Modèles de régression gaussienne pour des distributions en entrée

49è Journées de Statistiques

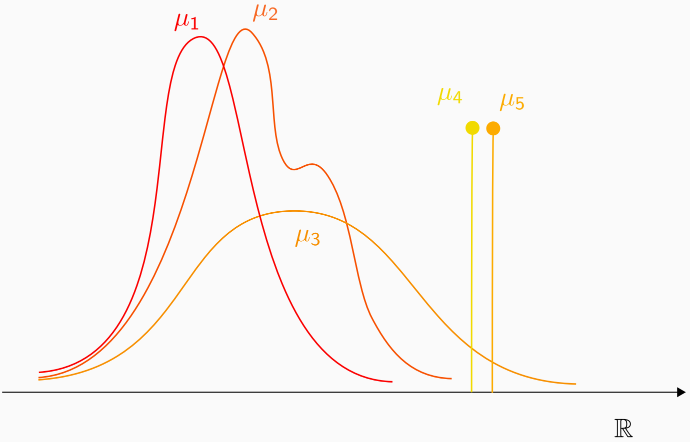
N. Venet[*,⋆], F. Bachoc[⋆], F. Gamboa[⋆], J.-M. Loubes[⋆]

30 mai 2017, Avignon

[*] CEA Tech, [⋆] Institut de Mathématiques de Toulouse

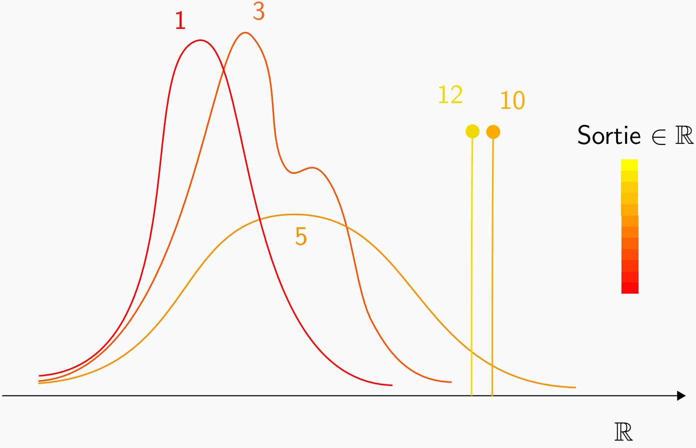
Le problème de régression sur des distributions

On dispose de n couples *entrées / sorties* $(\mu_i, y_i) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}$, et on cherche à associer une sortie à une distribution μ_{n+1}.



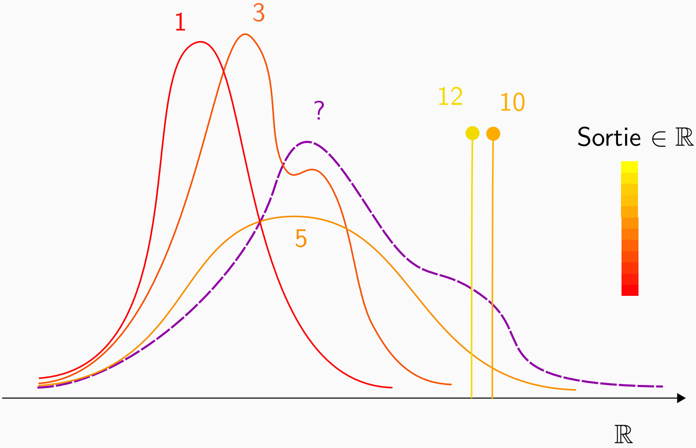
Le problème de régression sur des distributions

On dispose de n couples *entrées / sorties* $(\mu_i, y_i) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}$, et on cherche à associer une sortie à une distribution μ_{n+1}.



Le problème de régression sur des distributions

On dispose de n couples *entrées / sorties* $(\mu_i, y_i) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}$, et on cherche à associer une sortie à une distribution μ_{n+1}.



Motivations

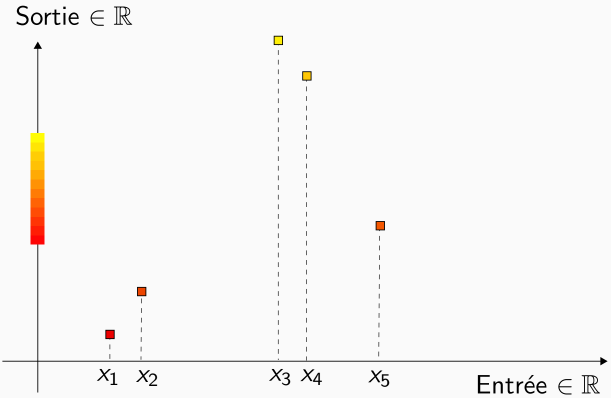
Nos motivations sont doubles : considérer des modèles dont les entrées sont

1. aléatoires ou
2. fonctionnelles (spectres, histogrammes, ...)
   - avec la restriction de positivité et de masse 1,
   - ... qui en contrepartie autorise l'intervention d'outils comme la *distance de Wasserstein*.

Plan de la présentation

Régression gaussienne

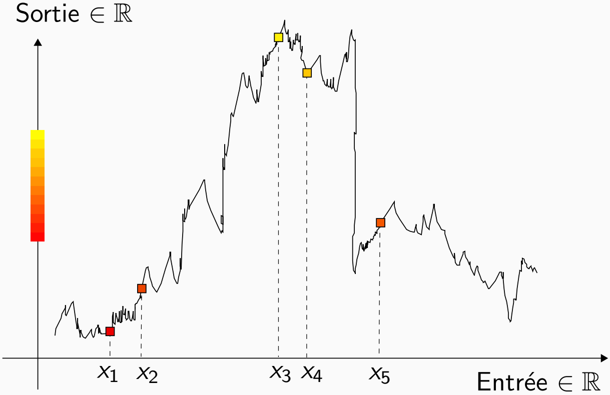
Régression gaussienne (Krigeage)



On choisit un processus aléatoire $(Y_x)_{x\in\mathbb{R}}$ et on considère

$$\hat{Y}(x) := \mathbb{E}(Y_x | Y_{x_1} = y_1, \cdots, Y_{x_n} = y_n)$$

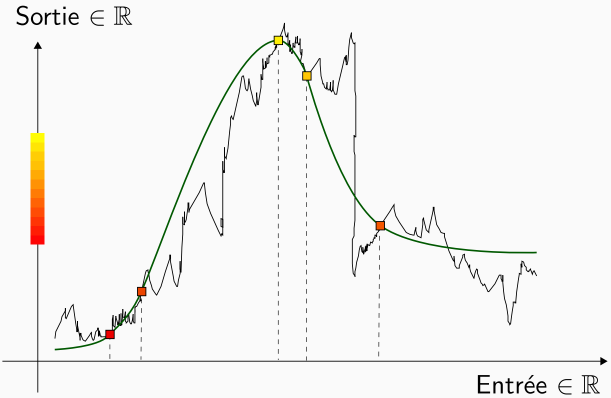
Régression gaussienne (Krigeage)



On choisit un processus aléatoire $(Y_x)_{x\in\mathbb{R}}$ et on considère

$$\hat{Y}(x) := \mathbb{E}(Y_x | Y_{x_1} = y_1, \cdots, Y_{x_n} = y_n)$$

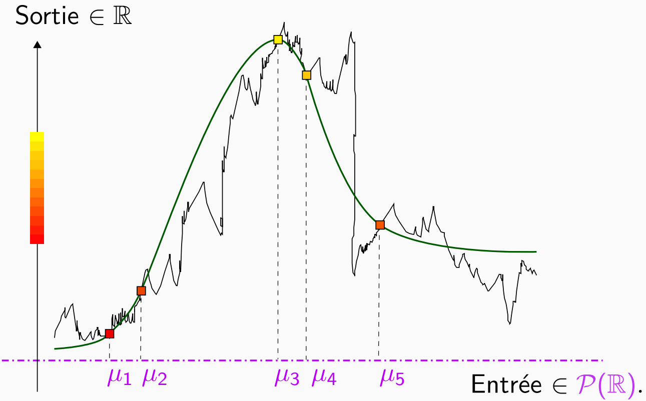
Régression gaussienne (Krigeage)



On choisit un processus aléatoire $(Y_x)_{x\in\mathbb{R}}$ et on considère

$$\hat{Y}(x) := \mathbb{E}(Y_x | Y_{x_1} = y_1, \cdots, Y_{x_n} = y_n)$$

Régression gaussienne (Krigeage)



Ici il faut un processus aléatoire $(Y_\mu)_{\mu \in \mathcal{P}(\mathbb{R})}$ pour considérer

$$\hat{Y}(\mu) := \mathbb{E}(Y_\mu | Y_{\mu_1} = y_1, \cdots, Y_{\mu_n} = y_n)$$

Noyaux sur l'espace de Wasserstein

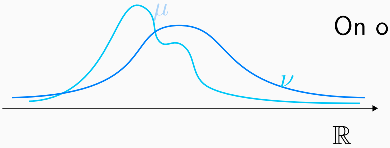
Distance de Wasserstein

La *distance de Wasserstein* entre deux probabilités μ et ν admettant des moments d'ordre deux est définie par :

$\mathbb{R}$

$$W_2(\mu,\nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

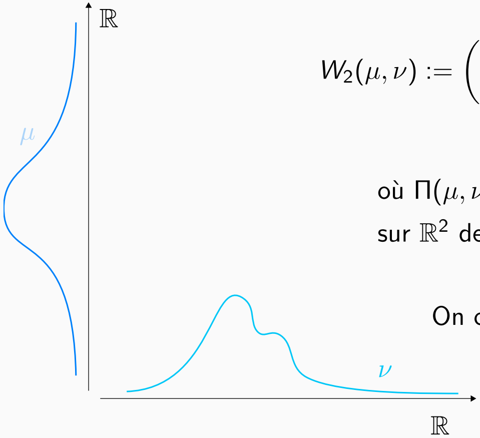
où $\Pi(\mu,\nu)$ est l'ensemble des probabilités sur $\mathbb{R}^2$ de marginales μ et ν.



On obtient un espace métrique $\mathcal{W}_2(\mathbb{R})$.

Distance de Wasserstein

La *distance de Wasserstein* entre deux probabilités μ et ν admettant des moments d'ordre deux est définie par :

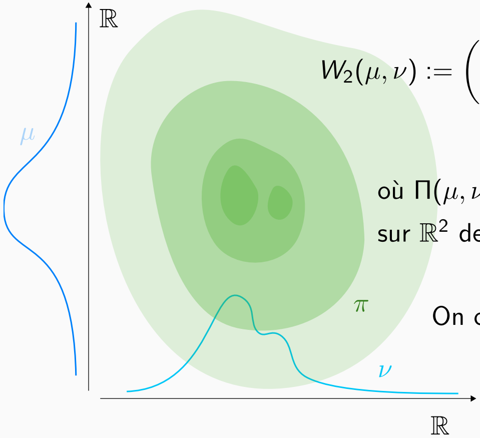


$$W_2(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

où $\Pi(\mu, \nu)$ est l'ensemble des probabilités sur $\mathbb{R}^2$ de marginales μ et ν.

On obtient un espace métrique $\mathcal{W}_2(\mathbb{R})$.

La *distance de Wasserstein* entre deux probabilités μ et ν admettant des moments d'ordre deux est définie par :



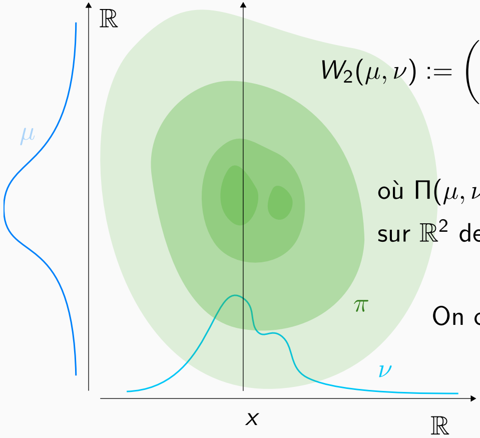
$$W_2(\mu,\nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

où $\Pi(\mu,\nu)$ est l'ensemble des probabilités sur $\mathbb{R}^2$ de marginales μ et ν.

On obtient un espace métrique $\mathcal{W}_2(\mathbb{R})$.

Distance de Wasserstein

La *distance de Wasserstein* entre deux probabilités μ et ν admettant des moments d'ordre deux est définie par :



$$W_2(\mu,\nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

où $\Pi(\mu,\nu)$ est l'ensemble des probabilités sur $\mathbb{R}^2$ de marginales μ et ν.

On obtient un espace métrique $\mathcal{W}_2(\mathbb{R})$.

Une remarque

Pour $\mu, \nu \in \mathcal{W}_2(\mathbb{R})$ et F_μ^{-1}, F_ν^{-1} les fonctions quantiles associées,

$$W_2(\mu, \nu) = \left(\int_{[0,1]} \left(F_\mu^{-1}(u) - F_\nu^{-1}(u) \right)^2 du \right)^{1/2}. \tag{1}$$

- Ce couplage optimal spécifique à la dimension 1 permet l'évaluation numérique de la distance de Wasserstein.
- C'est aussi l'ingrédient central des preuves des Théorèmes 1 et 2.

Noyaux sur l'espace de Wasserstein i

Théorème 1 (Champs browniens fractionnaires)

Pour tout $0 \leq H \leq 1$ *et* $\sigma_0 \in \mathcal{W}_2(\mathbb{R})$,

$$K^{H,\sigma}(\mu,\nu) = \frac{1}{2}\left(W_2^{2H}(\sigma_0,\mu) + W_2^{2H}(\sigma_0,\nu) - W_2^{2H}(\mu,\nu)\right) \qquad (2)$$

est une fonction de covariance sur $\mathcal{W}_2(\mathbb{R})$. *De plus cette fonction est non-dégénérée si et seulement si* $0 < H < 1$.

- On a un champ brownien fractionnaire indexé par $\mathcal{W}_2(\mathbb{R})$. Il est à accroissements stationnaires, nul en l'*origine* σ_0.
- Le paramètre de Hurst H gouverne l'auto-similarité, la régularité des trajectoires et la mémoire à longue distance.

Noyaux sur l'espace de Wasserstein ii

Théorème 2 (Processus stationnaires)

Pour $F : \mathbb{R}^+ \to \mathbb{R}^+$ *complètement monotone et* $0 < H \leq 1$*,*

$$(\mu, \nu) \mapsto F\left(W_2^{2H}(\mu, \nu)\right) \tag{3}$$

est la fonction de covariance d'un processus gaussien stationnaire indexé par $\mathcal{W}_2(\mathbb{R})$*.*

- Rappelons que $F \in C^\infty(\mathbb{R}^+, \mathbb{R}^+)$ est complètement monotone si $(-1)^n F^{(n)}$ est à valeurs positives pour tout $n \in \mathbb{N}$.
- En particulier pour $\sigma^2, \ell > 0$ et $0 \leq H \leq 1$,

$$K_{\sigma^2,\ell,H}(\nu_1, \nu_2) = \sigma^2 \exp\left(-\frac{W_2(\nu_1, \nu_2)^{2H}}{\ell}\right) \tag{4}$$

est une covariance.

Résultats asymptotiques pour l'EMV et la régression

Condition 1 (Cadre asymptotique “en expansion”)

On considère une matrice triangulaire de points d’observations de $\mathcal{W}_2(\mathbb{R})$ $\{\mu_1, ..., \mu_n\} = \{\mu_1^{(n)}, ..., \mu_n^{(n)}\}$ *tels que pour tout* $n \in \mathbb{N}$ *et* $1 \leq i \leq n$, μ_i *est de support inclus dans* $[i, i+K]$, *où* $K < \infty$ *est fixe.*

Condition 2 (Modèle paramétrique stationnaire)

Le modèle de fonctions de covariance $\{K_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ *est tel que*

$$\forall \theta \in \Theta,\ K_\theta(\mu, \nu) = F_\theta\left(W_2(\mu, \nu)\right) \ \text{et} \quad \sup_{\theta \in \Theta} |F_\theta(t)| \leq \frac{A}{1 + |t|^{1+\tau}},$$

avec $A < \infty$ *et* $\tau > 1$ *des constantes.*

Condition 3 (Cas “bien spécifié”)

Nous disposons d'observations $y_i = Y(\mu_i)$, $i = 1, \cdots, n$ *du processus aléatoire gaussien* Y, *centré et de covariance* K_{θ_0} *pour un* $\theta_0 \in \Theta$.

Condition 4 (Non-dégénérescence asymptotique)

La suite de matrices $R_\theta = (K_\theta(\mu_i, \mu_j))_{1 \leq i,j \leq n}$ *est telle que*

$$\lambda_{\inf}(R_\theta) \geq c$$

pour une constante $c > 0$, *où* $\lambda_{\inf}(R_\theta)$ *désigne la plus petite valeur propre de* R_θ.

Condition 5

$\forall \alpha > 0, \liminf_{n \to \infty} \inf_{\|\theta - \theta_0\| \geq \alpha} \frac{1}{n} \sum_{i,j=1}^{n} \left[K_\theta(\mu_i, \mu_j) - K_{\theta_0}(\mu_i, \mu_j)\right]^2 > 0.$

Consistance de l'EMV

Théorème 3 (Consistance de l'EMV))

Sous les conditions 1 à 5 l'estimateur par maximum de vraisemblance est consistant, c'est-à dire :

$$\hat{\theta}_{ML} \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0.$$

Conditions supplémentaires

Condition 6 (Régularité du modèle)

- *$\forall t \geq 0$, $F_\theta(t)$ est $\mathcal{C}^1$ en θ et vérifie $\sup_{\theta\in\Theta} \max_{i=1,\cdots,p} \left|\frac{\partial}{\partial\theta_i} F_\theta(t)\right| \leq \frac{A}{1+t^{1+\tau}}$, où A, τ sont définis dans la Condition 2.*
- *Pour tout $t \geq 0$, $F_\theta(t)$ est $\mathcal{C}^3$ en θ et $\forall q \in \{2,3\}$, $\forall i_1 \cdots i_q \in \{1, \cdots p\}$,*

$$\sup_{\theta\in\Theta} \max_{i=1,\cdots,p} \left|\frac{\partial}{\partial\theta_{i_1}} \cdots \frac{\partial}{\partial\theta_{i_q}} F_\theta(t)\right| \leq \frac{A}{1+|t|^{1+\tau}}.$$

- $\forall(\lambda_1 \cdots, \lambda_p) \neq (0, \cdots, 0)$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \left(\sum_{k=1}^{p} \lambda_k \frac{\partial}{\partial_{\theta_k}} K_{\theta_0}(\mu_i, \mu_j)\right)^2 > 0.$$

Normalité asymptotique de l'EMV

Théorème 4

Soit M_{ML} la matrice de taille $p \times p$ définie par

$$(M_{ML})_{i,j} = \frac{1}{2n} Tr \left(K_{\theta_0}^{-1} \frac{\partial K_{\theta_0}}{\partial \theta_i} K_{\theta_0}^{-1} \frac{\partial K_{\theta_0}}{\partial \theta_j} \right).$$

Sous les conditions 1 à 6, l'estimateur par maximum de vraisemblance est asympotiquement normal. Plus précisément :

$$\sqrt{n}\, M_{ML}^{1/2} \left(\hat{\theta}_{ML} - \theta_0 \right) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, I_p).$$

De plus

$0 < \liminf_{n \to \infty} \lambda_{min}(M_{ML}) \leq \limsup_{n \to \infty} \lambda_{max}(M_{ML}) < +\infty.$

Krigeage sous le modèle estimé par EMV

Théorème 5

Sous les conditions 1 à 6, l'estimateur par Krigeage sous $\hat{\theta}_{ML}$ est asymptotiquement optimal :

$$\forall \mu \in \mathcal{W}_2(\mathbb{R}), \; \left| \hat{Y}_{\hat{\theta}_{ML}}(\mu) - \hat{Y}_{\theta_0}(\mu) \right| = o_{\mathbb{P}}(1).$$

Résultats numériques

Performances sur des données simulées

- On note $m_k(\nu)$ le k-ième moment de ν et considère la fonction

$$F : \mathcal{W}_2(\mathbb{R}) \to \mathbb{R}$$

$$F(\nu) = \frac{m_1(\nu)}{0.05 + \sqrt{m_2(\nu) - m_1(\nu)^2}}, \tag{5}$$

qu'on va chercher à interpoler.

Performances sur des données simulées

- On note $m_k(\nu)$ le k-ième moment de ν et considère la fonction

$$F : \mathcal{W}_2(\mathbb{R}) \to \mathbb{R}$$

$$F(\nu) = \frac{m_1(\nu)}{0.05 + \sqrt{m_2(\nu) - m_1(\nu)^2}}, \tag{5}$$

  qu'on va chercher à interpoler.
- On génère $\nu_1, \cdots, \nu_{100}$ qui sont des gaussiennes de moyennes et de variances tirées uniformément, perturbées aléatoirement afin d'exhiber des irrégularités.

Performances sur des données simulées

- On note $m_k(\nu)$ le k-ième moment de ν et considère la fonction

$$F : \mathcal{W}_2(\mathbb{R}) \to \mathbb{R}$$

$$F(\nu) = \frac{m_1(\nu)}{0.05 + \sqrt{m_2(\nu) - m_1(\nu)^2}}, \tag{5}$$

qu'on va chercher à interpoler.

- On génère $\nu_1, \cdots, \nu_{100}$ qui sont des gaussiennes de moyennes et de variances tirées uniformément, perturbées aléatoirement afin d'exhiber des irrégularités.
- Maximum de vraisemblance sur $\hat{\sigma}^2, \hat{\ell}, \hat{H}$ pour le modèle gaussien paramétrique

$$K_{\sigma^2,\ell,H}(\nu_1, \nu_2) = \sigma^2 \exp\left(-\frac{W_2(\nu_1, \nu_2)^{2H}}{\ell}\right). \tag{6}$$

Performances sur données simulées

- On évalue la méthode sur un jeu de données test $(\nu_{t,i})_{i=1}^{500}$ généré de la même manière que les ν_i,

Performances sur données simulées

- On évalue la méthode sur un jeu de données test $(\nu_{t,i})_{i=1}^{500}$ généré de la même manière que les ν_i, avec les critères :

$$RMSE^2 = \frac{1}{500} \sum_{i=1}^{500} \left(F(\nu_{t,i}) - \hat{F}(\nu_{t,i}) \right)^2 ,$$

$$CIR_\alpha = \frac{1}{500} \sum_{i=1}^{500} \mathbf{1} \left\{ \left| F(\nu_{t,i}) - \hat{F}(\nu_{t,i}) \right| \leq q_\alpha \hat{\sigma}(\nu_{t,i}) \right\} .$$

Performances sur données simulées

- On évalue la méthode sur un jeu de données test $(\nu_{t,i})_{i=1}^{500}$ généré de la même manière que les ν_i, avec les critères :

$$RMSE^2 = \frac{1}{500}\sum_{i=1}^{500}\left(F(\nu_{t,i}) - \hat{F}(\nu_{t,i})\right)^2,$$

$$CIR_\alpha = \frac{1}{500}\sum_{i=1}^{500}\mathbf{1}\left\{\left|F(\nu_{t,i}) - \hat{F}(\nu_{t,i})\right| \leq q_\alpha \hat{\sigma}(\nu_{t,i})\right\}.$$

modèle	RMSE	$CIR_{0.9}$
"Wasserstein"	0.094	0.92
"Legendre" ordre 5	0.49	0.92
"Legendre" ordre 10	0.34	0.89
"Legendre" ordre 15	0.29	0.91
"PCA" ordre 5	0.63	0.82
"PCA" ordre 10	0.52	0.87
"PCA" ordre 15	0.47	0.93

Merci pour votre attention.

Références i

F. Bachoc.
Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes.
Journal of Multivariate Analysis, 125 :1–35, 2014.

F. Bachoc, F. Gamboa, J.-M. Loubes, and N. Venet.
Gaussian process regression model for distribution inputs.
arXiv preprint arXiv :1701.09055, 2017.

C. Berg, J. P. R. Christensen, and P. Ressel.
Harmonic analysis on semigroups.
Springer-Verlag, 1984.

J. Istas.
Manifold indexed fractional fields.
ESAIM Probab. Stat., 16 :222–276, 2012.

Références ii

N. Venet.
Nonexistence of fractional brownian fields indexed by cylinders.
arXiv preprint, 2016.

N. Venet.
On the existence of fractional brownian fields indexed by manifolds with closed geodesics.
arXiv preprint, 2016.

C. Villani.
Optimal transport : old and new, volume 338.
Springer Science & Business Media, 2009.

Avis de recherche :

Jeux de données spatiales sur le cylindre $\mathbb{S}^1 \times \mathbb{R}$.