\left[e^{\theta f(X)}|F(X,X')|\right] < \infty \\ \exists B, C \geq 0, \ \Delta(X) \leq B f(X) + C \end{cases}$,

alors pour tout $t \geq 0$,

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

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

<u>Rq</u> : • À $f(X)$ fixée, toute la concentration revient à trouver les bons F et distribution jointe (X,X').

• $\Delta(X) \leq B f(X) + C$ revient à dire que "Variance ≤ Espérance", d'où la déviation sous Poissonienne à droite.

• Comme $0 \leq \Delta(X) \leq B f(X) + C$, on a $f(X) \geq -\frac{C}{B}$, d'où la déviation sous-Gaussienne à gauche.

Voici un exemple avant d'attaquer la preuve.

<u>Application 1</u> : Une inégalité de Hoeffding

$$X = \sum_{i=1}^{n} Y_i \quad \text{avec} \quad \begin{cases} Y_i \in L^2 \text{ indépendantes} \\ \mathbb{E} Y_i = 0 \\ |Y_i| \leq c_i \quad \text{et } \sigma_i^2 = Var(Y_i) \end{cases}$$

<u>construction de X'</u> : Pareil que précédemment, $X' = X + (Y_I' - Y_I)$ où $(Y_i')_i \perp\!\!\!\perp (Y_i) \perp\!\!\!\perp I \sim \text{Uniforme}\{1, \dots, n\}$.

On note $F(x, x') = n(x - x')$ et on vérifie que

$$\begin{aligned} f(X) &= \mathbb{E}[F(X,X') \mid X] \\ &= n\,\mathbb{E}[Y_I - Y_I' \mid X] \\ &= n\,\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - Y_i' \mid X\right] \\ &= X - \mathbb{E}\left[\sum Y_I' \mid X\right] \\ &= X \end{aligned}$$

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

$$= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[(Y_i - Y_i')^2 \mid X\right]$$

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

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

D'après le théorème, on obtient

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq t\right) \leq \exp\left(-\frac{t^2}{\sum c_i^2 + \sigma_i^2}\right)$$

⊕ Même borne à gauche.

Un peu plus précis que le Hoeffding habituel.

Démonstration du théorème :

Commençons par un calcul préliminaire important : Pour $h : \mathcal{X} \to \mathbb{R}$ telle que $\mathbb{E}\left[|h(X)F(X,X')|\right] < \infty$, on a

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

(X,X') échangeable

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

F antisymetrique

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

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

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

. $h = e^{\theta f}$: (Pour $\theta \in \mathbb{R}$ fixé) On pose $m(\theta) = \mathbb{E}\left[e^{\theta f(X)}\right]$.

↳ But = borner la transformée de laplace.

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

Avec un peu de convexité, on montre que $|e^a - e^b| \leq \frac{1}{2}(e^a + e^b)|a-b|$, d'où

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

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

F antisymétrique

$$= |\theta|\,\mathbb{E}\left[\Delta(X)e^{\theta f(X)}\right]$$
$$\leq |\theta|\,\mathbb{E}\left[(Bf(X)+C)e^{\theta f(X)}\right],$$

autrement dit,

$$\boxed{\forall \theta \in \mathbb{R},\ |m'(\theta)| \leq B|\theta|\,m'(\theta) + C|\theta|\,m(\theta)}$$

En étudiant cette inéquation différentielle, on obtient

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

(Déviation à droite ⇐)

(Déviation à gauche ⇐)

D'où la concentration par un argument de laplace - Chernoff.

$\mathbb{P}(f(X) \geq t) = \mathbb{P}\left(e^{\theta f(X)} \geq e^{\theta t}\right) \leq \dfrac{\mathbb{E}\left[e^{\theta f(X)}\right]}{e^{\theta t}}$

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

Lien avec l'équation de Stein - Construction de F

Si (X,X') est échangeable, on pose

$$Ph(x) = \mathbb{E}\left[h(X') \mid X = x\right] \quad \text{pour } h : \mathcal{X} \to \mathbb{R} \text{ intégrable}$$

Dans l'optique de faire de la concentration sur $f(X)$, on résout

$$g - Pg = f.$$

≈ "Équation de Stein"

En notant $F(x,x') = g(x) - g(x')$, on a

- F est antisymétrique
- $$\begin{aligned} \mathbb{E}\left[F(X,X') \mid X\right] &= g(X) - \mathbb{E}\left[g(X') \mid X\right] \\ &= g(X) - Pg(X) \\ &= f(X) \end{aligned}$$

Formellement, on prend $g = \sum_{k \geq 0} P^k f$, ce qui donne un choix canonique et explicite de F (pourvu que la paire (X,X') soit construite).

Application 2 : Modèle de Curie-Weiss de magnétisation

Soient $\beta \geq 0$ et $h \in \mathbb{R}$. Pour $\sigma = (\sigma_1, \dots, \sigma_n) \in \{-1, 1\}^n$, la loi $\mathbb{P}$ est définie 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

Constante de normalisation

Interprétation
- σ_i : Spin de la i^e particule
- h : champ magnétique externe
- β : inverse d'une température

On s'intéresse à la concentration de la magnétisation globale

$$\boxed{m(\sigma) = \frac{1}{n} \sum_{i=1}^{n} \sigma_i,} \quad \text{où } \sigma \sim \mathbb{P}$$

Construction de σ' : $\sigma' = (\sigma_1, \dots, \sigma_{I-1}, \sigma_I', \sigma_{I+1}, \dots, \sigma_n)$, où

$$\begin{cases} I \sim \text{Uniforme}\{1, \dots, n\} \\ \mathcal{L}(\sigma_I') = \mathcal{L}(\sigma_I \mid \sigma_i, i \neq I) \end{cases}$$

- (σ, σ') est échangeable
- On note $F(\sigma, \sigma') = \sum_{i=1}^{n} (\sigma_i - \sigma_i')$ et on vérifie que

$$f(\sigma) = \mathbb{E}[F(\sigma, \sigma') \mid \sigma] = m(\sigma) - \frac{1}{n} \sum \tanh(\beta m_i(\sigma) + \beta h)$$

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

On obtient finalement

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

Rq : $m(\sigma)$ est concentrée autour des racines de l'équation $x = \tanh(\beta x + \beta h)$!

Application 3 : Un problème d'appariement généralisé

Soit $\{a_{ij}\}_{ij}$ une matrice réelle $n \times n$. Pour une permutation π tirée uniformément dans $\mathfrak{S}_n$, on s'intéresse à $X = \sum_{i=1}^{n} a_{i\pi(i)}$

"Statistique de Hoeffding"

Ex : . $a_{ij} = \mathbb{1}_{i=j}$: X = nombre de points fixes de π

. $a_{ij} = |i-j|$: X = "Spearman's footrule"

Construction de X' : Soient I, J uniformes sur $\{1, \dots, n\}$ et indépendants, et

$\pi' = \pi \circ (I,J)$, où (I,J) est la transposition de I et J.

On vérifie aisément que (π, π') est échangeable, et donc que (X, X') aussi, où

$$X' = \sum_{i=1}^{n} a_{i\pi'(i)}.$$

Pour $F(x,x') = \frac{1}{2} n (x - x')$,

. $$\mathbb{E}[F(X,X') | \pi] = \frac{n}{2} \mathbb{E}\left[a_{I\pi(I)} + a_{J\pi(J)} - a_{I\pi(J)} - a_{J\pi(I)} \,\middle|\, \pi\right]$$
$$= \frac{1}{n} \sum_{i,j} a_{i\pi(i)} - a_{i\pi(j)}$$
$$= X - \mathbb{E}X = f(X)$$

. $\Delta(X) \leq X + \mathbb{E}[X] = f(X) + 2\mathbb{E}[X]$, d'où

$$\boxed{\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2 \exp\left(-\frac{t^2}{4\mathbb{E}X + 2t}\right)}$$

Ex : X = nombre de points fixes de π : $a_{ij} = \mathbb{1}_{i=j} \longrightarrow \mathbb{P}(|X-1| \geq t) \leq 2 \exp\left(-\frac{t^2}{4+2t}\right)$.

Rq : Pour de vrais résultats de matrices aléatoires avec Stein, voir Mackey, Jordan et al 2014, "Matrix concentration inequalities via the method of exchangeable pairs"