Améliorations de la concentration sur les espaces produits

I/ Entropie et produit

Conventions d'écriture (formel)

Soit X un ensemble produit, $X = X_1 \times \cdots \times X_n$ $(n \geq 1)$

Un élément $x \in X$ s'écrit $x = (x_1, -, x_n)$ avec $x_i \in X_i$

→ Soit $f : X \to \mathbb{R}$

On note f_i la fonction définie sur X_i par

$$f_i(t) = f(x_1, -, x_{i-1}, t, x_{i+1}, -, x_n)$$

les $x_1, -, x_{i-1}, x_{i+1}, -, x_i$ étant fixés.

C'est à dire qu'on s'intéresse à $f(x) = f_i(x_i)$ comme fonction de x_i.

Il peut être utile de noter, pour $i \leq n$, $\hat{X}_i = \prod_{j \neq i} X_j$

(de sorte qu'à l'ordre près, on a "$X = X_i \times_i \hat{X}_i$")

De même, pour $x \in X$, $\hat{x}_i = (x_1, -, x_{i-1}, x_{i+1}, -, x_n) \in \hat{X}_i$

Ainsi la fonction $x_i \to f_i(x_i)$ est définie pour $\hat{x}_i \in \hat{X}_i$ donné.
[$\hat{x}_i = \hat{y}_i \Leftrightarrow x$ et y ne diffèrent (éventuellement) que de la $i^{\text{ème}}$ coordonnée]

Supposons maintenant que l'on munisse X d'une <u>proba</u> produit $\mu = \mu_1 \otimes - \otimes \mu_n$ avec

μ_i proba sur X_i.

On note $\widehat{\mu_i} = \mu_1 \otimes - \otimes \mu_{i-1} \otimes \mu_{i+1} \cdots \otimes \mu_n$

↳ proba sur $\hat{X}_i$

Pour un élément de $\hat{X}_i$ fixé (notons le $\hat{x}_i$) on a

$$\int_{X_i} f_i(t)\, d\mu_i(t) = \int_{X_i} f(x)\, d\mu_i(x)$$

i.e. on intègre $f(x)$ par rapport à la variable x_i et la mesure $d\mu_i(x_i)$

Par Fubini (structure produit) on a :

$$\int_{\hat{X}_i} \left(\int_{X_i} f_i\, d\mu_i \right) d\hat{\mu}_i = \int_X f\, d\mu$$

On a mieux, car on travaille avec des proba. (espérance conditionnelle)
En effet, la fonction

$$\begin{aligned} \hat{X}_i &\longrightarrow \mathbb{R} \\ \hat{x}_i &\longrightarrow \int_{X_i} f_i(x_i)\, d\mu_i(x_i) \end{aligned}$$

peut être vue comme une fonction sur X ne dépend que de $\hat{x}_i$

$$\begin{aligned} X &\longrightarrow \mathbb{R} \\ x &\longrightarrow \int_{X_i} f_i(x_i)\, d\mu(x_i) \end{aligned}$$

fct de $\hat{x}_i$ seulement

et alors on a :

$$\boxed{\int_X \left(\int_{X_i} f_i\, d\mu_i \right) = \int_{\hat{X}_i} \left(\int_{X_i} f_i\, d\mu_i \right) d\hat{\mu}_i = \int_X f\, d\mu}$$

Entropie

Soit (Ω, μ) un espace de proba et $f : \Omega \longrightarrow \mathbb{R}^+$. On définit

$$\mathrm{Ent}_\mu(f) := \int f \log f \, d\mu - \int f \, d\mu \times \log \int f \, d\mu$$

i.e.

$$\boxed{\mathrm{Ent}_\mu(f) = \int f \log\left(\frac{f}{\int f \, d\mu}\right) d\mu} \qquad = \int f \log f \, d\mu \quad \text{si } \int f \, d\mu = 1$$

Notez que pour $F : \Omega \longrightarrow \mathbb{R}^+$, on a :

$$\mathrm{Ent}_\mu(e^F) = \int F e^F \, d\mu - \int e^F \, d\mu \log \int e^F \, d\mu$$

Soit comme précédement X un espace de proba produit

$$X = X_1 \times \dots \times X_n$$
$$\mu = \mu_1 \otimes \dots \otimes \mu_n, \qquad \mu_i \text{ proba sur } X_i$$

Soit $f : X \longrightarrow \mathbb{R}$. Pour tout $\hat{x}_i \in \hat{X}_i$ fixé, on peut définir $\mathrm{Ent}_{\mu_i}(f_i)$. Plus généralement, pour $x \in X$,

$$\begin{aligned}
\mathrm{Ent}_{\mu_i}(f_i)(x) &= \mathrm{Ent}_{\mu_i}(f_i)(\hat{x}_i) \\
&= \int_{X_i} f_i(t) \log f_i(t) \, d\mu_i(t) - \int_{X_i} f_i(t) \, d\mu_i(t) \log \int_{X_i} f_i(t) \, d\mu_i(t) \\
&= \int_{X_i} f(x) \log f(x) \, d\mu_i(x_i) - \int_{X_i} f(x) \, d\mu_i(x_i) \log \int_{X_i} f(x) \, d\mu_i(x_i)
\end{aligned}$$

Histoire de notations, uniquement

et avec ces notations

$$\int_X \mathrm{Ent}_{\mu_i}(f_i)\, d\mu = \int_{\hat{X}_i} \mathrm{Ent}_{\mu_i}(f_i)\, d\hat{\mu}_i$$

<u>Theo A</u> (Tensorisation de l'entropie)

Soit (X,μ) le produit des espaces de proba (X_i,μ_i), $i=1,\dots,n$.
Pour $f : X \longrightarrow \mathbb{R}^+$ on a

$$\mathrm{Ent}_\mu(f) \leq \sum_{i=1}^n \int_X \mathrm{Ent}_{\mu_i}(f_i)\, d\mu$$

à comprendre

<u>Démo</u>. On raisonne par récurrence. Il suffit de montrer pour $n=2$

$\mu = \mu_1 \otimes \mu_2$
$X = X_1 \times X_2$

$d\mu(x_1,x_2) = d\mu_1(x_1)\, d\mu_2(x_2)$

Soit $f : X_1 \times X_2 \longrightarrow \mathbb{R}^+$

Ecrivons, pour $(x_2, x_2) \in X_2$ $\quad \hat{x}_1 = x_2$

$$f_1(x_1) := f^{x_2}(x_1) := f(x_1,x_2)$$

On voit donc f^{x_2} comme une fonction sur X_1.

$$\mathrm{Ent}_\mu(f) = \int_X f \log f\, d\mu - \int_X f\, d\mu \cdot \log \int_X f\, d\mu$$

$$= \int_{X_2}\int_{X_1} f \log f\, d\mu_1(x_1)\, d\mu_2(x_2) - \int_{X_2} \left(\int_{X_1} f\, d\mu_1 \cdot \log \int_X f\, d\mu \right) d\mu_2$$

On tente de faire apparaître $\mathrm{Ent}_{\mu_1}(f)$:

$$\mathrm{Ent}_\mu(f) = \int_{X_2}\left(\int_{X_1} f^{x_2}(x_1) \log \frac{f^{x_2}(x_1)}{\int_{X_1} f^{x_2}(x_1)d\mu_1(x_1)} d\mu_1(x_1)\right) d\mu_2(x_2)$$

$$+ \int_{X_2}\left(\int_{X_1} f^{x_2} d\mu_1\right) \log\left(\frac{\int_{X_1} f^{x_2} d\mu_1}{\int_{X_2}\left(\int_{X_1} f^{x_2} d\mu_1\right)}\right) d\mu_2(x_2)$$

$$= \int_{X_2} \mathrm{Ent}_{\mu_1}(f^{x_2})\, d\mu_2(x_2) + \mathrm{Ent}_{\mu_2}\left(\int_{X_1} f\, d\mu_1\right)$$

On conclut en remarquant que l'entropie est une fonctionnelle convexe, ie. $\forall \mu$ proba

$f \longrightarrow \mathrm{Ent}_\mu(f)$ est convexe sur l'ensemble des fcts ≥ 0

($\downarrow$ voir + loin)

On a donc :

$$\mathrm{Ent}_{\mu_2}\left(\int_{X_1} f\, d\mu_1\right) \leq \int_{X_1} \mathrm{Ent}_{\mu_2}(f)\, d\mu_1$$

On a donc bien

$$\mathrm{Ent}_\mu(f) \leq \int_{X_2} \mathrm{Ent}_{\mu_1}(f_1)\, d\mu_2 + \int_{X_1} \mathrm{Ent}_{\mu_2}(f_2)\, d\mu_1$$

$$= \int_X \mathrm{Ent}_{\mu_1}(f_1)\, d\mu + \int_X \mathrm{Ent}_{\mu_2}(f_2)\, d\mu.$$

$\blacksquare$

Pour la convexité de la fonctionnelle entropie, traitez l'

Exercice 5.1

Soit (Ω, μ) un espace de proba et $f : \Omega \to \mathbb{R}^+$

Montrer que

$$\mathrm{Ent}_\mu(f) = \sup \left\{ \int f g \, d\mu \; ; \; g : \Omega \to \mathbb{R} \text{ avec } \int e^g \, d\mu \leq 1 \right\}$$

En déduire la convexité de l'entropie.

La tensorisation de l'entropie est très utile. En voici une première conséquence, utile également.

Théo B

Soit (X, μ) le produit des espace de proba (X_i, μ_i), $i = 1, \dots, n$.

Pour $F : X \to \mathbb{R}$, on a :

$$\mathrm{Ent}_\mu(e^F) \leq \sum_{i=1}^{n} \int \mathcal{R}_i(e^{F_i}) \, d\mu$$

où, pour $i = 1, \dots, n$ et $x \in X$

$$\mathcal{R}_i(e^{F_i})(x) = \mathcal{R}_i(e^{F_i})(\hat{x}_i)$$

$$= \iint_{\{F_i(x_i) \geq F_i(y_i)\}} \left[F_i(x_i) - F_i(y_i) \right]^2 e^{F_i(x_i)} \, d\mu_i(x_i) \, d\mu_i(y_i)$$

$d\mu_i(x_i) \, d\mu_i(y_i)$

Rq : ⚠ $F_i(y_i) = F(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)$

Rq : Evidement $\mathcal{R}_i(e^{F_i})(x) \leq \iint_{X_i \times X_i} \left[F_i(x_i) - F_i(y_i) \right]^2 e^{F_i(x_i)} \, d\mu_i(x_i) \, d\mu_i(y_i)$

Rq :

Démo

Si on a montré l'inégalité pour $n=1$ (un seul espace), alors l'inégalité sur le produit découle immédiatement du theo A

Notons $X = X_1 \quad \mu = \mu_1$

On doit montrer que

$$(*) \quad Ent_\mu(e^F) \leq \iint (x-y)^2 \left(\frac{\partial F}{\partial x}\right)^2_{(x)} e^{-F(x)} d\mu(x)\, d\mu(y)$$

Par Jensen on a :

$$Ent_\mu(e^F) = \int F e^F d\mu - \int e^F d\mu \log \int e^F d\mu$$
$$\leq \int F e^F d\mu - \int e^F d\mu \int F d\mu$$

On remarque que

$$\int F e^F d\mu - \int e^F d\mu \int F d\mu = \frac{1}{2} \iint [F(x)-F(y)][e^{F(x)} - e^{F(y)}] d\mu(x) d\mu(y)$$

symétrique en x,y

$$= \frac{1}{2} \iint_{\{F(x) \geq F(y)\}} + \frac{1}{2} \iint_{\{F(y) \geq F(x)\}}$$

identique $(x \leftrightarrow y)$

$$= \iint_{\{F(x) \geq F(y)\}} [F(x)-F(y)][e^{F(x)} - e^{F(y)}] d\mu(x) d\mu(y)$$

On conclut en remarquant que (exo)

Pour $s, t \in \mathbb{R}$ avec $s \geq t$:

$$(s-t)(e^s - e^t) \leq \frac{1}{2}(s-t)^2(e^s + e^t) \leq (s-t)^2 e^s$$

II/ Entropie et concentration sur les espaces produit

Notre but est de montrer une concentration gaussienne pour les fonctions convexes et lip. sur $[0,1]^n$ (comme le résultat obtenu par la propriété (τ) - convexe).

On commence par l'observation suivante

Prop C

Soit P une proba produit sur $\mathbb{R}^n$ et $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction séparément convexe (càd que chaque f_i est convexe sur $\mathbb{R}$, ce qui est moins demander que f convexe, bien sûr)

Alors

$$\mathrm{Ent}_P\left(e^f\right) \leq \iint \sum_{i=1}^n (x_i - y_i)^2 \left(\frac{\partial f}{\partial x_i}(x)\right)^2 e^{f_i(x_i)}$$

Rq : On rappelle qu'une fonction convexe est localement lip et que donc elle admet presque partout un gradient ∇f. Cela dit, comme la mesure P est ici générale, il est prudent de supposer que f est C^1 dans cet énoncé (ce n'est pas gênant pour la suite)

Démo : Ecrivons $P = \mu_1 \otimes \dots \otimes \mu_n$ avec μ_i proba sur $\mathbb{R}$.

D'après le theo B, (et Fubini, encore) il est suffisant de montrer que

$$\mathcal{R}_i(e^{f_i})(x) \leq \iint_{\mathbb{R}\times\mathbb{R}} (x_i-y_i)^2 f_i'(x_i)^2 e^{f_i(x_i)} \, d\mu_i(x_i)\, d\mu_i(y_i)$$

(en effet, l'intégrale de droite intégrée % $d\hat{\mu}_i(\hat{y}_i)$ est la même que celle intégrée % $d\mu(y)$ sur X)

Comme f_i est convexe sur $\mathbb{R}$ on a

$$f_i(x_i) - f_i(y_i) \leq (x_i - y_i) f_i'(x_i)$$

↳ loi constante (indep. de $\hat{y}_i$)

Si $f_i(x_i) \geq f_i(y_i)$ alors on peut prendre les carrés et on obtient

$$\mathcal{R}_i(e^{f_i}) \leq \iint_{\{f_i(x_i)\geq f_i(y_i)\}} (x_i-y_i)^2 f_i'(x_i) e^{f_i(x_i)} \, d\mu_i(x_i)\, d\mu_i(y_i)$$

$$\leq \iint_{X_i\times X_i} \quad \text{''} \quad \text{''} \quad \text{''}$$

$\blacksquare$

Voici une conséquence de cette proposition

theo D

Soit P une proba. produit sur $[a,b]^n$ avec $a,b\in\mathbb{R}$ et $d=|b-a|$.
Pour toute fonction $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ séparément convexe et 1-lip (par rapport à la norme euclidienne) on a : $\forall r \geq 0$,

$$P\left(\left\{F \geq \int F\,dP + r\right\}\right) \leq e^{-r^2/4d^2}$$

En général ce résultat est énoncé sur $[0,1]^n$ (i.e $d=1$)
Par ailleurs, il suffit que F soit définie sur $[a,b]^n$.

Evidemment, d est beaucoup plus petit que le diamètre de $[a,b]^n$ (ici, d est vraiment une constante)

Dans ce théo, on gagne par rapport à la concentration produit car on est en mesure d'utiliser que F est 1-lip % la métrique euclidienne. Le prix à payer pour pouvoir utiliser cette hypothèse forte est qu'on doit imposer F séparément convexe

[Notez que la concentration produit donnerait un terme en n,

$$\forall r \geq 0 \quad P(\{F \geq \int F dP + r\}) \leq e^{-r^2/4d^2 n} \longleftarrow \text{mauvais} \quad]$$

<u>Démo</u> (méthode de Herbst)

OPS que $d = 1$, i.e : on travaille sur $[0,1]^n$

Par ailleurs, il suffit de montrer le théo dans le cas où f est C^1

[En effet, pour f 1-lip générale, on peut convoler avec un noyau Gaussien (g_ε), $f_\varepsilon = f * g_\varepsilon$. On a alors f_ε C^∞ avec $|\nabla f_\varepsilon| \leq 1$. On applique le résultat à f_ε et on passe à la limite par convergence ~~uniforme~~ sur $[0,1]^n$.

On a donc $|\nabla f| \leq 1$ sur $\mathbb{R}^n$ i.e : $\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 (x) \leq 1, \forall x$

Soit $\boxed{H(\lambda) = \int e^{\lambda F} dP}$ On veut majorer $H(\lambda)$.

On applique la Prop C à la fonction sép. convexe λF et on trouve

$$\mathrm{Ent}_P(e^{\lambda F}) \leq \iint \sum \underbrace{(x_i - y_i)^2}_{\substack{\leq 1 \text{ sur le} \\ \text{support de } P \otimes P}} \lambda^2 \left(\frac{\partial F}{\partial x_i}\right)^2 (x)\, e^{\lambda F(x)} dP(x) dP(y)$$

$$\leq \lambda^2 \int e^{\lambda F(x)} dP(x)$$

Par ailleurs

$$H'(\lambda) = \frac{dH}{d\lambda}(\lambda) = \int F e^{\lambda F} dP$$

Donc $\quad Ent_P(e^{\lambda F}) = \lambda H'(\lambda) - H(\lambda) \log H(\lambda)$

On a donc

$$\lambda H'(\lambda) - H(\lambda) \log H(\lambda) \leq \lambda^2 H(\lambda)$$

On conclut à une majoration de H par un argument à la Gronwall.

Pour le voir, posons $\alpha(\lambda) := \frac{1}{\lambda} \log H(\lambda)$ $\qquad \alpha(0) = H'(0)$ pour que α soit continue

On a exactement :

$$\alpha'(\lambda) \leq 1$$

. Donc $\qquad \alpha(\lambda) \leq \lambda + \alpha(0)$

et donc $\qquad H(\lambda) \leq e^{\lambda^2 + \lambda \alpha(0)}$

Par ailleurs : $\qquad H(\lambda) = 1 + \lambda \int F dP + o(\lambda)$

donc $\qquad \alpha(0) = H'(0) = \int F dP$

On a donc

$$\boxed{\int e^{\lambda [F - \int F dP]} dP \leq e^{\lambda^2}} \qquad \underline{\forall \lambda \geq 0}$$

La suite est classique. !

On a :
$\forall r \geq 0, \qquad P(\{F - \int F dP \geq r\}) \leq e^{\lambda^2 - \lambda r}$ $\qquad$ 2 λ optimal $\lambda = \frac{r}{2}$

soit encore : $\qquad \underline{P(\{F \geq \int F dP + r\}) \leq e^{-r^2/4}}$

Applications du résultat

Evidement, dans le cas de $f(x) = x \cdot a = \sum_{i=1}^{n} a_i x_i$

sur (Ω_n, σ_n), cela ne donne rien de mieux.

En effet, on a $\|f\|_{Lip} = |a|$ (euclidien) et donc, puisque f est convexe

on retrouve

$$\mathbb{P}\left(\sum_{i=1}^{n} a_i \varepsilon_i \geq r \right) \leq e^{-r^2/4|a|^2}$$

(Rq : on ne retrouve qu'un des sens de la concentration)

Prenons E un Banach, $(E, \|\cdot\|)$ de dual $(E^*, \|\cdot\|_*)$

Soit $v_1, \ldots, v_k \in E$ et $\varepsilon_1, \ldots, \varepsilon_k$ des

v.a. de Bernoulli ± 1 indépendantes.

Posons

$$S = \sum_{i=1}^{k} \varepsilon_i v_i \quad \text{(vecteur aléatoire de } E \text{)}$$

que peut-on dire de $\|S\|$?

1) Si on utilise la concentration produit sur (Ω_k, σ_k)

Pour $f(x) = \| \sum x_i v_i \|$

On a : si x et y varient de la coord. i seulement

$$|f(x) - f(y)| \leq 2 \|v_i\|$$

Donc :

$$\mathbb{P}(\|S\| \geq \mathbb{E}\|S\| + r) \leq e^{-r^2/8K^2}$$

$$\text{où } K^2 = \sum_{i=1}^{n} \|v_i\|^2$$

2) Si on utilise le <u>theo D</u> avec le cas

$$f(x) = \| \sum_{i=1}^{k} x_i v_i \|$$

On remarque que f est <u>convexe</u> sur $\mathbb{R}^n$

quelle est sa constante de lip ?

$$f(x) = \sup_{\substack{\xi \in E^* \\ \|\xi\|_* \leq 1}} \sum_{i=1}^{n} x_i \, \xi(v_i)$$

On voit que $\|f\|_{lip} \leq \sigma$

où $\sigma := \sup_{\|\xi\|_* \leq 1} \sqrt{\sum_{i=1}^{k} \xi(v_i)^2}$

~~Ou~~ On a alors

$$\mathbb{P}(\|S\| \geq \mathbb{E}\|S\| + r) \leq e^{-r^2/4\sigma^2}$$

<u>Or</u> : $\sigma^2 \leq \sum_{i=1}^{k} \|v_i\|^2$

en général, il y a une grande différence.

[Revoir concentration Gaussienne version Pisier]

Donc le <u>theo D</u> donne la bonne concentration gaussienne pour la norme d'une somme <u>vectorielle</u> de v.a. indep.