Transport optimal régularisé pour l'estimation des poids dans des modèles de mélange, et applications à la cytométrie

Soutenance de thèse

Paul Freulon,

Directeur : Jérémie Bigot
Coencadrant : Boris Hejblum
Collaborateur scientifique : Arthur Leclaire

28 mars 2023

Université de Bordeaux

Données de cytométrie

Cytométrie en flux : Mesures caractérisant différents paramètres sur des cellules dans un échantillon biologique

Principe

- Mesure de d caractéristiques biologiques sur chaque cellule
  Exemple : protéines exprimées à la surface de la cellule
- Une observation $X_i \in \mathbb{R}^d \to$ valeurs mesurées sur une cellule

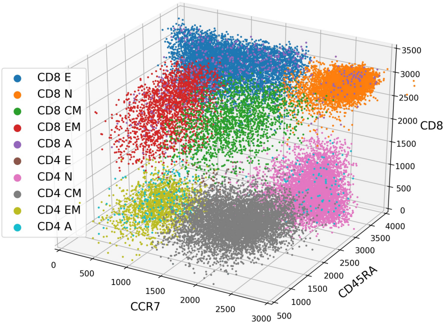
CCR7	CD4	CD45RA	CD3	HLADR	CD38	CD8
717.3	1146.5	3094.8	2526.3	1333.1	1510.2	3203.7

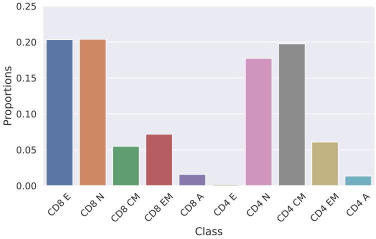
Une observation $X_i \in \mathbb{R}^d$ correspondant à une cellule biologique

En pratique : mesures sur n cellules $\longrightarrow$ n observations $X_1, \dots, X_n$

Données disponibles : $n \in [10^5, 10^7]$ **et** $d \in \{5, 30\}$

Quantité d'intérêt : les proportions par classe



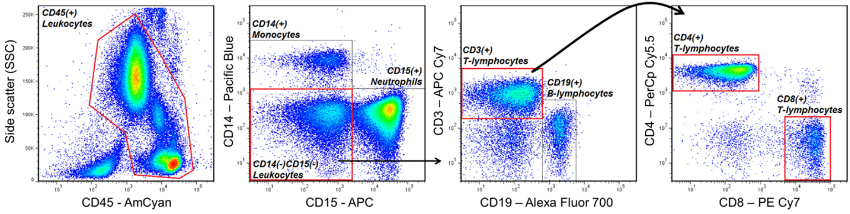


Gauche : projection selon 3 bio-marqueurs et classification en $K = 10$ sous-populations. Droite : proportions relatives des 10 sous-populations.

- $X_1, \ldots, X_n$ se répartissent en K sous-populations $C_1, \ldots, C_K$
- Dans cette présentation : **on estime les proportions**
- Application en immunologie

Analyse manuelle des données

Classification des cellules en sous-populations en fonction des paramètres mesurés



Exemple de traitement manuel (Verschoor et al. 2015)

- Suite de projections selon 2 paramètres biologiques
- Identification des régions à forte densité
- Annotation $+/-$ d'un sous groupe en fonction de l'intensité du signal associé au paramètre biologique à chaque étape
- **Résultat** : classification en K sous-populations $C_1, \ldots, C_K$

Estimation supervisée

Classification $X_1, \ldots, X_n$ en K classes $C_1, \ldots, C_K$ connue

- K connu
- Mesure empirique de C_k

$$\hat{\mu}_k := \frac{1}{n_k} \sum_{i : X_i \in C_k} \delta_{X_i}, \quad \text{où} \quad n_k = \# C_k$$

- $\to \hat{\mu}_1, \ldots, \hat{\mu}_K$ connues

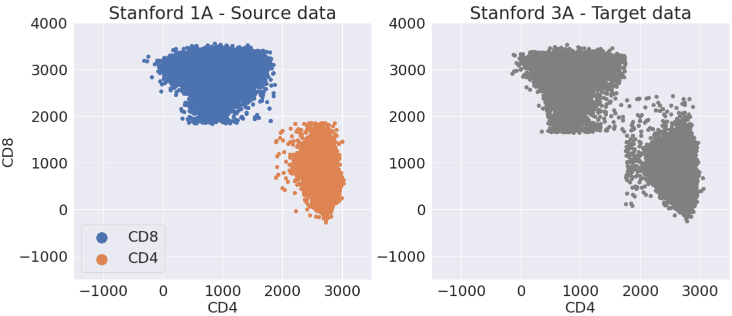
Modèle de mélange avec composantes $\hat{\mu}_1, \ldots, \hat{\mu}_K$

$$\widehat{\mathcal{M}}_K := \left\{ \hat{\mu}_n(\theta) := \sum_{k=1}^{K} \theta_k \hat{\mu}_k \mid \theta \in \Sigma_K \right\}$$

- Paramètres : poids $\theta \in \Sigma_K = \{\theta \in \mathbb{R}_+^K \mid \sum_{k=1}^K \theta_k = 1\}$
- **Absence d'hypothèse** sur la loi de $X_1, \ldots, X_n$

Observations non-classifiées : $Y_1, \ldots, Y_m$

Objectif : proportions dans un jeu de données non classifiées



- Classification : $K = 2 \rightarrow \hat{\mu}_1,\ \hat{\mu}_2,\ \theta = (\theta_1, 1-\theta_1) \in \Sigma_2$

$$\hat{\mu}_n(\theta) = \theta_1 \left(\frac{1}{n_1} \sum_{i : X_i \in C_1} \delta_{X_i} \right) + (1-\theta_1) \left(\frac{1}{n_2} \sum_{i : X_i \in C_2} \delta_{X_i} \right)$$

- Observations non-classifiées (droite) $Y_1, \ldots, Y_m$
  $\longrightarrow \hat{\nu}_m := \frac{1}{m} \sum_{j=1}^{m} \delta_{Y_j}$

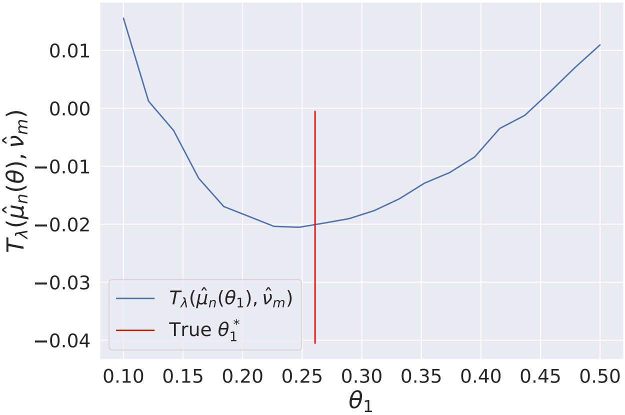
Estimateur des proportions

Classification supposée inconnue dans $Y_1, \ldots, Y_m$

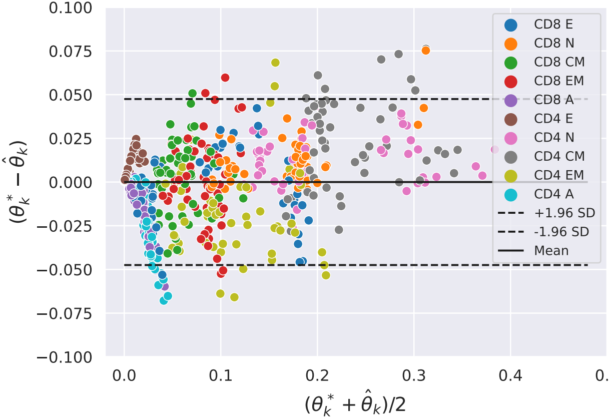
Objectif : proportions par classe dans $Y_1, \ldots, Y_m$

$$\hat{\theta}_\lambda = (\hat{\theta}_1, \ldots, \hat{\theta}_K) := \underset{\theta \in \Sigma_K}{\arg\min}\, \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \hat{\nu}_m)$$

$\mathcal{T}_\lambda : \mathcal{P}(\mathbb{R}^d)^2 \to \mathbb{R}_+$ coût de transport régularisé (Redko et al., 2019)



Résultats sur données HIPC



$K = 10$ sous-populations, $d = 7$ paramètres, $n \in [10^5, 10^6]$ cellules

- Diagramme de Bland-Altman - référence : méthode manuelle
- Un jeu de données : $K = 10$ points $((\hat{\theta}_k + \theta_k^*)/2, \hat{\theta}_k - \theta_k^*)_{1 \leq k \leq 10}$
- Points proches de l'axe $y = 0$: bons résultats

Ça fonctionne mais...

- **CytOpT** : $\hat{\theta}_\lambda = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \hat{\nu}_m)$
  (Freulon, Bigot, Hejblum, 2022, AOAS)
- Si λ trop grand, la méthode **échoue**
- Si $Y_1, \ldots, Y_m \sim_{\text{i.i.d.}} \nu$ où $\nu = \sum_{k=1}^{K} \theta_k^* \mu_k$

Questions

- Proximité $\hat{\theta}_\lambda$ et poids θ^* ?
- Comment choisir λ ?
- Calcul efficace de $\hat{\theta}_\lambda$?

Plan

1. Transport optimal régularisé

2. Estimation des poids et analyse statistique

3. Expériences et calcul efficace d'un estimateur des poids

Transport optimal régularisé

Problème de transport optimal de Kantorovich (1942)

μ, ν : **mesures de probabilités** sur $\mathbb{R}^d$ à support **compact**
Coût (au sol) : $c(x, y) = \|x - y\|_2^2$.

Coût de transport optimal entre μ et ν

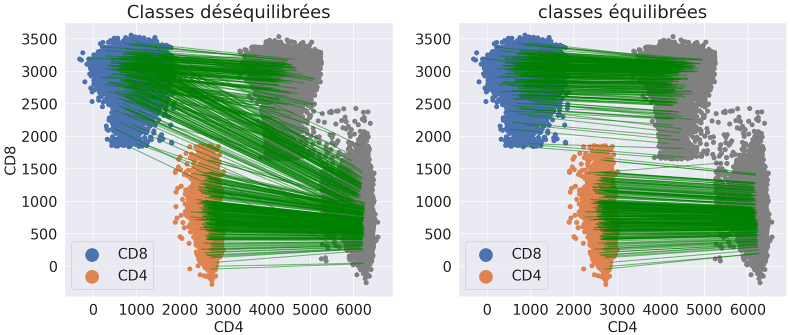
$$\mathcal{T}_0(\mu, \nu) := \underbrace{\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} ||x - y||_2^2 \mathrm{d}\pi(x, y)}_{\text{problème de Kantorovich}}$$

$\Pi(\mu, \nu)$ mesures de probabilité ayant pour marginales μ et ν

- $\sqrt{\mathcal{T}_0(\mu, \nu)}$ distance de transport
- Plan de transport optimal : π^* solution problème
- Existence π^* : oui
- Unicité π^* : pas nécessairement

Transport discret : $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$ et $\nu = \sum_{j=1}^{m} b_j \delta_{y_j}$

$$\mathcal{T}_0(\mu, \nu) = \min_{\pi \in \mathbb{R}_+^{n \times m}} \sum_{i,j} \|x_i - y_j\|^2 \pi_{i,j} \quad \text{où} \quad \pi 1_m = a \text{ et } \pi^T 1_n = b$$



- Gauche $\mathcal{T}_0(\hat{\mu}_n, \hat{\nu}_m) \longrightarrow a_i = 1/n$ et $b_j = 1/m$
- Droite $\mathcal{T}_0(\hat{\mu}_n(\hat{\theta}_\lambda), \hat{\nu}_m) \longrightarrow a_i = a_i(\hat{\theta}_\lambda)$ et $b_j = 1/m$

Coût algorithmique important : $m = n$, $\mathcal{T}_0(\mu, \nu) \Rightarrow O(n^3 \log(n))$

Transport optimal régularisé (M.Cuturi, 2013)

$$\mathcal{T}_{\lambda}(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^2 \mathrm{d}\pi(x,y) + \underbrace{\lambda\,\mathrm{KL}(\pi|\mu\otimes\nu)}_{\text{régularisation}}$$

$\lambda \geq 0$ **paramètre de régularisation**

$\mathrm{KL}(\pi|\mu\otimes\nu) = \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\mu\otimes\nu}(x,y)\right)\mathrm{d}\pi(x,y)$ si $\pi \ll \mu\otimes\nu$

- Bénéfice algorithmique : cas discret $m = n$
  Calcul de $\mathcal{T}_{\lambda}(\mu,\nu)$ en $O(n^2\log(n))$ opérations
- Problème mieux posé : existence **et** unicité de π^*_{λ} solution
- **Cette présentation** : étude des performances statistiques de $\mathcal{T}_{\lambda}(\mu,\nu)$ par rapport à $\mathcal{T}_{0}(\mu,\nu)$

Problème : choix de λ

Formulation duale

Transport classique et transport régularisé

$$\mathcal{T}_\lambda(\mu,\nu) = \max_{\substack{\varphi \in L^\infty(\mathcal{X}), \\ \psi \in L^\infty(\mathcal{Y})}} \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu - \int_{\mathcal{X}\times\mathcal{Y}} m_\lambda(\varphi+\psi) \mathrm{d}\mu \mathrm{d}\nu$$

Variables duales: φ, ψ

- Si $\lambda = 0$, alors $m_0(\varphi+\psi) = +\infty 1_{\varphi(x)+\psi(y) > \|x-y\|^2}$
- Si $\lambda > 0$, alors $m_\lambda(\varphi+\psi) = \lambda \exp\left(\frac{\varphi(x)+\psi(y)-\|x-y\|^2}{\lambda}\right)$

Intérêt de la formulation duale

- **Solutions numériques** : $\lambda = 0$ algorithme des enchères
  $\lambda > 0$ algorithme de **Sinkhorn**, algorithme **stochastique**
- **Analyse statistique** : contrôle de $|\mathcal{T}_\lambda(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|$ par $\sup_{\varphi\in\mathcal{F}} |\int \varphi \mathrm{d}(\nu - \hat{\nu}_n)|$
  $\mathcal{F}$ contient les variables duales optimales

Estimation des poids et analyse statistique

Application du transport optimal en statistique

Objectif: Approcher une mesure de probabilité inconnue ν avec un modèle paramétrique $\{\mu_\theta \mid \theta \in \Theta\}$ où $\Theta \subset \mathbb{R}^D$

Ici: Modèle de mélange à K composantes $\mu_1, \ldots, \mu_K$:

$$\mathcal{M}_K := \left\{ \mu(\theta) := \sum_{k=1}^{K} \theta_k \mu_k \mid \theta \in \Sigma_K \right\}$$

Projection de ν sur $\mathcal{M}_K$ au sens du transport optimal

$$\theta^* \in \underset{\theta \in \Sigma_K}{\arg\min}\, \mathcal{T}_0(\mu(\theta), \nu) \quad \text{i.e.} \quad \forall \theta \in \Sigma_K,\ \mathcal{T}_0(\mu(\theta^*), \nu) \leq \mathcal{T}_0(\mu(\theta), \nu)$$

Problème : en pratique ν inconnue

Estimation de $\theta^* = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_0(\mu(\theta), \nu)$

Hypothèse :

- modèle $\left\{ \mu(\theta) := \sum_{k=1}^{K} \theta_k \mu_k \mid \theta \in \Sigma_K \right\}$ connu
- observations $Y_1, \ldots, Y_n \sim_{\text{i.i.d.}} \nu$ disponibles

On remplace ν par $\hat{\nu}_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{Y_j}$
$\to$ famille d'estimateurs $(\hat{\theta}_\lambda)_{\lambda \geq 0}$

$$\hat{\theta}_\lambda = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n) \quad \text{où} \quad \lambda \geq 0.$$

- Convergence de $\hat{\theta}_\lambda = \hat{\theta}_\lambda^{(n)}$ vers θ^* quand n augmente ?
- Comment choisir λ ?
- Solutions numériques pour calculer $\hat{\theta}_\lambda$ en pratique ?
  En cytométrie $n \in [10^5, 10^7]$.

Écart entre $\hat{\theta}_\lambda$ et θ^*

Objectif : contrôler l'écart entre

$$\theta^* = \operatorname*{arg\,min}_{\theta \in \Sigma_K} \mathcal{T}_0(\mu(\theta), \nu) \quad \text{et} \quad \hat{\theta}_\lambda = \operatorname*{arg\,min}_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)$$

Critère : $r(\hat{\theta}_\lambda, \theta^*) := \mathcal{T}_0(\mu(\hat{\theta}_\lambda), \nu) - \mathcal{T}_0(\mu(\theta^*), \nu)$

Fait :

$$0 \leq r(\hat{\theta}_\lambda, \theta^*) \leq 2 \sup_{\theta \in \Sigma_K} |\mathcal{T}_0(\mu(\theta), \nu) - \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)|$$

Fixons $\theta \in \Sigma_K$ et regardons

$$|\mathcal{T}_0(\mu(\theta), \nu) - \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)|$$

Estimation du coût de transport $\mathcal{T}_0(\mu, \nu)$

- Mesure μ **connue**
- $Y_1, \ldots, Y_n \sim_{\text{i.i.d.}} \nu \to \hat{\nu}_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{Y_j}$

Contrôle de

$$|\mathcal{T}_0(\mu, \nu) - \mathcal{T}_\lambda(\mu, \hat{\nu}_n)|$$

$\mathcal{T}_\lambda(\mu, \hat{\nu}_n)$ est un estimateur de $\mathcal{T}_0(\mu, \nu)$

- Résultat non-asymptotique

$$\mathbb{E}\left[|\mathcal{T}_0(\mu, \nu) - \mathcal{T}_\lambda(\mu, \hat{\nu}_n)|\right] \leq B(n, d, \lambda)$$

Puis en **déduire un choix de** λ

Estimation de $\mathcal{T}_0(\mu, \nu)$: état de l'art

Théorème (Chizat et al., 2020)

Si μ et ν sont à support compact et $d > 4$

$$\mathbb{E}\left[\left|\mathcal{T}_0(\mu, \nu) - \mathcal{T}_0(\mu, \hat{\nu}_n)\right|\right] \leq C_0 n^{-2/d}$$

Peut-on espérer mieux ?

Estimation de $\mathcal{T}_0(\mu,\nu)$: état de l'art

Théorème (Chizat et al., 2020)

Si μ et ν sont à support compact et $d > 4$

$$\mathbb{E}\left[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_0(\mu,\hat{\nu}_n)|\right] \leq C_0 n^{-2/d}$$

Peut-on espérer mieux ?

Théorème (Manole and Niles-Weed, 2021)

Borne inférieure "minimax"

$$\inf_{\widehat{\mathcal{T}}_n} \sup_{\substack{\mu\in\mathcal{P}(\mathcal{X}),\\ \nu\in\mathcal{P}(\mathcal{Y})}} \mathbb{E}\left[\left|\mathcal{T}_0(\mu,\nu) - \widehat{\mathcal{T}}_n\right|\right] \geq C_1(n\log n)^{-2/d}$$

infimum calculé sur l'ensemble des estimateurs de $\mathcal{T}_0(\mu,\nu)$

Pour l'estimateur régularisé $\mathcal{T}_\lambda(\mu,\hat{\nu}_n)$ avec $\lambda > 0$?

$$\mathbb{E}\left[|\mathcal{T}_\lambda(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|\right] \lesssim \left(1 + \frac{1}{\lambda^{d/2}}\right)\frac{1}{\sqrt{n}}$$

Décomposition approximation-estimation

$$|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)| \leq \overbrace{|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_\lambda(\mu,\nu)|}^{\textit{approximation}} + \overbrace{|\mathcal{T}_\lambda(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|}^{\textit{estimation}}$$

Contrôle décomposition approximation-estimation

À partir de (Genevay et al., 2019) et (Chizat et al, 2020)

$$\mathbb{E}[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|] \lesssim \underbrace{\lambda \log(\lambda^{-1})}_{\textit{approximation}} + \underbrace{\left(1 + \frac{1}{\lambda^{d/2}}\right)\frac{1}{\sqrt{n}}}_{\textit{estimation}}$$

Compromis approximation-estimation:

$$\lambda_n = n^{-1/(d+2)} \Rightarrow \mathbb{E}[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_{\lambda_n}(\mu,\hat{\nu}_n)|] \lesssim n^{-1/d+2}$$

Choix de λ, mais vitesse sous-optimale ($n^{-2/d}$ avec $\mathcal{T}_0(\mu,\hat{\nu}_n)$)

Point faible : borne d'estimation $(1+\lambda^{-d/2})/\sqrt{n}$

Une erreur d'estimation indépendante de λ

Théorème (Bigot, Freulon, Hejblum, Leclaire)

Pour $\lambda \geq 0$, en dimension $d > 4$

$$\mathbb{E}\left[|\mathcal{T}_\lambda(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|\right] \leq C_0 n^{-2/d}$$

C_0 dépend des supports, de la dimension, **pas** de λ

Retour à la décomposition approximation-estimation :

$$\mathbb{E}[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|] \lesssim \underbrace{\lambda \log(\lambda^{-1})}_{\text{approximation}} + \underbrace{n^{-2/d}}_{\text{estimation}}$$

- Choisir λ_n t.q. $|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_{\lambda_n}(\mu,\nu)| \lesssim \lambda_n \log(\lambda_n^{-1}) \leq n^{-2/d}$

Corollaire (Bigot, Freulon, Hejblum, Leclaire)

Avec $\lambda_n = \frac{n^{-2/d}}{d}$ si $d > 4$

$$\mathbb{E}\left[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_{\lambda_n}(\mu,\hat{\nu}_n)|\right] \lesssim n^{-2/d} \log(n)$$

Vitesse quasiment optimale

Idées de preuve concernant l'erreur d'estimation

En notant $\hat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ où $Y_1, \ldots, Y_n \sim_{\text{i.i.d.}} \nu$, on a

$$|\mathcal{T}_\lambda(\mu, \nu) - \mathcal{T}_\lambda(\mu, \hat{\nu}_n)| \leq \sup_{\varphi \in \mathcal{F}} \left| \int \varphi \mathrm{d}(\nu - \hat{\nu}_n) \right|$$

$\mathcal{F}$ contient les variables optimales pour le problème dual
$\mathcal{T}_\lambda(\mu, \nu) = \max_{\varphi,\psi} \int_{\mathcal{X}} \psi \mathrm{d}\mu + \int_{\mathcal{Y}} \varphi \mathrm{d}\nu - \int_{\mathcal{X}\times\mathcal{Y}} m_\lambda(\varphi + \psi) \mathrm{d}\mu \mathrm{d}\nu$

Fait : $\|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2$
Nouveau problème de transport $\to$ coût $s(x, y) = -\langle x, y\rangle$

$$\mathcal{T}_\lambda^s(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \underbrace{-\langle x, y\rangle}_{s(x,y)} \mathrm{d}\pi(x, y) + \lambda\, \mathrm{KL}(\pi | \mu \otimes \nu)$$

Lemme : $\mathcal{F}^s = \{$fonctions convexes et Lipschitziennes$\}$
Résultat connu (Chizat et al. 2020) :

$$\mathbb{E}\left[\sup_{\varphi \in \mathcal{F}^s} \left| \int \varphi \mathrm{d}(\nu - \hat{\nu}_n) \right| \right] \lesssim n^{-2/d}$$

Bilan sur le choix de λ

$$\text{On a} \quad \mathbb{E}\left[|\mathcal{T}_0(\mu,\nu) - \mathcal{T}_\lambda(\mu,\hat{\nu}_n)|\right] \lesssim \lambda \log(\lambda^{-1}) + n^{-2/d}$$

Théorie : plus λ est petit, plus $\mathcal{T}_\lambda(\mu,\hat{\nu}_n)$ est proche de $\mathcal{T}_0(\mu,\nu)$

Pratique : prendre λ petit ralentit le calcul numérique de $\mathcal{T}_\lambda(\mu,\hat{\nu}_n)$

Compromis : erreur d'approximation – gain algorithmique

Ce dont on ne parlera pas :

- Réduction de l'erreur d'approximation $\to \mathcal{S}_\lambda(\mu,\nu)$
- Réduction du coût algorithmique : limitation du nombre d'itérations de l'algorithme de Sinkhorn

Choix de λ pour l'estimation des poids

Retour sur les modèles de mélange :

$$\theta^* = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_0(\mu(\theta), \nu) \quad \text{et} \quad \mu(\theta) = \sum_{k=1}^{K} \theta_k \mu_k$$

Analyse similaire pour $\hat{\theta}_\lambda = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)$

Théorème (Bigot, Freulon, Hejblum, Leclaire)

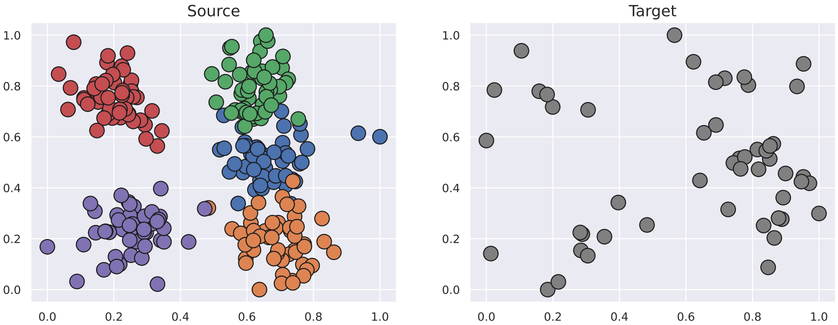
$\text{supp}(\nu)$ compact et $\forall k \in \{1, \ldots, K\}$, $\text{supp}(\mu_k)$ compact.
Avec $\lambda_n = \frac{n^{-2/d}}{d}$

$$0 \leq \mathbb{E}\left[\mathcal{T}_0(\mu(\hat{\theta}_{\lambda_n}), \nu) - \mathcal{T}_0(\mu(\theta^*), \nu)\right] \lesssim n^{-2/d} \log(n)$$

Extension : probablement vrai pour un modèle $(\mu_\theta)_{\theta \in \Theta}$ s'il existe $\mathcal{K}$ borné tel que pour tout $\theta \in \Theta$, $\text{supp}(\mu_\theta) \subset \mathcal{K}$

Expériences et calcul efficace d'un estimateur des poids

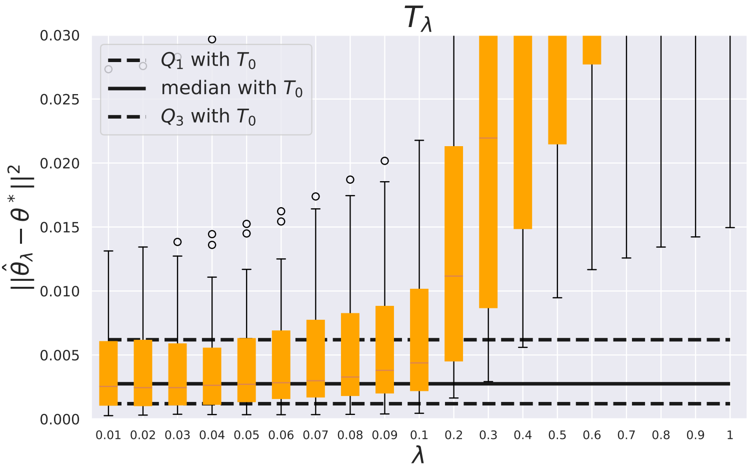
Données synthétiques



Projection 2D d'observations selon un mélange gaussien de $K = 5$ composantes en dimension $d = 6$

- Dans ces expériences, λ varie entre $\lambda = 0.01$ et $\lambda = 1$
- Calcul $\hat{\theta}_\lambda = \min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)$: descente de gradient
- $\nu = \sum_{k=1}^{K} \theta_k^* \mu_k$, donc θ^* connu !

Évolution des performances de $\hat{\theta}_\lambda$ en fonction de λ

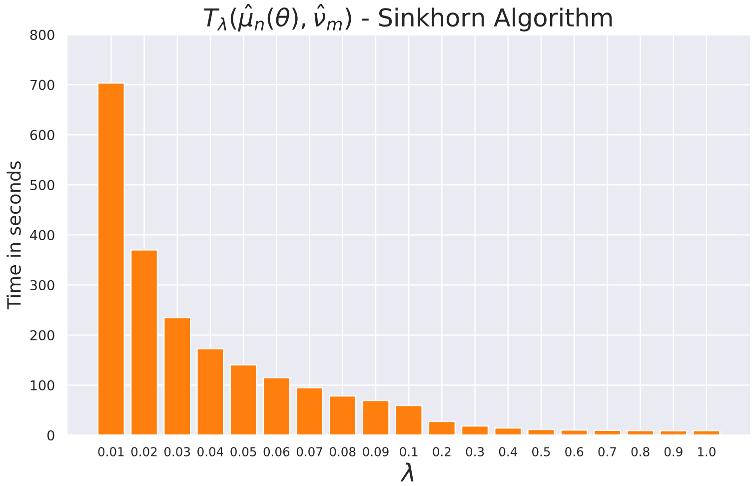


Pour λ fixé, on calcule 50 estimateurs $\hat{\theta}_\lambda^{[1]}, \ldots, \hat{\theta}_\lambda^{[50]}$ de θ^*

Expérimentalement : λ petit $\Rightarrow \|\hat{\theta}_\lambda - \theta^*\|^2$ petit

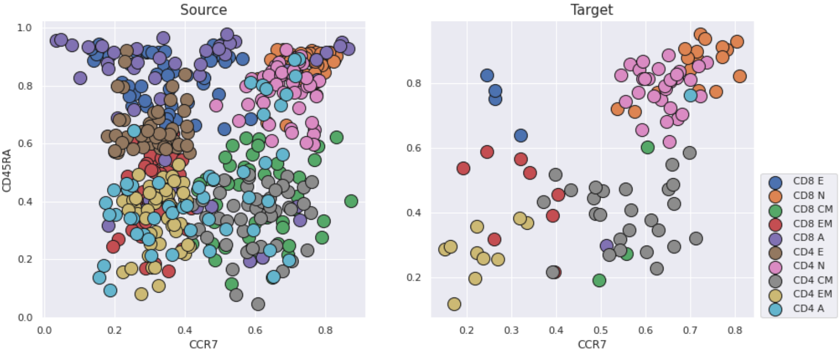
Ici $n = 50,\ d = 6$ donc $\lambda_n \approx 0.05$

Temps de calcul de $\hat{\theta}_\lambda$ en fonction de λ



Temps nécessaire pour calculer 50 estimateurs des poids $\hat{\theta}_\lambda$

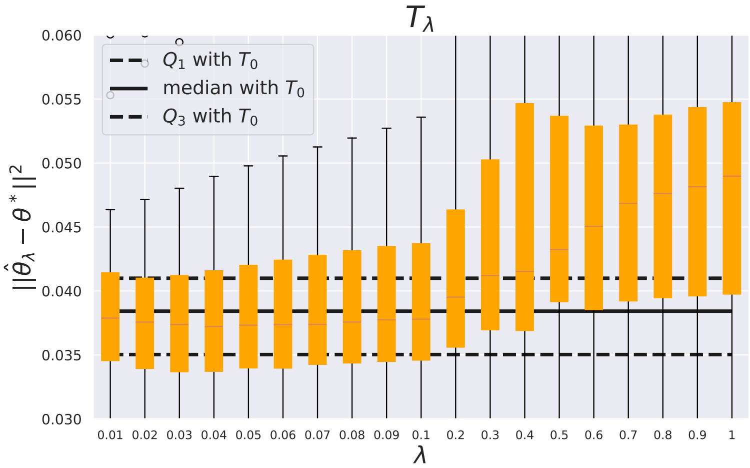
Résultats numériques sur données de cytométrie en flux



Projection 2D des mesures de cytométrie en flux. $K = 10$ sous-populations cellulaires identifiées. $d = 7$ paramètres biologiques sont mesurés

$\lambda > 0$ varie entre $\lambda = 0.01$ et $\lambda = 1$

Performance de $\hat{\theta}_\lambda$ en fonction de λ (données de cytométrie)

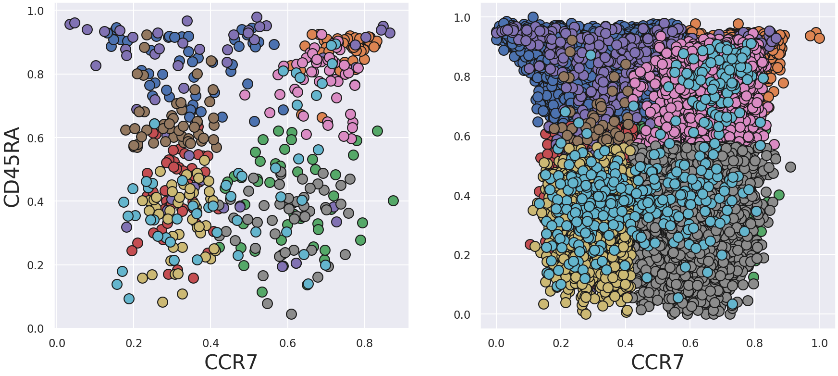


Calcul de $\nabla_\theta \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)$: algorithme de Sinkhorn

Expérimentalement : λ petit $\Rightarrow \|\hat{\theta}_\lambda - \theta^*\|^2$ petit

Ici $n = 100$, $d = 7$, $\lambda_n \approx 0.04$

Calcul et minimisation d'un coût de transport



Gauche : sous-échantillon $n = 500$. droite : jeu complet $n \approx 31000$

Calcul de $\mathcal{T}_\lambda(\mu, \hat{\nu}_n)$ ou de $\nabla_\theta \mathcal{T}_\lambda(\mu(\theta), \hat{\nu}_n)$

- Discret - discret : Sinkhorn (1964), Cuturi (2013)
- Discret - continu (ou grand support) : Genevay et al. (2016), Bercu et al. (2020)

Calcul de l'estimateur $\hat{\theta}_\lambda = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu)$

Problème d'optimisation :

$$\min_{\theta \in \Sigma_K} \overbrace{\max_{\varphi \in \mathbb{R}^n} \mathcal{J}_\lambda(\varphi, \theta)}^{\mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu)}$$

$\mathcal{J}_\lambda$ fonction duale $\longrightarrow \mathcal{J}_\lambda(\varphi, \theta) = \int_{\mathcal{X}} \varphi \mathrm{d}\hat{\mu}_n(\theta) + \int_{\mathcal{Y}} \varphi^{c,\lambda} \mathrm{d}\nu$

Calcul de l'estimateur $\hat{\theta}_\lambda = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu)$

Problème d'optimisation :

$$\min_{\theta \in \Sigma_K} \overbrace{\max_{\varphi \in \mathbb{R}^n} \mathcal{J}_\lambda(\varphi, \theta)}^{\mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu)}$$

$\mathcal{J}_\lambda$ fonction duale $\longrightarrow \mathcal{J}_\lambda(\varphi, \theta) = \int_{\mathcal{X}} \varphi \mathrm{d}\hat{\mu}_n(\theta) + \int_{\mathcal{Y}} \varphi^{c,\lambda} \mathrm{d}\nu$

Régularisation sur l'espace des paramètres (Ballu et al. 2020)

$$\hat{\theta}_{\lambda,\tau} := \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu) + \tau\, \mathrm{H}(\theta)$$

où $\mathrm{H}(\theta) = \sum_{k=1}^{K} \theta_k \log(\theta_k)$ et $\tau \geq 0$

- Modèle discret $\{\hat{\mu}_n(\theta) = \sum_{k=1}^{K} \theta_k \hat{\mu}_k \mid \theta \in \Sigma_K\}$ (données classifiées)
- On peut simuler $Y_1, Y_2 \ldots \sim_{\text{i.i.d.}} \nu$ (mesures de cytométrie arrivent en ligne)

Régularisation $\tau\,\mathsf{H}$: conséquences

$$
\begin{aligned}
\min_{\theta\in\Sigma_K} \mathcal{T}_{\lambda}(\hat{\mu}_n(\theta),\nu) + \tau\,\mathsf{H}(\theta) &= \min_{\theta\in\Sigma_K} \max_{\varphi\in\mathbb{R}^n} \mathcal{J}_{\lambda}(\varphi,\theta) + \tau\,\mathsf{H}(\theta) \\
&= \max_{\varphi\in\mathbb{R}^n} \min_{\theta\in\Sigma_K} \mathcal{J}_{\lambda}(\varphi,\theta) + \tau\,\mathsf{H}(\theta) \\
&= \max_{\varphi\in\mathbb{R}^n} \mathcal{J}_{\lambda}(\varphi,\chi_{\tau}(\varphi)) + \tau\,\mathsf{H}(\chi_{\tau}(\varphi))
\end{aligned}
$$

Régularisation $\tau \, \mathsf{H}$: conséquences

$$\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu) + \tau \, \mathsf{H}(\theta) = \min_{\theta \in \Sigma_K} \max_{\varphi \in \mathbb{R}^n} \mathcal{J}_\lambda(\varphi, \theta) + \tau \, \mathsf{H}(\theta)$$
$$= \max_{\varphi \in \mathbb{R}^n} \min_{\theta \in \Sigma_K} \mathcal{J}_\lambda(\varphi, \theta) + \tau \, \mathsf{H}(\theta)$$
$$= \max_{\varphi \in \mathbb{R}^n} \mathcal{J}_\lambda(\varphi, \chi_\tau(\varphi)) + \tau \, \mathsf{H}(\chi_\tau(\varphi))$$

Conséquence : $\hat{\theta}_{\lambda,\tau} = \chi_\tau(\varphi^*)$ où φ^* solution de

$$\max_{\varphi \in \mathbb{R}^n} \mathcal{J}_\lambda(\varphi, \chi_\tau(\varphi)) + \tau \, \mathsf{H}(\chi_\tau(\varphi)) = \max_{\varphi \in \mathbb{R}^n} \mathbb{E}_{Y \sim \nu} [h_{\lambda,\tau}(Y, \varphi)]$$

- Algorithme de Robbins-Monro : $(\varphi_\ell)_{\ell \geq 0}$
- Retour vers l'espace des paramètres : $\theta_\ell = \chi_\tau(\varphi_\ell)$

Convergence vers $\hat{\theta}_{\lambda,\tau} = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_{\lambda}(\hat{\mu}_n(\theta), \nu) + \tau\, \mathrm{H}(\theta)$

Théorème (Freulon)

Si $\sup_{\varphi \in \mathbb{R}^n} \mathbb{E}_{Y \sim \nu}\left[h_{\lambda,\tau}(Y, \varphi)\right]$ admet un maximum, et $\nu \ll$ Leb.
On peut calculer une suite $(\theta_\ell)_{\ell \geq 0}$ telle que

$$\theta_\ell \underset{\ell \to +\infty}{\longrightarrow} \hat{\theta}_{\lambda,\tau} \quad \text{presque sûrement}$$

Preuve : même technique que (Bercu et al. 2020)

Remarque : dans le résultat précédent $\mathcal{T}_\lambda = \mathcal{T}_\lambda^{\mathfrak{M} \otimes \mathrm{Leb}}$

Convergence vers $\hat{\theta}_{\lambda,\tau} = \arg\min_{\theta \in \Sigma_K} \mathcal{T}_\lambda(\hat{\mu}_n(\theta), \nu) + \tau \, \mathrm{H}(\theta)$

Théorème (Freulon)

Si $\sup_{\varphi \in \mathbb{R}^n} \mathbb{E}_{Y \sim \nu}\left[h_{\lambda,\tau}(Y, \varphi)\right]$ admet un maximum, et $\nu \ll \mathrm{Leb}$.
On peut calculer une suite $(\theta_\ell)_{\ell \geq 0}$ telle que

$$\theta_\ell \underset{\ell \to +\infty}{\longrightarrow} \hat{\theta}_{\lambda,\tau} \quad \text{presque sûrement}$$

Preuve : même technique que (Bercu et al. 2020)

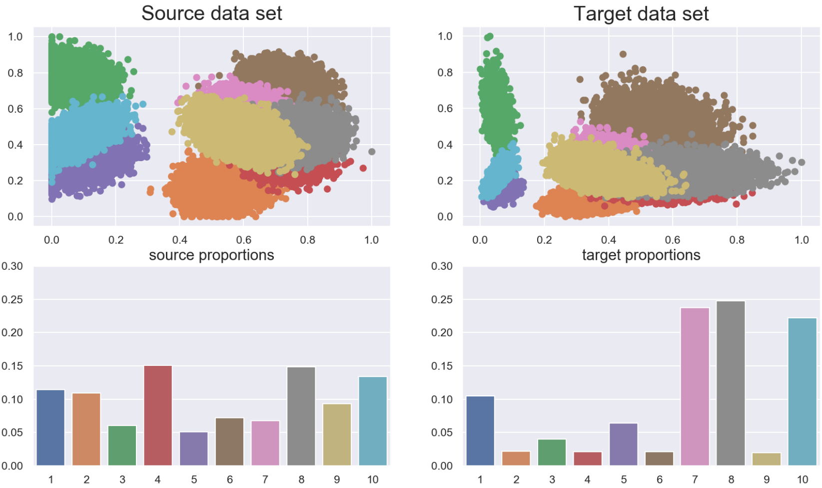
Remarque : dans le résultat précédent $\mathcal{T}_\lambda = \mathcal{T}_\lambda^{\mathfrak{M} \otimes \mathrm{Leb}}$

Heuristique (Freulon)

Avec $\lambda_n = \frac{n^{-2/d}}{d}$ et $\tau_n = \frac{n^{-2/d}}{K}$, l'estimateur $\hat{\theta}_{\lambda,\tau}$ vérifie

$$R(\hat{\theta}_{\lambda_n,\tau_n}, \theta^*) = \mathbb{E}\left[\mathcal{T}_0(\mu(\hat{\theta}_{\lambda_n,\tau_n}), \nu) - \mathcal{T}_0(\mu(\theta^*), \nu)\right] \lesssim n^{-2/d} \log(n)$$

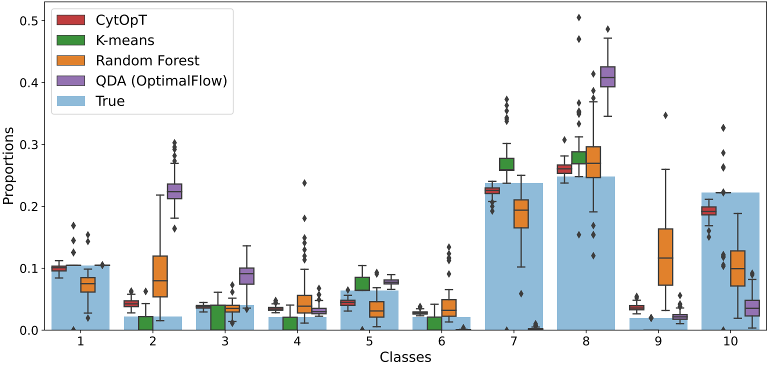
Simulation



- $K = 10, d = 10$ et $n \in [10^5, 10^6]$
- Distributions source et cible différentes

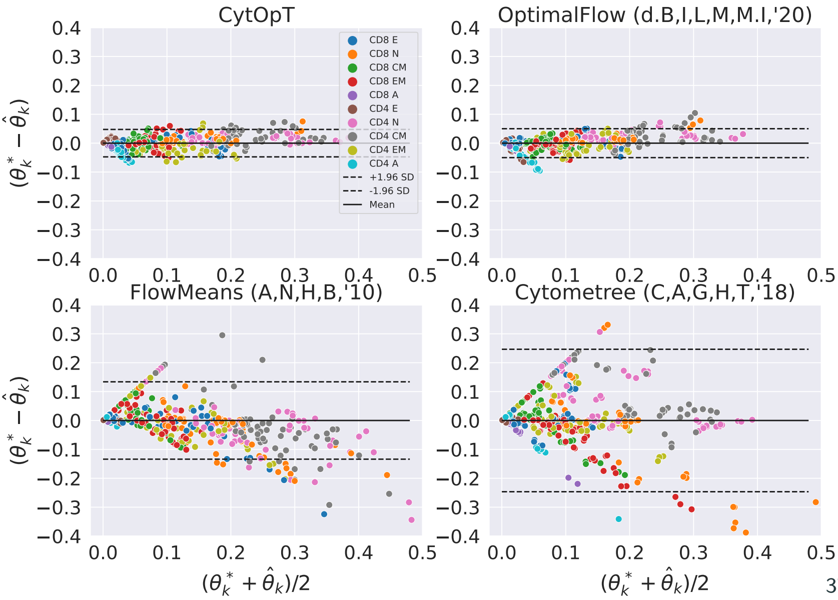
- Comparaison avec des méthodes de classification
- Avec l'objectif d'estimer les proportions

Simulation : résultats

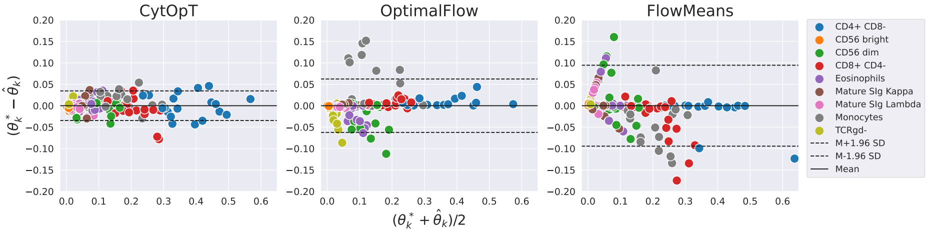


- Calcul de 100 estimateurs $\hat{\theta}^{[1]}_{\lambda,\tau}, \ldots, \hat{\theta}^{[100]}_{\lambda,\tau}$
- Estimation stable

Résultats cytométrie (HIPC)



Résultats cytométrie (OptimalFlow, Del Barrio et al. 2020)



Conclusion et perspectives

Bilan

- Développement d'une nouvelle méthode d'estimation des poids dans un modèle de mélange.
- Minimisation efficace d'un coût de transport régularisé par algorithme stochastique.
- Poursuite des travaux existants sur l'impact statistique du paramètre de régularisation λ.
- Choix du paramètre de régularisation dans des méthodes basées sur la minimisation d'une distance de transport.

Conclusion et perspectives

Bilan

- Développement d'une nouvelle méthode d'estimation des poids dans un modèle de mélange.
- Minimisation efficace d'un coût de transport régularisé par algorithme stochastique.
- Poursuite des travaux existants sur l'impact statistique du paramètre de régularisation λ.
- Choix du paramètre de régularisation dans des méthodes basées sur la minimisation d'une distance de transport.

Valorisation scientifique

- Freulon, Bigot, Hejblum, 2022, AOAS
(participation au développement du package associé avec K.Ba)
- Bigot, Freulon, Hejblum, Leclaire, 2022, soumis

Conclusion et perspectives

Bilan

- Développement d'une nouvelle méthode d'estimation des poids dans un modèle de mélange.
- Minimisation efficace d'un coût de transport régularisé par algorithme stochastique.
- Poursuite des travaux existants sur l'impact statistique du paramètre de régularisation λ.
- Choix du paramètre de régularisation dans des méthodes basées sur la minimisation d'une distance de transport.

Valorisation scientifique

- Freulon, Bigot, Hejblum, 2022, AOAS (participation au développement du package associé avec K.Ba)
- Bigot, Freulon, Hejblum, Leclaire, 2022, soumis

Merci pour votre attention !