Etude de modèles de déformations entre distributions avec la distance de Wasserstein

Hélène Lescornel [1] & Jean-Michel Loubes [2]

[1] *Institut de Mathématiques de Toulouse*
118 route de Narbonne
31000 Toulouse
helene.lescornel@math.univ-toulouse.fr
[2] *Institut de Mathématiques de Toulouse*
118 route de Narbonne
31000 Toulouse
loubes@math.univ-toulouse.fr

Résumé.
Nous présenterons l'étude de modèles de déformation entre distributions, en nous placant sous différents points de vue.

On suppose tout d'abord que les observations proviennent du modèle suivant

$$\begin{cases} \varepsilon_{i1} \quad 1 \leqslant i \leqslant n \\ X_i = \varphi_{\theta^\star}(\varepsilon_{i2}) \quad 1 \leqslant i \leqslant n. \end{cases} \tag{1}$$

On note μ la loi de ε et F la fonction de répartition associée. La déformation est modélisée au travers de la fonction $\varphi_{\theta^\star}$.

On suppose connue la forme de la déformation, c'est à dire la fonction φ mais pas son importance représentée par le paramètre $\theta^\star \in \Theta \subset \mathbb{R}^d$. La loi μ est également inconnue.

Le but de l'exposé est de proposer un estimateur pour la quantité $\theta^\star$ puis de présenter ses propriétés asymptotiques.

L'idée est d'aligner la loi de la variable X sur la loi de ε. Pour cela, pour $\theta \in \Theta$ on considère la variable aléatoire

$$Z_i(\theta) = \varphi_\theta^{-1}(X_i) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2}) \sim \mu_\star(\theta) \quad \forall 1 \leqslant i \leqslant n.$$

En faisant varier θ, nous faisons coincider la distribution $\mu_\star(\theta)$ avec la distribution μ. En effet, on remarque que $\mu_\star(\theta^\star) = \mu$ et cela nous permet donc de retrouver le paramètre de déformation.

Afin de quantifier l'alignement entre ces distributions, nous considérons leur distance de Wassserstein d'ordre 2 qui a pour expression

$$W_2^2(\mu_\star(\theta), \mu) = \int_0^1 \left((F_\theta)^{-1}(t) - (F)^{-1}(t)\right)^2 dt \tag{2}$$

avec F_θ la fonction de répartition de $\mu_\star(\theta)$.

Les données nous permettent d'approcher ce critère théorique en remplacant les lois par leur version empirique. On obtient ainsi le critère suivant, exprimé avec les statistiques d'ordre des échantillons.

$$M_n(\theta) = W_2^2\left(\mu_\star^n(\theta), \mu_1^n\right) = \frac{1}{n}\sum_{i=1}^{n}\left[Z_{(i)}(\theta) - \varepsilon_{(i)1}\right]^2. \tag{3}$$

Nous sommes donc amenés à étudier le M-estimateur du paramètre de déformation

$$\widehat{\theta}^n \in \arg\min_{\theta\in\Theta} M_n(\theta) \tag{4}$$

qui permet également de proposer un estimateur $\widehat{\mu}_n$ pour la densité μ basé sur les observations $(X_i)_{1\leqslant i\leqslant n}$.

Nous exposerons les résultats de consistance pour ces estimateurs obtenus principalement sous des hypothèses de régularité sur les fonctions de déformations. Les preuves suivent les schémas classiques de la théorie de la M-estimation avec des contraintes relativement faibles sur la loi μ.

Nous présenterons ensuite un résultat de convergence en loi pour l'estimateur du paramètre de déformation. Ce résultat, basé sur une Delta-Méthode, requiert des conditions plus fortes sur les distributions étudiées.

Dans un second temps, nous présenterons une extension de ce modèle au cas où les observations proviennent du modèle suivant

$$X_{ij} = \varphi_{\theta_j^\star}(\varepsilon_{ij}) \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant J. \tag{5}$$

Nous proposerons une autre procédure d'estimation pour le paramètre $\theta^\star = (\theta_1^\star, \ldots, \theta_J^\star)$, obtenue en cherchant à aligner les lois sur la distribution moyenne. D'autres techniques de démonstration permettent d'obtenir un théorème de convergence presque sûre de l'estimateur lorsque les observations ne sont plus à valeurs dans $\mathbb{R}$. Ce résultat a été obtenu en collaboration avec J. Cuesta de l'Université de Cantabrie.

Pour finir, nous présenterons des travaux réalisés avec E. Del Barrio permettant de tester l'adéquation de données au modèle (5).

Mots-clés. Distance de Wasserstein , M-estimation, modèle semi-parametrique

Abstract.

We present a study concerning models of deformations between distributions. First, we will assume that the observations are coming from the following model

$$\begin{cases} \varepsilon_{i1} \quad 1 \leqslant i \leqslant n \\ X_i = \varphi_{\theta^\star}(\varepsilon_{i2}) \quad 1 \leqslant i \leqslant n. \end{cases}$$

We denote by μ the law of ε and F the associated distribution function. The deformation is modelled through the function $\varphi_{\theta^\star}$. Here we consider that the shape of the deformation is available through the function φ but that its amount represented by the parameter $\theta^\star \in \Theta \subset \mathbb{R}^d$ is unknown. Our aim is to propose an estimator for the quantity $\theta^\star$ and to present its asymptotic properties.

The idea is to align the distribution of X onto the distribution of ε. For that we consider the random variable

$$Z_i(\theta) = \varphi_\theta^{-1}(X_i) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2}) \sim \mu_\star(\theta) \quad \forall 1 \leqslant i \leqslant n$$

where $\theta \in \Theta$. By varying the parameter θ, we will align the distribution $\mu_\star(\theta)$ with μ. Indeed, as $\mu_\star(\theta^\star) = \mu$ this allows us to identify the deformation parameter.

To quantify the alignment of the distributions, we consider their Wassserstein of order 2 which is given by

$$W_2^2(\mu_\star(\theta), \mu) = \int_0^1 \left((F_\theta)^{-1}(t) - (F)^{-1}(t)\right)^2 dt$$

where F_θ is the distribution function of $\mu_\star(\theta)$.

The observations permit to estimate this criterion with the order statistics of the samples

$$M_n(\theta) = W_2^2(\mu_\star^n(\theta), \mu_1^n) = \frac{1}{n}\sum_{i=1}^n \left[Z_{(i)}(\theta) - \varepsilon_{(i)1}\right]^2.$$

Then we will study the M-estimator defined as

$$\widehat{\theta}^n \in \arg\min_{\theta\in\Theta} M_n(\theta).$$

It allows to propose an estimator $\widehat{\mu}_n$ of the distribution μ based on the observations $(X_i)_{1\leqslant i\leqslant n}$.

We will present consistency results for these estimators obtained mainly with regularity assumptions on the deformations. The proofs follow classical guidelines of M-estimation theory and do not necessitate strong assumptions on μ. Then we will present a result of convergence in distribution for the estimator of the deformation parameter. This one is based on a Delta-method and necessitates stronger assumptions on the distributions of the observations.

In a second time, we will generalize our procedure to the following model

$$X_{ij} = \varphi_{\theta_j^\star}(\varepsilon_{ij}) \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant J.$$

We will propose another estimation procedure for the parameter $\theta^\star = (\theta_1^\star, \ldots, \theta_J^\star)$, by aligning the distributions on their mean. Others techniques of proof allow to obtain a

result of almost sure convergence without assuming that the observations belong to $\mathbb{R}$. This work was done in collaboration with J. Cuesta.

To conclude, we will present the works realized with E. Del Barrio which allow to build a goodness of fit test to the model (5).

Mots-clés. Wasserstein distance, M-estimation, Semi-parametric model

1 Description de la communication

Depuis plusieurs années, les statisticiens se sont intéressés à l'étude de données déformées avec différents objectifs : reconstruire les déformations, retrouver les données structurelles ou encore extraire une moyenne sur les différentes déformations. On peut par exemple citer dans le cadre de données fonctionnelles, la procédure Dynamic Time Warping (D.T.W.) de Sakoe et Chiba dans [3], permettant d'aligner différentes courbes par une renormalisation de l'axe du temps, et ainsi de retrouver les déformations subies par les fonctions.

Nous éloignant du cadre fonctionnel, nous considérons ici un modèle où les données proviennent de différentes déformations d'une même distribution. Plus précisément, on s'intéresse au cas où on dispose de réalisations de variables aléatoires réelles de la forme

$$X_j = \varphi_j(\varepsilon), \quad 1 \leqslant j \leqslant J.$$

On note μ la distribution structurelle de ce modèle, qui est la loi de ε. La déformation est modélisée au travers de la fonction φ_j. On rencontre par exemple cette situation en biologie dans l'étude des puces A.D.N. où les variables X_j représentent les niveaux d'expression de gènes. Le but est alors d'obtenir une distribution moyenne sur les J différentes déformations. Un procédé très largement utilisé est la normalisation quantile, dont les propriétés statistiques ont été étudiées dans [2].

Dans notre cas, nous modélisons le phénomène par un modèle semi-paramétrique en commencant tout d'abord par considérer un modèle plus simple où l'on observe

$$\begin{cases} \varepsilon \\ X = \varphi_{\theta^\star}(\varepsilon) \end{cases}$$

On suppose connue la forme de la déformation, φ, mais pas son importance représenté par le paramètre $\theta^\star$.

Le but de l'exposé est de proposer des estimateurs pour les quantités $\theta^\star$ et μ puis de présenter leurs propriétés asymptotiques.

On suppose pour cela disposer des observations i.i.d. :

$$\begin{cases} \varepsilon_{i1} \quad 1 \leqslant i \leqslant n \\ X_i = \varphi_{\theta^\star}(\varepsilon_{i2}) \quad 1 \leqslant i \leqslant n. \end{cases}$$

En suivant le principe du D.T.W., l'idée est d'aligner la distribution de la variable X sur celle de ε. Pour cela, pour $\theta \in \Theta$ on considère la variable aléatoire

$$Z_i(\theta) = \varphi_\theta^{-1}(X_i) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2}) \sim \mu_\star(\theta) \quad \forall 1 \leqslant i \leqslant n.$$

En faisant varier θ, on essaie de faire coincider la distribution $\mu_\star(\theta)$ sur μ. Elles seront en effet alignées pour le paramètre de déformation inconnu : $\mu_\star(\theta^\star) = \mu$. Afin de quantifier l'alignement entre ces distributions, nous considérons leur distance de Wassserstein d'ordre 2 notée

$$M(\theta) = W_2(\mu_\star(\theta), \mu).$$

On trouve par exemple dans [1] une revue des principales propriétes de cette distance. Cette métrique liée aux problèmes de transport optimal (cette relation est par exemple décrite dans [5]) est adaptée à l'idée de déplacer la masse chargée par $\mu_\star(\theta)$ sur celle de μ. De plus elle est facilement calculable dans notre cadre :

$$M(\theta) = W_2^2(\mu_\star(\theta), \mu) = \int_0^1 \left((F_\theta)^{-1}(t) - (F)^{-1}(t)\right)^2 dt$$

avec F_θ la fonction de répartition de $\mu_\star(\theta)$. Ce critère permet ainsi une caractérisation du paramètre d'intérêt : $M(\theta^\star) = 0 = \min_{\theta\in\Theta} M(\theta)$.

Les données nous permettent d'approcher ce critère théorique en remplacant les lois par leur version empirique. On obtient ainsi la quantité suivante, exprimée grâce aux fonctions de répartition empiriques des échantillons, et qui fait donc intervenir leurs statistiques d'ordre.

$$M_n(\theta) = W_2^2(\mu^n(\theta), \mu_1^n) = \frac{1}{n}\sum_{i=1}^n \left[Z_{(i)}(\theta) - \varepsilon_{(i)1}\right]^2.$$

Nous sommes donc amenés à étudier le M-estimateur du paramètre de déformation

$$\widehat{\theta}^n \in \arg\min_{\theta\in\Theta} M_n(\theta)$$

qui permet également de proposer un estimateur $\widehat{\mu}_n$ pour la densité μ basé sur les observations $(X_i)_{1\leqslant i\leqslant n}$.

Le résultat de consistance pour les paramètres de déformation est obtenu en suivant les schémas classiques de la théorie de la M-estimation (décrite dans [4]). Les hypothèses requises concernent la régularité des fonctions de déformation et une condition d'enveloppe sur celles-ci. Notamment, aucune condition sur le support de la loi μ n'est nécessaire.

Nous présenterons ensuite un résultat de convergence en loi pour l'estimateur du paramètre de déformation. Ce résultat, basé sur une Delta-Méthode requiert des conditions plus fortes sur les distributions étudiées afin d'assurer la Hadamard différentiabilité de fonctionnelles liées aux fonctions quantiles.

Dans un second temps, en revenant vers le problème initial, nous présenterons une extension de cette procédure d'estimation au cas où les observations proviennent du modèle suivant

$$X_{ij} = \varphi_{\theta_j^\star}(\varepsilon_{ij}) \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant J.$$

Nous donnerons une autre procédure d'estimation pour le paramètre $\theta^\star = (\theta_1^\star, \ldots, \theta_J^\star)$, obtenue en cherchant à aligner les lois sur la distribution moyenne. D'autres techniques de démonstration permettent d'obtenir un théorème de convergence presque sûre de l'estimateur lorsque les observations ne sont plus à valeurs dans $\mathbb{R}$. Ce résultat a été obtenu en collaboration avec J. Cuesta de l'Université de Cantabrie.

Pour finir, nous présenterons l'ébauche de travaux réalisés avec E. Del Barrio permettant de tester l'adéquation de données au modèle (5).

Bibliographie

[1] J. A. Cuesta and C. Matran, (1989), *Notes on the Wasserstein metric in Hilbert spaces.* Annals of Probability, 17(3):1264-1276.

[2] S. Gallon, J.-M. Loubes, and E. Maza, (2013),*Statistical properties of the quantile normalization method for density curve alignment.* Mathematical Biosciences, (0):-

[3] H. Sakoe and S. Chiba, (1978), *Dynamic programming algorithm optimization for spoken word recognition.* IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1):43 -49.

[4] A. Van der Vaart, (2000) *Asymptotic statistics.* Number 3. Cambridge Univ Pr.

[5] C. Villani, (2009) *Optimal transport*, volume 338 of Grundlehren der Mathematischen Wis- senschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin. Old and new.