Operations with Concentration Inequalities

Cosme Louart

Assistant Professor at CUHK Shenzhen

SCHOOL OF DATA SCIENCE

Content

I - Motivation in machine learning

Given a data set $\mathcal{D} = ((x_1, y_1), \dots, (x_n, y_n))$ n independent drawings of $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}$

Look for a mapping $\Phi_{\mathcal{D}} : \mathbb{R}^p \to \mathbb{R}$ that "minimizes":

$$L(\Phi_{\mathcal{D}}(X), Y) \qquad \text{For a given a loss } L : \mathbb{R}^2 \to \mathbb{R}_+$$

Behavior of the loss $L(\Phi_{\mathcal{D}}(X), Y)$?

Ex: $\mathcal{D}, X, Y \to L(\Phi_{\mathcal{D}}(X), Y)$ $\lambda_{n,p}$-Lipschitz:

$$\mathbb{P}\left(|L(\Phi_{\mathcal{D}}(X), Y) - L'| \geq t\right) \leq \alpha\left(\frac{t}{\lambda_{np}}\right)$$

Idea: Consider $\alpha : t \mapsto \sup\left\{\mathbb{P}\left(|f(X) - f(X')| \geq t\right),\ f : \mathbb{R}^p \to \mathbb{R}, \text{1-Lipschitz}\right\}$.

Question: $\lim_{t\to\infty} \alpha(t) = 0$? Depends on p?

I - Motivation in machine learning

Theorem: Given $Z \sim \mathcal{N}(\mu, I_n)$, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq 2e^{-\frac{t^2}{2}} \qquad Z, Z' \ i.i.d.$$

Given $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ λ-Lipschitz and $f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z'))| \geq t\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{\lambda} f(\Phi(Z)) - \frac{1}{\lambda} f(\Phi(Z'))\right| \geq \frac{t}{\lambda}\right) \leq 2e^{-\frac{t^2}{2\lambda^2}}.$$

$$\|\Phi(Z) - \Phi(Z')\| \leq \Lambda \|Z - Z'\| \quad a.s.$$

Random

Theorem: (Talagrand)
Given $Z = (Z_1, \dots Z_n) \in [0,1]^n$ s.t. $Z_1, \dots, Z_n$ independent
$\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz and convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq 2e^{-\frac{t^2}{4}}.$$

Michel Talagrand (1995) *Concentration of measure and isoperimetric inequalities in product spaces.* Publications mathématiques de l'IHÉS, 104:905–909.

I - Motivation in machine learning

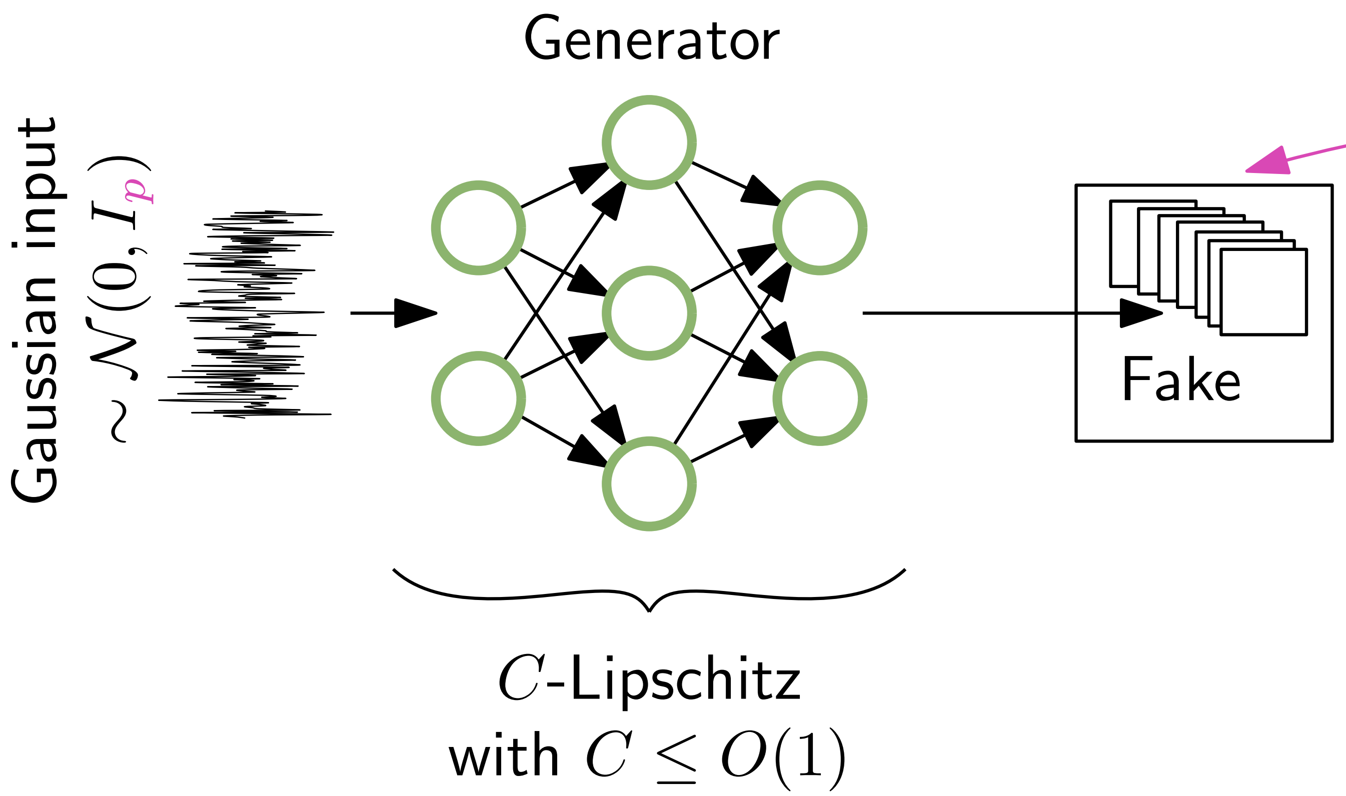
Theorem: Given $Z \sim \mathcal{N}(\mu, I_n)$, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq 2e^{-\frac{1}{2}t^2}$$

Recall: $\forall \Phi : \mathbb{R}^n \to \mathbb{R}^q$ λ-Lipschitz, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t\right) \leq 2e^{-\frac{1}{2}(t/\lambda)^2}.$$

GAN generated images are concentrated vectors



Concentrated by construction

I - Motivation in machine learning

Outside from Gaussian contration:
Possible to set heavy tailed concentration depending on the dimension

Proposition:
Consider $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$, $Z = (Z_1, \ldots, Z_n) \sim \mathcal{N}(0, I_n)$, $\phi_i \in \mathcal{L}_1(\mathbb{R})$ and $\exists h : \mathbb{R}_+ \to \mathbb{R}_+$, increasing s.t.:

- $\forall i \in [n] : \ X_i = \phi_i(Z_i)$
- $\forall t \in \mathbb{R}$, $\forall i \in [n]$: $|\phi_i'(t)| \leq h(|t|)$
- for all $a > 2\log(2n)$, $b > 0$:
  $$h(\sqrt{a+b}) \leq h(\sqrt{a})h(\sqrt{b}).$$

Then $\forall f \in \mathcal{L}_1(\mathbb{R}^n, \mathbb{R})$:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq 3\mathcal{E}_2 \circ (\mathrm{Id} \cdot h)^{-1} \circ \left(\frac{t}{\underbrace{h(\sqrt{2\log(n)})}_{\eta_n}}\right)$$

Where $\mathcal{E}_2 : t \mapsto 2e^{-t^2/2}$

Example: Consider the case $\phi_i|_{\mathbb{R}_+} : t \mapsto e^{t^2/2q} - 1$, $h = \phi_i|_{\mathbb{R}_+}'$.

"Conjecture" : If $\forall r \leq 1 : \ \mathbb{E}[|X_{i,j}|^r] \leq \infty$ Then $\eta_n \leq o(\sqrt{n})$

II - Operation with concentration inequalities.

Definition: $\alpha \boxplus \beta = (\alpha^{-1} + \beta^{-1})^{-1}$

Proposition: Given $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$, two random variables $X, Y \in \mathbb{R}$ such that $\forall t \in \mathbb{R}$:

$$\mathbb{P}\left(X \geq t\right) \leq \alpha(t) \quad \text{and} \quad \mathbb{P}\left(Y \geq t\right) \leq \beta(t)$$

Then $\mathbb{P}\left(X + Y \geq t\right) \leq 2\alpha \boxplus \beta(t)$

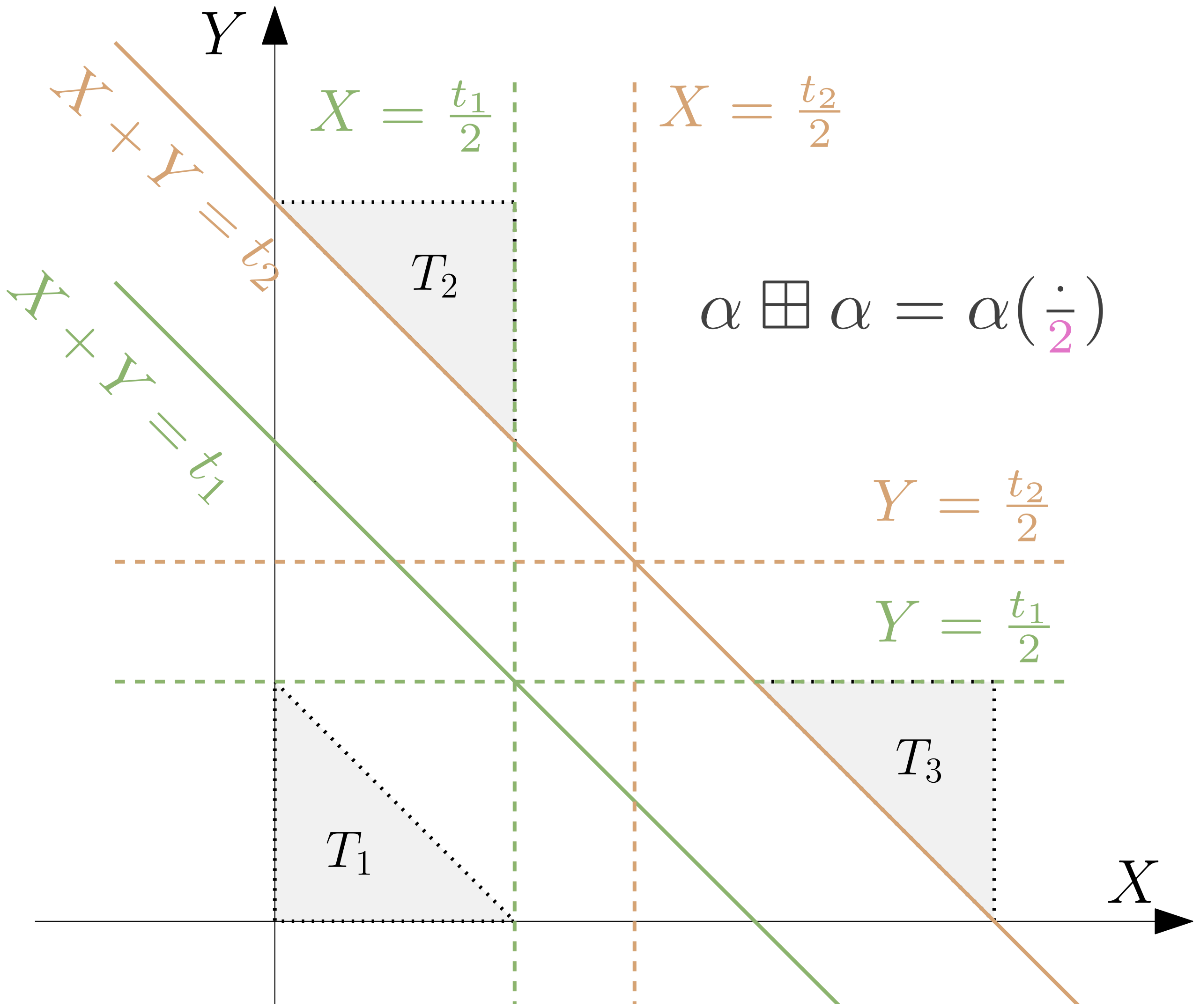
Proof: Denoting $\gamma \equiv \alpha \boxplus \beta$, for any $t \in \mathbb{R}$:

In particular: $\alpha^{-1}(\gamma(t)) + \beta^{-1}(\gamma(t)) = t$

$$\begin{aligned}\mathbb{P}\left(X + Y \geq t\right) &\leq \mathbb{P}\left(X + Y \geq \alpha^{-1}(\gamma(t)) + \beta^{-1}(\gamma(t))\right)\\ &\leq \mathbb{P}\left(X \geq \alpha^{-1}(\gamma(t))\right) + \mathbb{P}\left(Y \geq \beta^{-1}(\gamma(t))\right)\\ &\leq 2\gamma(t)\end{aligned}$$

$\forall t \in [t_1, t_2]$:

$$\mathbb{P}(X + Y \geq t) = \tfrac{2}{3} = \mathbb{P}(X \geq \tfrac{t}{2}) + \mathbb{P}(Y \geq \tfrac{t}{2})$$



Uniform distribution of (X, Y) on T_1, T_2, T_3

II - Operation with concentration inequalities.

Definition: $\alpha \boxtimes \beta \equiv (\alpha^{-1} \cdot \beta^{-1})^{-1}$

Proposition: Given $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$, $X, Y > 0$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(X \geq t\right) \leq \alpha(t) \quad \text{and} \quad \mathbb{P}\left(Y \geq t\right) \leq \beta(t)$$

Then $\mathbb{P}\left(X \cdot Y \geq t\right) \leq 2\alpha \boxtimes \beta(t)$

Proof: Denoting $\gamma \equiv \alpha \boxtimes \beta = (\alpha^{-1} \cdot \beta^{-1})^{-1}$, $\forall t > 0$:

$$\begin{aligned}
\mathbb{P}\left(X \cdot Y \geq t\right) &\leq \mathbb{P}\left(X \cdot Y \geq \alpha^{-1}(\gamma(t)) \cdot \beta^{-1}(\gamma(t))\right) \\
&\leq \mathbb{P}\left(X \geq \alpha^{-1}(\gamma(t))\right) + \mathbb{P}\left(Y \geq \beta^{-1}(\gamma(t))\right) \\
&\leq 2\gamma(t)
\end{aligned}$$

II - Operation with concentration inequalities.

Theorem: Consider $Z \in \mathbb{R}^n$, random, s.t. $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t) \qquad (Z, Z' \text{ i.i.d.})$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(\Lambda(Z) \geq t\right) \leq \beta(t)$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ s.t.:

$$\|\Phi(Z) - \Phi(Z')\| \leq \max(\Lambda(Z), \Lambda(Z'))\|Z - Z'\| \quad a.s.$$

Then: $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz, $\forall t > 0$:

$$\forall t > 0 : \quad \mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq 3\, \alpha \boxtimes \beta(t)$$

Proof: Denote $\Lambda = \Lambda(Z)$, $\Lambda' = \Lambda(Z')$, $\gamma \equiv \alpha \boxtimes \beta = (\alpha^{-1} \cdot \beta^{-1})^{-1}$, $\theta \equiv \beta^{-1}(\gamma(t))$

$$\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq \underbrace{\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t, \max(\Lambda, \Lambda') \leq \theta\right)}_{\leq \mathbb{P}(|h(\Phi(Z)) - h(\Phi(Z')| \geq t) \ \leq \alpha\left(\frac{t}{\beta^{-1}(\gamma(t))}\right)} + \underbrace{\mathbb{P}\left(\max(\Lambda, \Lambda') \geq \theta\right)}_{\leq 2\beta(\beta^{-1}(\gamma(t))}$$

With $h : x \mapsto \sup_{\Lambda(z) \leq \theta} f \circ \Phi(z) - \theta d(x, z)$

$\to$ equal to $f \circ \phi$ on $\{z \in \mathbb{R}^n, \Lambda(z) \leq \theta\}$.

$\to$ $\beta^{-1}(\gamma(t))$-Lipschitz on $\mathbb{R}^n$

$$\leq \alpha(\alpha^{-1}(\gamma(t)))$$

(Since $\forall t > 0 : \ \alpha^{-1}(\gamma(t)) \cdot \beta^{-1}(\gamma(t)) = t$)

II - Operation with concentration inequalities.

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t)$$

with $\alpha : \mathbb{R}^+ \to \mathbb{R}_+$.

Then, Given $d \in \mathbb{N}$, $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ d-times differentiable:

$$\mathbb{P}\left(|\Phi(Z) - m_0| \geq t\right) \leq C_d \ \alpha \circ \beta_0(c_d \ t),$$

where, $\forall k \in [d-1]$, we introduced m_k, a median of $\|d^k\Phi|_Z\|$ and $m_d \ \equiv \sup_{z \in \mathbb{R}^n} \|d^d\Phi|_z\|$.

Proof: Denote: $\beta_k \equiv \left(\frac{\mathrm{Id}}{m_{k+1}}\right)^{\frac{1}{1}} \boxplus \cdots \boxplus \left(\frac{(d-k)!\,\mathrm{Id}}{m_d}\right)^{\frac{1}{d-k}}$

Strategy: Show recursively for $k = d-1, \ldots, 0$:

$$\mathbb{P}\left(\left|\|d^k\Phi|_Z\| - m_k\right| \geq t\right) \leq C \ \alpha\left(c \ \beta_k(t)\right),$$

Last step $k = 0$: Given $t \geq 0$, denote:

$$\mathcal{A}_t \equiv \left\{z \in \mathbb{R}^n : \forall l \in [d], \left|\|d^l\Phi|_z\| - m_l\right| \leq \beta_l^{-1}\left(\beta_0(t)\right)\right\}.$$

Core inference: Φ is ω_t-continuous on $\mathcal{A}_t$ with certain $\omega_t : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\omega_t^{-1}(t) \geq c\beta_0(t)$

- $\mathbb{P}\left(|\Phi(Z) - m_0| \geq t, \ Z \in \mathcal{A}_t\right) \leq 2\alpha \circ \omega_t^{-1}(t) \leq C\alpha(c\beta_0(t))$
- $\mathbb{P}\left(Z \notin \mathcal{A}_t\right) \leq \sum_{l=1}^{d} \mathbb{P}\left(\left|\|d^l\Phi|_z\| - m_l\right| \geq \beta_l^{-1}\left(\beta_k(t)\right)\right) \leq \sum_{l=1}^{d} C \ \alpha\left(c\beta_l \circ \beta_l^{-1} \circ \beta_0(t)\right) \leq C'\alpha(c\beta_0(t))$

Application 1: Heavy tailed concentration

Proposition:
Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$, $\phi_i \in \mathcal{L}_1(\mathbb{R})$ and $\exists h : \mathbb{R}_+ \to \mathbb{R}_+$, increasing s.t.:

- $\forall i \in [n] : \; X_i = \phi_i(Z_i)$
- $\forall t \in \mathbb{R}$, $\forall i \in [n]$: $|\phi_i'(t)| \leq h(|t|)$
- for all $a > 2\log(2n)$, $b > 0$: $h(\sqrt{a+b}) \leq h(\sqrt{a})h(\sqrt{b})$.

Then $\forall f \in \mathcal{L}_1(\mathbb{R}^n, \mathbb{R})$:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq 3\mathcal{E}_2 \circ (\mathrm{Id} \cdot h)^{-1} \circ \left(\frac{t}{h(\sqrt{2\log(n)})}\right)$$

Where $\mathcal{E}_2 : t \mapsto 2e^{-t^2/2}$

Lemma: Given $h : \mathbb{R}_+ \to \mathbb{R}_+$, increasing s.t.:

for all $a > 2\log(2n)$, $b > 0$: $\quad h(\sqrt{a+b}) \leq h(\sqrt{a})h(\sqrt{b})$.

$\min(1, n\mathcal{E}_2 \circ h^{-1}) \leq \mathcal{E}_2 \circ h^{-1} \circ \frac{\mathrm{Id}}{\eta_n}$, $\quad$ with: $\eta_n \equiv h(\sqrt{2\log(n)})$

Application 1: Heavy tailed concentration

Proposition:
Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$, $\phi_i \in \mathcal{L}_1(\mathbb{R})$ and $\exists h : \mathbb{R}_+ \to \mathbb{R}_+$, increasing s.t.:

- $\forall i \in [n] : \; X_i = \phi_i(Z_i)$
- $\forall t \in \mathbb{R}$, $\forall i \in [n]$: $|\phi_i'(t)| \leq h(|t|)$
- for all $a > 2\log(2n)$, $b > 0$:
$$h(\sqrt{a+b}) \leq h(\sqrt{a})h(\sqrt{b}).$$

Then $\forall f \in \mathcal{L}_1(\mathbb{R}^n, \mathbb{R})$:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq 3\mathcal{E}_2 \circ (\mathrm{Id} \cdot h)^{-1} \circ \left(\frac{t}{\underbrace{h(\sqrt{2\log(n)})}_{\eta_n}} \right)$$

Where $\mathcal{E}_2 : t \mapsto 2e^{-t^2/2}$

“Conjecture” : If $\forall r \leq 1 : \; \mathbb{E}[|X_{i,j}|^r] \leq \infty$ Then $\eta_n \leq o(\sqrt{n})$

Application 2: Hanson Wright Theorem

Theorem: **(Hanson Wright)** Given $A \in \mathcal{M}_n$ deterministic, $Z = (z_1, \dots, z_n) \in \mathbb{R}^n$ such that:

- $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz:
  $$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq C' e^{-c' t^2}$$
- $\|\mathbb{E}[Z]\| \leq K$

$C, c, C', c', K > 0$, independent with n

$$\mathbb{P}\left(\left|Z^T A Z - \mathbb{E}[Z^T A Z]\right| \geq t\right) \leq C e^{-\frac{ct^2}{\|A\|_F^2}} + C e^{-\frac{ct}{\|A\|}}$$

$\Phi(Z)$ satisfying: $|\Phi(Z) - \Phi(Z')| \leq \underbrace{(\|AZ\| + \|AZ'\|)}_{\Lambda(Z):\text{variations of } \Phi} \|Z - Z'\|$

Adamczak, Radosław (2014) *A note on the Hanson-Wright inequality for random vectors with dependencies.* Electronic Communications in Probability. 20. 10.1214/ECP.v20-3829.

Application 2: Hanson Wright Theorem

Theorem:
- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t) \qquad (Z, Z' \text{ i.i.d.})$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(\Lambda(Z) \geq t\right) \leq \beta(t)$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ s.t.:

$$\|\Phi(Z) - \Phi(Z')\| \leq \max(\Lambda(Z), \Lambda(Z'))\|Z - Z'\| \quad a.s.$$

Then: $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz, $\forall t > 0$:

$$\forall t > 0 : \quad \mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq 2\, \alpha \boxtimes \beta(t)$$

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq 2e^{-t^2/2}$$

$$\implies \mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq Ce^{-ct^2}$$

$$\implies \mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq C'e^{-c't}$$

For $C, C', c', c > 0$ numerical constant.

Question: Possible to replace $\begin{cases} f(Z') \\ f(\Phi(Z')) \end{cases}$ with $\begin{cases} \mathbb{E}[f(Z)] \\ \mathbb{E}[f(\Phi(Z))] \end{cases}$ **??**

Yes, **IF** $\alpha, \beta : t \mapsto 2e^{-t^2/2}$

Other choices for α, β ??

Application 2: Hanson Wright Theorem

Convex concentration setting (Talagrand's Theorem)

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(|\Lambda(Z) - \mathbb{E}[\Lambda(Z)]| \geq t\right) \leq \alpha\left(\frac{t}{\lambda}\right)$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ convex s.t.:

$$\|\Phi(Z) - \Phi(Z')\| \leq \max(\Lambda(Z), \Lambda(Z'))\|Z - Z'\| \quad a.s.$$

- Assume α independent with n and:

$$\sigma_\alpha \equiv \sqrt{\int_{\mathbb{R}_+} t\alpha(t)dt} \leq \infty \quad (\mathbb{E}[|f(Z) - \mathbb{E}[f(Z)]|^2] \leq \sigma_\alpha^2)$$

Then:

$$\forall t > 0 : \quad \mathbb{P}\left(|\Phi(Z) - \mathbb{E}[\Phi(Z)]| \geq t\right) \leq 2\ \alpha\left(\frac{t}{|\mathbb{E}[\Lambda(Z)]|}\right) + 2\ \alpha\left(\sqrt{\frac{t}{\lambda}}\right).$$

Lemma:

$$\alpha \boxtimes \alpha \circ \min\left(\mathrm{inc}_{\mathbb{E}[\Lambda(Z)]}, \frac{\mathrm{Id}}{\lambda}\right) = \alpha \circ \min\left(\frac{\mathrm{Id}}{\mathbb{E}[\Lambda(Z)]}, \sqrt{\frac{\mathrm{Id}}{\lambda}}\right)$$

Application 2: Hanson Wright Theorem

Convex concentration setting (Talagrand's Theorem)

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(|\|AZ\| - \mathbb{E}[\|AZ\|]| \geq t\right) \leq \alpha\left(\frac{t}{\|A\|}\right)$$

Then: $\forall A \in \mathcal{M}_n$:

$$\forall t \geq 0 : \quad \mathbb{P}\left(\left|Z^T AZ - \mathbb{E}[Z^T AZ]\right| \geq t\right) \leq 2\,\alpha\left(\frac{t}{\mathbb{E}[\|AZ\|]}\right) + 2\,\alpha\left(\sqrt{\frac{t}{\|A\|}}\right).$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ convex s.t.:

$$\left\|Z^T AZ - Z'^T AZ'\right\| \leq 2\max(\underbrace{\|AZ\|}_{\Lambda(Z)}, \|AZ'\|)\|Z - Z'\| \quad a.s$$

- Assume α independent with n and:

$$\sigma_\alpha \equiv \sqrt{\int_{\mathbb{R}_+} t\alpha(t)dt} \leq \infty$$

Application 2: Hanson Wright Theorem

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$$

with $\alpha : \mathbb{R}^+ \to \mathbb{R}_+$ independent with n.

- $\sigma_\alpha \equiv \sqrt{\int_{\mathbb{R}_+} t\alpha(t)dt} \leq \infty$

Then: $\forall A \in \mathcal{M}_n$, $\forall t > 0$:

$$\mathbb{P}\left(\left|Z^T AZ - \mathbb{E}[Z^T AZ]\right| \geq t\right) \leq C\ \alpha\left(\frac{ct}{\|\Sigma\| \|A\|_F}\right) + C\alpha\left(\sqrt{\frac{ct}{\|A\|}}\right).$$

$$\begin{aligned}\mathbb{E}[\|AZ\|] &\leq \sqrt{\mathbb{E}[\|AZ\|^2]}\\ &= \sqrt{\mathbb{E}[\mathrm{Tr}(A^T AZZ^T)]}\\ &= \|A\|_F\sqrt{\|\mathbb{E}[ZZ^T]\|}\\ &= \|A\|_F\sqrt{\|\Sigma\|}\end{aligned}$$

Comparison *Adamczak's result:* $\alpha : t \mapsto e^{-\frac{t^2}{2{\sigma_\alpha}^2}}$

Application 3: Random matrix concentration

Given $x_1, \dots, x_n \sim \mathcal{N}(0, \Sigma)$, i.i.d. random vectors, note $X \equiv (x_1, \dots, x_n) \in \mathbb{R}^{n \times p}$.

Goal: Eigen value distribution of $\frac{1}{n} X X^T$: $\mu \equiv \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}$ **??**

Eigen values of $\frac{1}{p} X X^T$

$\left(\text{Sp} \left(\frac{1}{p} X X^T \right) = \{ \lambda_1, \dots, \lambda_p \} \right)$

- Correspondance $\mu \longleftrightarrow m : z \mapsto \int_{\mathbb{R}} \frac{1}{z - \lambda} d\mu(\lambda)$

"Steiltjes Transform" (similar to Cauchy Transform)

- Link with the *"Resolvent"*: $m(z) = \frac{1}{p} \text{Tr} Q(z)$, where $Q(z) \equiv \left(z I_p - \frac{1}{n} X X^T \right)^{-1}$.

Strategy: Find deterministic $\tilde{Q} \in \mathcal{M}_p$ such that $Q \approx \tilde{Q}$

Application 3: Random matrix concentration

1- Concentration of $Q = (zI_n - \frac{1}{n}XX^T)^{-1}$

$\alpha : t \mapsto \sup \{\mathbb{P}(|f(X) - f(X')| \geq t), f \in \mathcal{L}_1(\mathcal{M}_p, \mathbb{R})\}$

$M \mapsto (zI_p - \frac{1}{n}MM^T)^{-1}$ is $\frac{C}{\sqrt{n}}$-Lipschitz

$\rightarrow \forall f \in \mathcal{L}_1(\mathcal{M}_p, \mathbb{R})$:

$$\mathbb{P}\left(|f(Q) - \mathbb{E}[f(Q)]| \geq t\right) \leq \alpha(\sqrt{n}t/C)$$

2- Find deterministic computable $\tilde{Q}$ close to $\mathbb{E}[Q]$.

Will deduce: $\forall A \in \mathcal{M}_p$ deterministic:

$$\mathbb{P}\left(|\mathrm{Tr}(A(Q - \tilde{Q}))| \geq t\right) \leq C\alpha(?)$$

Application 3: Random matrix concentration

Goal: Approach $\mathbb{E}[Q] = \mathbb{E}\left[\left(zI_p - \frac{1}{n}XX^T\right)^{-1}\right]$

- Of course $\mathbb{E}[Q]$ far from $(zI_p - \Sigma)^{-1}$ $\qquad \Sigma \equiv \frac{1}{n}\sum_{i=1}^n \Sigma_i$ where $\Sigma_i = \mathbb{E}\left[\frac{1}{n}x_i x_i^T\right]$, $\forall i \in [n]$

Solution: Look for $\tilde{Q} \equiv \left(zI_p - \Sigma^\Delta\right)^{-1}$ $\qquad \Sigma^\Delta \equiv \frac{1}{n}\sum_{i=1}^n \Delta_i \Sigma_i$, Δ to be determined

Given $A \in \mathcal{M}_p$, deterministic:

Dependence between Q and x_i

$$\mathrm{Tr}\left(A(\mathbb{E}[Q] - \tilde{Q})\right) = \mathbb{E}\left[\mathrm{Tr}\left(AQ\left(\Sigma^\Delta - \frac{1}{n}XX^T\right)\tilde{Q}\right)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathrm{Tr}\left(\Delta_i AQ\Sigma_i\tilde{Q} - AQx_i x_i^T\tilde{Q}\right)\right]$$

Application 3: Random matrix concentration

$$\mathrm{Tr}\left(A(\mathbb{E}[Q]-\tilde{Q}_\delta)\right)=\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[\mathrm{Tr}\left(\left(\Delta_i-\frac{1}{1+\frac{1}{n}x_i^TQ_{-i}x_i}\right)AQ_{-i}x_ix_i^T\tilde{Q}^\Delta\right)\right]+O\left(\frac{1}{\sqrt{n}}\right)$$

Use the *Schur Formula:* $Qx_i=\dfrac{Q_{-i}x_i}{1+\frac{1}{n}x_i^TQ_{-i}x_i}$, with $Q_{-i}\equiv\left(zI_p-\frac{1}{n}XX^T-x_ix_i^T\right)^{-1}$.

Independent with x_i

1. Chose $\Delta_i^{(1)}\equiv\mathbb{E}\left[\frac{1}{1-\frac{1}{n}x_i^TQ_{-i}x_i}\right]\approx\frac{1}{1-\frac{1}{n}\mathrm{Tr}(\Sigma\tilde{Q}^{\Delta^{(1)}})}$ independent with p,n.

   Relies on **Hanson-Wright Inequality:**

   $$\mathbb{P}\left(\left|x_i^T\tilde{Q}^\Delta AQ_{-i}x_i-\mathbb{E}[x_i^T\tilde{Q}^\Delta AQ_{-i}x_i]\right|\geq t\right)\leq Ce^{-c(t/\|A\|_{HS})^2}+Ce^{-ct/\|A\|}$$

2. Chose $\Delta^{(2)}$ solution to $\Delta_i^{(2)}=\frac{1}{1-\frac{1}{n}\mathrm{Tr}(\Sigma\tilde{Q}^{\Delta^{(2)}})}$

Application 3: Random matrix concentration

Recall the objects: $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$

$Q = \left(zI_p - \frac{1}{n}XX^T\right)^{-1}$ $\quad\quad$ $\tilde{Q} \equiv \left(zI_p - \Sigma^\Delta\right)^{-1}$ $\quad$ $\Sigma^\Delta \equiv \frac{1}{n}\sum_{i=1}^n \Delta_i \Sigma_i$

With Δ solution to $\Delta_i^{(2)} = \frac{1}{1 - \frac{1}{n}\mathrm{Tr}(\Sigma \tilde{Q}^\Delta)}$

$\dot{\alpha} : t \mapsto \sup \left\{\mathbb{P}(|f(x_i) - f(x_i')| \geq t), i \in [n], f \in \mathcal{L}_1(\mathbb{R}^p, \mathbb{R})\right\}$
$\alpha : t \mapsto \sup \left\{\mathbb{P}(|f(X) - f(X')| \geq t), f \in \mathcal{L}_1(\mathcal{M}_p, \mathbb{R})\right\}$

Theorem: Assume:

- $x_1, \dots, x_n$ independents
- $\|\Sigma_i\| \leq C$
- $\forall t > 0 : \alpha(\sqrt{n}t) \leq \dot{\alpha}(t)$

Then: For all $f : \mathcal{M}_p \to \mathbb{R}$:

$$\|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_{HS} \leq C\frac{\dot{\tau}_4}{\sqrt{n}} \qquad \|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_* \leq C\sqrt{p}\dot{\tau}_1 + \dot{\tau}_2,$$

where: $\dot{\tau}_1 = \int \dot{\alpha}(t)dt$, $\dot{\tau}_2 = 2\int t\dot{\alpha}(t)dt$ and $\dot{\tau}_3 = 3\int t^2\dot{\alpha}(t)dt$

$o(\sqrt{p})$ $\quad$ $o(p)$ $\quad$ $o(p^{\frac{1}{3}})$ $\quad$ (Heavy tailed concentration)

Application 3: Random matrix concentration

Theorem:

Assume:
- $x_1, \ldots, x_n$ independents
- $\|\Sigma_i\| \leq C$
- $\forall t > 0 : \alpha(\sqrt{n}t) \leq \dot{\alpha}(t)$

Then: $\|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_{HS} \leq C \frac{\dot{\tau}_4}{\sqrt{n}}$ $\quad \|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_* \leq C\sqrt{p}\dot{\tau}_1 + C\dot{\tau}_2,$

where: $\dot{\tau}_1 = \int \dot{\alpha}(t)dt$, $\dot{\tau}_2 = \int t\dot{\alpha}(t)dt$ and $\dot{\tau}_4 = \int t^3\dot{\alpha}(t)dt$

- Stieltjes transform $m(z) = \frac{1}{p}\mathrm{Tr}(Q)$ in heavy tailed setting:

$$\mathbb{P}\left(\left|\frac{1}{p}\mathrm{Tr}(Q) - \frac{1}{p}\mathrm{Tr}(\tilde{Q}^\Delta)\right| \geq t\right) \leq \dot{\alpha}\left(\left(\frac{p}{\dot{\tau}_2} + \frac{\sqrt{p}}{\dot{\tau}_1}\right)t\right) \qquad \text{Because } \left\|\frac{1}{p}I_p\right\| = \frac{1}{p}$$

"Conjecture" : If $\exists \phi \in \mathcal{L}_1(\mathbb{R})$: $X_{i,j} = \Phi(Z_{i,j}, Z \sim \mathcal{N}(0, I_{pn})$ and $\forall r \leq 1$: $\mathbb{E}[|X_{i,j}|^r] \leq \infty$
Then $\dot{\tau}_1 \leq o(\sqrt{p})$ and $\dot{\tau}_2 \leq o(p)$

- In machine learning:

$$\mathbb{P}\left(\left|\frac{1}{n}Y^TQ^2Y - \frac{1}{n}Y^T\tilde{Q}_2^\Delta Y\right| \geq t\right) \leq \dot{\alpha}\left(\frac{\sqrt{nt}}{\dot{\tau}_4}t\right) \qquad \text{Because } \left\|\frac{1}{n}YY^T\right\|_{HS} \leq C$$

Application 1: Heavy tailed concentration

Proposition:
Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$, $\phi_i \in \mathcal{L}_1(\mathbb{R})$ and $\exists h : \mathbb{R}_+ \to \mathbb{R}_+$, increasing s.t.:

- $\forall i \in [n] : \; X_i = \phi_i(Z_i)$
- $\forall t \in \mathbb{R}$, $\forall i \in [n]$: $|\phi_i'(t)| \leq h(|t|)$
- for all $a > 2\log(2n)$, $b > 0$:
$$h(\sqrt{a+b}) \leq h(\sqrt{a})h(\sqrt{b}).$$

Then $\forall f \in \mathcal{L}_1(\mathbb{R}^n, \mathbb{R})$:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq 3\mathcal{E}_2 \circ (\mathrm{Id} \cdot h)^{-1} \circ \left(\frac{t}{h(\sqrt{2\log(n)})} \right)$$

Where $\mathcal{E}_2 : t \mapsto 2e^{-t^2/2}$

Example: Consider the case $\phi_i|_{\mathbb{R}_+} : t \mapsto e^{t^2/2q} - 1$, $h = \phi_i|_{\mathbb{R}_+}'$.

Then $\mathbb{E}[X_i^r] \leq \infty \iff r < q$

$\eta_n = \frac{1}{q} n^{\frac{1}{q}} \sqrt{2\log(n)}$, and for $q > 1$, $r < q$: $\quad \mathbb{E}[|f(X) - \mathbb{E}[f(X)]|^r] \leq C{\eta_n}^r$

Then $\mathbb{E}[X_i^2] \leq \infty \iff q > 2 \implies \eta_n \underset{n \to \infty}{\leq} o(\sqrt{n})$