Transport optimal entropique

Armand LEY

Existence et unicité pour le problème de Monge-Kantorovitch :

Proposition :

L'ensemble $U(a, b)$ est un convexe compact, qui contient $a \otimes b := (a_i b_j)_{i,j}$

Existence :

Le problème de Monge-Kantorovitch admet toujours une solution.

Unicité :

Pas d'unicité en général.

Entropie et entropie relative

Définitions :

$\mathrm{Ent}(p) = -\sum_{x \in E} p_x \log(p_x)$.

$\mathrm{Ent}(p|q) = \sum_{x \in E} p_x \log\left(\frac{p_x}{q_x}\right)$.

Propriétés de l'entropie

Propriétés :

1. Si u est la probabilité uniforme sur E, alors
$$\mathrm{Ent}(p|u) = \log(|E|) - \mathrm{Ent}(p).$$
2. Pour tout couple $(p, q) \in \mathcal{P}(E)^2$,
$$\mathrm{Ent}(p|q) \geq 0,$$
avec égalité si et seulement si $p = q$.
3. Pour tout $q \in \mathcal{P}(E)$, l'application
$$\mathrm{Ent}(\cdot|q)$$
est strictement convexe.

Proposition

L'entropie n'est pas une distance sur $\mathcal{P}(E)$.

Contre-exemples :

1. Si $|E| = 2$, $p = (1/2, 1/2)$ et $q = \delta_1$, alors $\mathrm{Ent}(p|q) = +\infty$.
2. Si $|E| = 3$, $p = (1/2, 1/4, 1/4)$ et $q = (1/3, 1/3, 1/3)$, on a

$$6(\mathrm{Ent}(p|q) - \mathrm{Ent}(q|p)) = \log(3/2) + \log(4/3) > 0.$$

3. Si on prend $p_1 = p$, $p_3 = q$ et $p_2 = (p_1 + p_3)/2$, on obtient $\mathrm{Ent}(p_1|p_3) \approx 0.06$ et $\mathrm{Ent}(p_1|p_2) + \mathrm{Ent}(p_2|p_3) \approx 0.03$

Proposition

L'entropie n'est pas une distance sur $\mathcal{P}(E)$.

Contre-exemples :

1. Si $|E| = 2$, $p = (1/2, 1/2)$ et $q = \delta_1$, alors $\mathrm{Ent}(p|q) = +\infty$.
2. Si $|E| = 3$, $p = (1/2, 1/4, 1/4)$ et $q = (1/3, 1/3, 1/3)$, on a

$$6(\mathrm{Ent}(p|q) - \mathrm{Ent}(q|p)) = \log(3/2) + \log(4/3) > 0.$$

3. Si on prend $p_1 = p$, $p_3 = q$ et $p_2 = (p_1 + p_3)/2$, on obtient $\mathrm{Ent}(p_1|p_3) \approx 0.06$ et $\mathrm{Ent}(p_1|p_2) + \mathrm{Ent}(p_2|p_3) \approx 0.03$

Problème de transport pénalisé

But:

Minimiser $J_\varepsilon : P \in U(a, b) \mapsto J(P) + \varepsilon \mathrm{Ent}(P|a \otimes b)$.

Existence et unicité pour le problème pénalisé

Proposition :

Pour tout $P \in U(a,b)$, $\mathrm{Ent}(P|a \otimes b) = \mathrm{Ent}(a) + \mathrm{Ent}(b) - \mathrm{Ent}(P)$.

Théorème :

L'application J_ε admet un unique minimiseur, qu'on note P_ε.

Comportement quand $\varepsilon \to 0$

Notation

- $\mathcal{O}(a, b)$ désigne l'ensemble des plans de transport optimaux.
- On note $W_\varepsilon(a, b) := \min_{P \in U(a,b)} J_\varepsilon(P)$ et $W(a, b) := W_0(a, b)$.

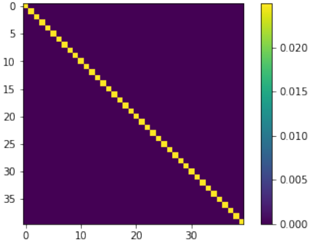
Théorème

1. La suite $(P_\varepsilon)_{\varepsilon > 0}$ converge vers l'unique minimiseur de $\mathrm{Ent}(\cdot | a \otimes b)$ sur $\mathcal{O}(a, b)$.
2. On a
$$W_\varepsilon(a, b) \xrightarrow[\varepsilon \to 0]{} W(a, b).$$

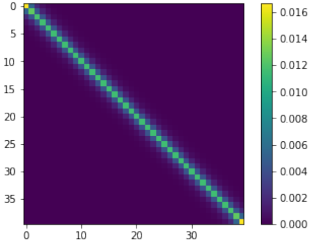
Comportement quand $\varepsilon \to +\infty$

Proposition :

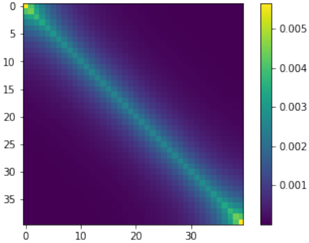
1. $P_\varepsilon \xrightarrow[\varepsilon\to+\infty]{} a \otimes b$
2. $W_\varepsilon(a, b) \xrightarrow[\varepsilon\to+\infty]{} J(a \otimes b)$.



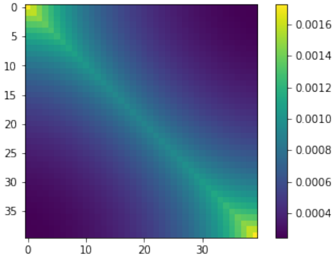
(a) eps = 0.1



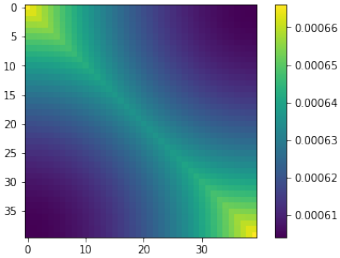
(b) eps =1



(a) eps = 5



(a) eps = 20



(b) eps = 400

Figure: Solutions du problème pénalisé

L'algorithme de Sinkhorn :

Définition du processus :

1. On normalise la matrice suivant les lignes de manière à obtenir a comme somme de ligne.
2. On normalise la matrice suivant les colonnes de manière à obtenir b comme somme de colonne.

Le théorème de Sinkhorn :

Énoncé :

Supposons que K soit à coefficients strictement positifs.

1. Il existe un couple $(u, v) \in {\mathbb{R}_+^*}^m \times {\mathbb{R}_+^*}^n$ tel que $\mathrm{Diag}(u) K \mathrm{Diag}(v) \in U(a, b)$.
2. Ce couple (u, v) est unique à une constante multiplicative près.
3. La suite $(K_l)_{l \geq 0}$ converge vers $\mathrm{Diag}(u) K \mathrm{Diag}(v)$.

Sinkhorn et problème de transport entropique

Notations

- $K^\varepsilon := (e^{-\frac{1}{\varepsilon}C_{i,j}})_{i,j}$.
- $(K_l^\varepsilon)_{l\geq 0}$ la suite de matrices fournie par l'algorithme de Sinkhorn

Théorème

1. On a $P_\varepsilon = \lim\limits_{l\to+\infty} K_l^\varepsilon$.
2. Une vitesse de convergence est donnée par :

$$\|\log(P_\varepsilon) - \log(K_l^\varepsilon)\|_\infty = O(\lambda(K^\varepsilon)^l)$$

Merci pour votre attention !