

# Modèles déformables et consistance statistique

S. Gadat

Institut de Mathématiques de Toulouse  
Université de Toulouse

**Mathématiques pour l'image, Orléans, Juin 2012**

Issus de collaborations avec J. Bigot, D. Bontemps, T. Klein, C. Marteau

## Introduction

Motivations

Modèle statistique déformable

## Méthodes d'estimation de forme moyenne

Données

Définition d'une distance au travers de l'action de groupe

Estimation par recalage + moyenne

Estimation par Moyenne de Fréchet

## Résultats de statistiques mathématiques

$M$  estimation

Sur les procédés de recalage

Approche par déconvolution

Estimation Bayésienne

## Conclusion

## Motivations

**Objectif principal d'un statisticien** : comparer des objets aléatoires qui partagent des caractéristiques communes et extraire des informations sur la distribution de ces objets

### Outils

- ▶ Statistiques du premier ordre : moyenne empirique
- ▶ Statistiques du second ordre : matrice de covariance - Analyse en composante principale (**ACP**)

Soient  $n$  variables aléatoires  $Y_m, m = 1, \dots, n$  iid. La moyenne empirique est

$$\bar{Y}^n = \frac{1}{n} \sum_{m=1}^n Y_m$$

Dans le cas  $\mathcal{H} = \mathbb{R}^p$ , l'ACP consiste à diagonaliser la covariance empirique

$$S = \frac{1}{n} \sum_{m=1}^n (Y_m - \bar{Y}^n)(Y_m - \bar{Y}^n)',$$

et à prendre les premiers vecteurs propres comme modes de variations principaux.

## Exemple pour des visages

**Données** : Images de  $N \times N$  pixels, i.e.  $\mathcal{H} = \mathbb{R}^{N \times N}$



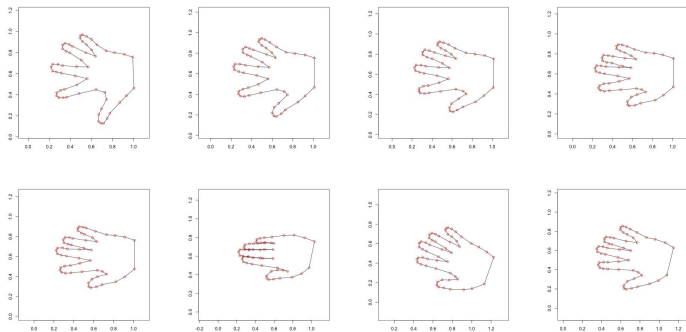
Moyenne empirique inconsistante



$\bar{Y}^n$  n'est pas un visage et l'ACP basée sur  $\bar{Y}^n$  est sans objet.

## Un exemple pour l'analyse de formes

**Données** : formes planes de mains pour  $k = 13$  landmarks  
(points rouges de  $\mathbb{R}^2$ )



La moyenne de ces formes ne sera pas un élément convaincant pour une approximation des données au premier ordre...

## Modes de variations dans les modèles déformables

**Problème :** Les approches standard ne sont pas adaptées dans le cas où les données  $Y_m$  représentent des courbes ou images similaires, déformées, puisque  $\bar{Y}^n$  n'est déjà pas une bonne estimation de la forme moyenne.

**Objectifs idéaux :** Proposer un estimateur "consistant" de la forme moyenne d'un ensemble de courbes ou images et estimer les modes de variations **linéaires et nonlinéaires** autour de la forme moyenne.

**Pour cet exposé :** Décrire ce qui est possible ou non d'un point de vue statistique, dans des cadres plus ou moins limitatifs.

## Modélisation Approche de Grenander(1993)

**Idée** : les  $n$  courbes ou images similaires sont modélisées comme la déformation d'une forme commune par l'action d'un groupe de déformation.

**Modèle déformable** : l'observation  $Y_m : \Omega \rightarrow \mathbb{R}$ ,  $m = 1, \dots, n$  avec  $\Omega \subset \mathbb{R}^d$  pour  $d = 1, 2, 3$  vérifie

$$Y_m(x) = \mathbf{f} \circ \phi_m(\mathbf{x}) + Z_m(\phi_m(x)) + W_m(x), \quad x \in \Omega, \text{ où}$$

- ▶  $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}$  est la forme commune inconnue
- ▶  $\phi_m : \mathbb{R}^d \rightarrow \mathbb{R}^d$  sont des déformations aléatoires qui modélisent des variations **non linéaires** individuelles de la forme.
- ▶  $Z_m(x)$  est un processus centré qui modélise les variations **linéaires** en intensité
- ▶  $W_m$  est un bruit additif indépendant de moyenne nulle

**Dans cet exposé** : comment *mathématiquement* estimer  $\mathbf{f}$  et  $\phi_m$  lorsque  $n \rightarrow +\infty$  ?

On supposera par la suite qu'il n'y a pas de variabilité photométrique.

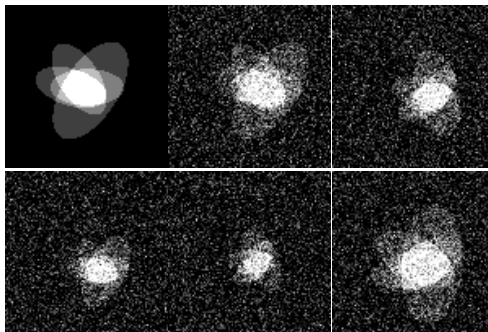
# Différents modèles de déformations

## Déformations rigides

- ▶ Translation :  $\phi(x) = x - \tau$  où  $b \in \mathbb{R}^d$
- ▶ Rotation + scaling (dans  $\mathbb{R}^2$ ) :  $\phi(x) = \frac{1}{a}A_\theta x$  avec  $a \in \mathbb{R}^+$  et

$$A_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

- ▶ Affine (Translation + rotation + scaling) :  $\phi(x) = \frac{1}{a}A_\theta(x - \tau)$ , soit 2D ou 3D
- ▶ Exemple sur le "fantôme de Shepp Logan"





# Différents modèles de déformations

## Déformations élastiques

- ▶ Grande déformations (i.e. **difféomorphismes**) :  $\phi$  est une déformation régulière et inversible de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$  (Cf Travaux de Grenander, Trounev, Younes, Miller,...)
- ▶ Exemple d'actions sur "Lena"



## Inconsistance de la moyenne euclidienne

**Modèle le plus simple** : observation de courbes shiftées aléatoirement  
 $Y_m : [0, 1] \rightarrow \mathbb{R}$ ,  $m = 1, \dots, n$  telles que

$$Y_m(x) = f(x - \tau_m) + W_m(x), \text{ for } x \in [0, 1], \text{ où}$$

- ▶  $f : [0, 1] \rightarrow \mathbb{R}$  est 1-périodique
- ▶ les  $\tau_m$  sont des shifts aléatoires iid de loi de densité  $g$
- ▶ les  $W_m$  sont des bruits additifs centrés indépendants des shifts

**La moyenne empirique est inconsistante** : si  $n \rightarrow +\infty$

$$\bar{Y}^n(x) = \frac{1}{n} \sum_{m=1}^n Y_m(x) \rightarrow \mathbb{E}f(x - \tau_1) = \int f(x - \tau)g(\tau)d\tau = f \star g(x)$$

## Introduction

Motivations

Modèle statistique déformable

## Méthodes d'estimation de forme moyenne

Données

Définition d'une distance au travers de l'action de groupe

Estimation par recalage + moyenne

Estimation par Moyenne de Fréchet

## Résultats de statistiques mathématiques

*M* estimation

Sur les procédés de recalage

Approche par déconvolution

Estimation Bayésienne

## Conclusion

## Données (courbes, images ou formes)

**Observations** : ensemble de  $n$  "objets"  $Y_j, j = 1, \dots, n$  qui peuvent être :

- ▶ courbes ou images i.e.  $Y_j : \Omega \rightarrow \mathbb{R}$  avec  $\Omega \subset \mathbb{R}^q$  e.g.  $\Omega = [0, 1]^q$  et  $q = 1, 2, 3$
- ▶ points d'une forme  $Y_j \in \mathbb{R}^{2k}$  (ensemble de  $k$  landmarks de  $\mathbb{R}^2$ )

**Notion importante** : quelle distance mettre sur les objets observés ?

Choix standard : **distance euclidienne**  $d$  sur  $\mathcal{H}$  associée à  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$

$$d(y, y') = \|y - y'\|_{\mathcal{H}} = \sqrt{\langle y - y', y - y' \rangle_{\mathcal{H}}} \text{ pour } y, y' \in \mathcal{H}$$

**Ce choix conditionne déjà les statistiques du premier ordre...**

Distance Euclidienne  $\Rightarrow$  moyenne empirique usuelle dans  $\mathcal{H}$

$$\bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j = \arg \min_{y \in \mathcal{H}} \sum_{j=1}^n \|Y_j - y\|_{\mathcal{H}}^2$$

**Problème** : ce choix n'est pas adapté dans notre cadre d'objets déformables

## Choix de l'action d'un groupe $G$ sur $\Omega$ - distance

**Données :**  $Y_j : \Omega \rightarrow \mathbb{R}$ ,  $j = 1, \dots, n$  avec  $\Omega \subset \mathbb{R}^q$

**Groupe de déformation :** on fixe (ou connaît) un groupe  $G$  qui agit sur  $\Omega$  et qui modélise **la variabilité géométrique** des observations

**Action du groupe :** si  $g \in G$  et  $x \in \Omega$

- ▶ on note  $\cdot$  l'action de  $G$  sur  $\Omega$
- ▶ on suppose que  $g \cdot x \in \Omega$  si  $(g, x) \in G \times \Omega$

**Exemple :** pour des courbes  $f \in \mathcal{H} = L^2_{per}([0, 1])$ , on peut choisir le groupe des translations

$$G = \mathbb{S}_1$$

(le tore de dimension 1) et pour  $g = g(\theta) \in \mathbb{S}_1$  avec  $\theta \in [0, 1[$  et  $f \in L^2_{per}([0, 1])$ , l'action est alors

$$g \cdot f(t) = f(t - \theta), \quad t \in [0, 1].$$

Notons que  $g$  joue alors le rôle de  $\phi_j$  dans le modèle décrit initialement par

$$Y_j = \mathbf{f} \circ \phi_j + W_j \tag{1}$$

## Idée naive : procédé de recalage

- ▶ On observe  $Y_j, j = 1 \dots n$  issue du modèle (1).
- ▶ On estime tous les paramètres de déformation  $g_j$  qu'on inverse (au sens de  $G$ ) pour obtenir  $\hat{g}_j^{-1}$ .
- ▶ On obtient alors une estimation en utilisant une moyenne empirique des observations recalées :

$$\bar{Y}^n = \frac{1}{n} \sum_{j=1}^n Y \cdot \hat{g}_j^{-1}$$

- ▶ Est-ce envisageable d'estimer correctement les  $g_j$  ? Sous quelle asymptotique (nombre de répétitions, niveau de bruit, ...)

## Une idée exploitant la moyenne de Fréchet

La moyenne empirique est solution de

$$\bar{Y}^n = \arg \min_{h \in \mathcal{H}} \sum_{j=1}^n \int_0^1 d(Y_j, h)^2$$

**Problème :** la distance Euclidienne  $d_E^2(Y_m, h) = \int_{\Omega} |Y_m(x) - h(x)|^2 dx$  n'est pas adaptée à l'analyse des modèles déformables statistiques.

**Choix d'une distance intrinsèque :** utiliser une "mesure de dissemblance" entre les  $Y_m$  et l'action de  $G$

$$d_S^2(Y, h) = \inf_{g \in G} d_E^2(Y, g^{-1} \cdot h)$$

et calculer la solution du problème de minimisation

$$\tilde{Y}^n \in \arg \min_{h \in \mathcal{H}} \sum_j d_S^2(Y_j, h)$$

**Choix d'une distance plus générale :** Pour  $y_1, y_2 \in \mathcal{H}^2$

$$d_G^2(y_1, y_2) = \inf_{g \in G} \left\{ \int_{\Omega} \left| y_1(t) - g^{-1} \cdot y_2(t) \right|^2 dt + \lambda D(g, e) \right\}, \text{ avec } \lambda \geq 0 \quad (2)$$

## Moyenne de Fréchet dans un espace métrique (1948)

Si  $Y_1, \dots, Y_n$  sont des v.a. iid dans un espace métrique  $\mathcal{M}$ , muni de la distance  $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$ , alors la **moyenne de Fréchet empirique** de  $Y_1, \dots, Y_n$  est définie par

$$\tilde{Y}^n \in \arg \min_{y \in \mathcal{M}} \frac{1}{n} \sum_{m=1}^n d^2(y, Y_m).$$

La **moyenne de Fréchet théorique** est définie par

$$Y^F = \arg \min_{y \in \mathcal{M}} \mathbb{E}_{W,g} d^2(y, Y_{W,g})$$

- ▶ Consistance et convergence de  $\tilde{Y}^n$  lorsque  $n \rightarrow +\infty$  démontrée dans le cas où  $\mathcal{M}$  est une variété Riemannienne de **dimension finie** et  $d$  distance géodésique (Bhattacharya & Patrangenaru A.O.S. 2003, A.O.S. 2005). Un exemple classique pour les courbes planes  $\mathcal{M}$  : espace de Kendall (J.R.S.S. 1997).
- ▶ **Problème** : les résultats sur les variétés ne sont généralement pas adaptés à l'étude des moyennes de Fréchet pour les courbes ou images paramétrées comme objets de dimension "infinie".



# Calcul de la moyenne de Fréchet empirique

**Observations :**  $Y_1, \dots, Y_n$  iid de  $L^2(\Omega)$

## Moyenne de Fréchet Empirique

**Étape 1** - Registration/warping des données

$$(\hat{g}_1, \dots, \hat{g}_n) = \arg \min_{(g_1, \dots, g_n) \in G^n} \left\{ \frac{1}{n} \sum_{j=1}^n \int_{\Omega} \left| g_j^{-1} \cdot Y_j(t) - \frac{1}{n} \sum_{m=1}^n g_m^{-1} \cdot Y_m(t) \right|^2 dt + \lambda D(g_j, e) \right\}$$

**Étape 2** - Alignement et moyennisation des données

$$\hat{y}_n = \frac{1}{n} \sum_{j=1}^n \hat{g}_j^{-1} \cdot Y_j$$

## Approximation de la moyenne de Fréchet par la moyenne Procruste

Parfois, la phase de minimisation précédente est délicate numériquement et on lui préfère un schéma de minimisation alternée itératif (Kneip & Gasser (A.O.S. 1988), Wang & Gasser (A.O.S. 1997))

- ▶ Initialisation :  $\hat{f}_0 = \frac{1}{n} \sum_{m=1}^n Y_m$

- ▶ Pour  $1 \leq i \leq i_{\max}$

- ▶ Pour  $1 \leq m \leq n$ , on calcule

$$\hat{g}_{m,i} = \arg \min_{g \in G} d(g \cdot Y_m, \hat{f}_{i-1})$$

- ▶ On estime alors  $\hat{f}_i = \frac{1}{n} \sum_{m=1}^n \hat{g}_{m,i} \cdot Y_m$

Stabilisation rapide ( $i_{\max} = 3$  suffit en général) mais est très sensible à l'initialisation  $\hat{f}_0$ .

## Exemple de moyenne de courbes

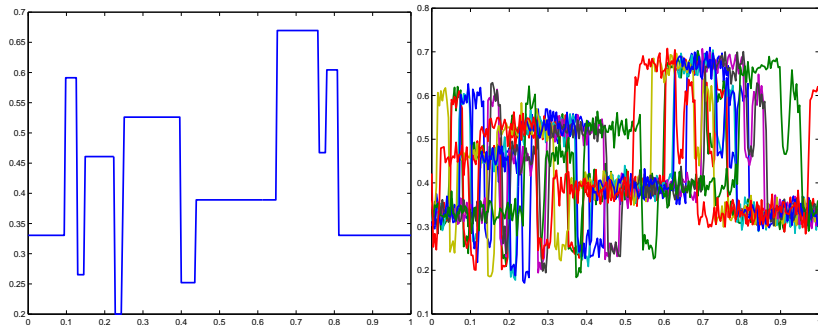
$G = \mathbb{S}^1$  - groupe des translations

**Données** : courbes translattées aléatoirement  $f \in L^2_{per}([0, 1])$  + modèle de bruit blanc gaussien

$Y_j(t) = f(t - \theta_j) + W_j(t)$ , for  $t \in [0, 1]$ , et  $\theta_j$  iid translations aléatoires,

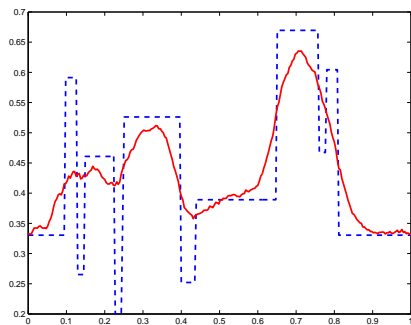
$j = 1, \dots, n$

Courbe  $f$  / Échantillon de 10 parmi  $n = 200$

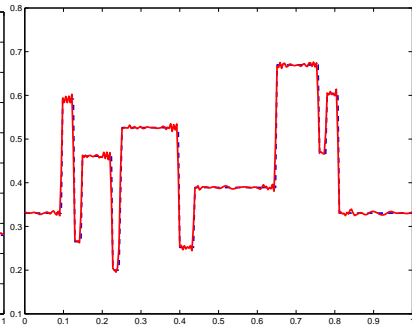


# Exemple de moyenne de courbes

## Moyenne euclidienne



## Moyenne de Fréchet



BG, AOS, '10

# Exemple de moyenne d'images par moyenne Procruste

**Moyenne euclidienne**

**Moyenne Procruste**



## Introduction

Motivations

Modèle statistique déformable

## Méthodes d'estimation de forme moyenne

Données

Définition d'une distance au travers de l'action de groupe

Estimation par recalage + moyenne

Estimation par Moyenne de Fréchet

## Résultats de statistiques mathématiques

*M* estimation

Sur les procédés de recalage

Approche par déconvolution

Estimation Bayésienne

## Conclusion

# Cadre du Modèle simplifié pour des résultats mathématiques

## Objectifs :

- ▶ Critiquer de manière constructive la pertinence des méthodes précédentes
- ▶ Proposer des méthodes d'estimation venues de statistiques mathématiques
- ▶ Quantifier la difficulté d'estimation lorsque  $n$  devient grand,  $\epsilon$  petit . . .

**Cadre :** On se restreindra souvent à la situation où on observe le modèle de bruit blanc gaussien suivant :

$$Y_j(t) = f(t - \theta_j^*) + \epsilon W_j(t), \text{ for } t \in [0, 1], \text{ et } \theta_j \text{ iid translations aléatoires de loi } g.$$

**Inconnus :**  $f$ , les paramètres de translation  $\theta_j^*$ , et parfois  $g$ .

## Résultat général de $M$ estimation

**Cadre :**

$$Y_j = \phi_j \cdot Y^* + \epsilon_j$$

avec  $\phi_j$  une déformation rigide, ou même élastique issue d'un modèle de champ de vecteurs décomposés sur une famille finie :

$$\forall j \in \{1 \dots n\} \quad \phi_j = \phi_j^{(1)} \text{ et } \frac{d\phi_j^{(t)}}{dt} = v_j(\phi_j^{(t)})$$

avec

$$v_j = \sum_{k=1}^K a_{j,k} e_k, \quad (a_{j,k})_{j,k} \text{ coefficients aléatoires et bornés.}$$

**Résultat :** Sous certaines conditions techniques, la moyenne de Fréchet empirique converge presque sûrement vers la moyenne de Fréchet théorique.

**Faiblesse du résultat :** En réalité, on ne dit rien sur la vitesse, sur la nature du minimum, . . . .

*BGL, JMIV '09*



## Estimation des paramètres de recalage dans le modèle de shift (1)

**Cadre :**  $Y_j(t) = f(t - \theta_j^*) + \epsilon W_j(t)$ , bruit blanc et  $\theta_j$  i.i.d.  $\sim g$ .

Le problème est clairement non identifiable en l'état, mais si on suppose

- ▶  $(H_g)$  :  $g$  est à support compact inclus dans  $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$  et de moyenne nulle.
- ▶  $(H_f)$  : La fonction  $f$  est telle que  $c_1(f) > 0$

### Theorem (BG, AOS '10)

Sous  $(H_f)$  et  $(H_g)$ , le modèle est identifiable à la fois en  $f$  et en  $g$ . Par ailleurs, on peut trouver des estimateurs par moyenne de Fréchet des décalages  $(\hat{\theta}_j)_{j=1\dots n}$  satisfaisants  $\sum \hat{\theta}_j = 0$  et

$$\mathbb{P} \left( \frac{1}{n} \sum_{m=2}^n (\hat{\theta}_m - \theta_m^*)^2 \geq C(f, \ell_0, \epsilon, n, t, g) \right) \leq 3 \exp(-t), \quad (3)$$

**Outils :**  $M$  estimation associée à l'inégalité de Bernstein

**Premier bug :** La constante  $C(f, \ell_0, \epsilon, n, t, g)$  ne tend pas vers 0 lorsque  $n$  augmente ! Cela ne se produit qu'uniquement lorsque  $\epsilon$  tend aussi vers 0 (niveau de bruit évanescent, ou augmentation de la résolution).

## Estimation des paramètres de recalage dans le modèle de shift (2)

**Cadre :**  $Y_j(t) = f(t - \theta_j^*) + \epsilon W_j(t)$ , bruit blanc et  $\theta_j$  i.i.d.  $\sim g$ .

**Deuxième bug :** En fait, on ne peut pas estimer les paramètres de recalage correctement à la vue du résultat suivant :

**Theorem (BG, AOS '10)**

Sous  $(H_f)$  et  $(H_g)$ , **quels que soient des estimateurs**  $(\hat{\theta}_1, \dots, \hat{\theta}_n)$  des vrais paramètres de recalage  $(\theta_1^*, \dots, \theta_n^*)$ , on a

$$\mathbb{E} \left( \frac{1}{n} \sum_{j=1}^n (\hat{\theta}_j - \theta_j^*)^2 \right) \geq \frac{\epsilon^2}{\sum_{k \in \mathbb{Z}} (2\pi k)^2 |c_k|^2 + \epsilon^2 \int_{\Theta} \left( \frac{\partial}{\partial \theta} \log g(\theta) \right)^2 g(\theta) d\theta} > 0.$$

**Ultime bug :**

**Theorem (BGKM '12, BC'11)**

Sous  $(H_f)$  et  $(H_g)$ , **quels que soient des estimateurs**  $(\hat{\theta}_1, \dots, \hat{\theta}_n)$  des vrais paramètres de recalage  $(\theta_1^*, \dots, \theta_n^*)$ , on a

$$\mathbb{E} \left\| \frac{1}{n} \sum_{j=1}^n f(\cdot - \hat{\theta}_j + \theta_j^*) - f \right\|_2^2 \geq C \epsilon^2 \frac{c_1(f)^2}{\|f'\|_2 + \epsilon^2 I(g)} > 0.$$

## Estimation des paramètres de recalage dans le modèle de shift (3)

**Cadre :**  $Y_j(t) = f(t - \theta_j^*) + \epsilon W_j(t)$  , bruit blanc et  $\theta_j$  *i.i.d.*  $\sim g$ .

### Conclusions :

- ▶ L'estimation des recalages en vue d'obtenir une estimation de  $f$  n'est possible que si  $\epsilon \mapsto 0$ .
- ▶ Ce genre de résultats a déjà été obtenu par BLV, PTRF '10 lorsqu'on augmente la fréquence d'échantillonnage de la courbes par exemple.
- ▶ Lorsque  $f$  est moins régulière, la borne inférieure devient plus permissive.
- ▶ La démonstration exploite astucieusement l'inégalité de van Tree (inégalité de Cramer-Rao "bayésienne").
- ▶ Cela condamne implicitement la réussite de méthode Procruste de minimisation alternée pour approcher la moyenne de Fréchet.
- ▶ Il est important dans ce genre de modèle de supprimer le bruit des données **avant** la phase de recalage.
- ▶ Ces résultats pourraient sans doute être étendus à des modèles plus généraux.

## Approche par déconvolution (1)

**Cadre :**  $n$  réalisations du modèle  $Y_1, \dots, Y_n$  :

$$Y_m = \tau_m \cdot f + \epsilon W_m(x), \quad m = 1, \dots, n, \text{ et } \tau_m \text{ i.i.d. } \sim g$$

On suppose par ailleurs que  $g$  est connue.

**Objectif :** estimer  $f$  ainsi qu'une borne inf et sup (lorsque  $n \rightarrow +\infty$ ) du risque minimax dans la situation des courbes shiftées

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f), \quad \text{où}$$

- ▶  $\mathcal{R}(\hat{f}_n, f) = \mathbb{E} \|\hat{f}_n - f\|^2 = \mathbb{E} \int_0^1 |\hat{f}_n(x) - f(x)|^2 dx$
- ▶  $\mathcal{F} \subset L^2([0, 1])$  boule de Sobolev
- ▶  $\hat{f}_n$  fonction mesurable du processus  $\{Y_m, m = 1, \dots, n\}$

**Comparaison au modèle sans déformation :** Si  $\mathcal{F} = H^s(A)$  (boule de Sobolev de rayon  $A$ ) de régularité  $s$  ("nombre de dérivées") et si les  $\tau_m$  sont nuls alors

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f) \sim Cn^{-\frac{2s}{2s+1}}$$

Pour l'étude du modèle lorsque  $\tau$  appartient à un groupe de Lie compact  
BCG, IEEE IT '12

## Approche par déconvolution (2)

**Déconvolution ?** L'espérance ponctuelle de chaque courbe :

$$\mathbb{E} [f(x - \tau_m)] = \int_{\mathbb{R}} f(x - \tau)g(\tau)d\tau = f \star g(x)$$

La moyenne des  $n$  courbes vérifie

$$dY(x) = f \star g(x)dx + \underbrace{\xi(x)dx}_{\text{Erreur non Gaussienne}} + \underbrace{\frac{\epsilon}{\sqrt{n}}dW(x)}_{\text{Erreur Gaussienne}}, \quad x \in [0, 1].$$

**Cas de déconvolution classique avec erreur Gaussienne :**

$$dY(x) = f \star g(x)dx + \frac{\epsilon}{\sqrt{n}}dW(x) \quad x \in [0, 1],$$

**Vitesse de convergence minimax :** si  $\gamma_\ell = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x}g(x)dx$  et si

$$\exists \nu > 0 \quad C_{\min}|\ell|^{-\nu} \leq |\gamma_\ell| \leq C_{\max}|\ell|^{-\nu}.$$

alors sur  $\mathcal{F} = H^s(A)$  :

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f) \sim Cn^{-\frac{2s}{2s+2\nu+1}} \quad (\text{au lieu de } n^{-\frac{2s}{2s+1}} \text{ dans le cas direct})$$

## Approche par déconvolution (3)

**Notre cadre :** Si  $\ell \in \mathbb{Z}$ , on pose  $\theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x) dx$  et  $c_{m,\ell} = \int_0^1 e^{-2i\ell\pi x} dY_m(x)$ . Alors

$$c_{m,\ell} = \theta_\ell e^{-i2\pi\ell\tau_m} + \epsilon_m z_{\ell,m} \text{ où } z_{\ell,m} \sim_{i.i.d.} N_{\mathbb{C}}(0, 1)$$

La moyenne des  $n$  coefficients de Fourier satisfait

$$\tilde{c}_\ell = \frac{1}{n} \sum_{m=1}^n c_{\ell,m} = \theta_\ell \gamma_\ell + \underbrace{\xi_\ell}_{\text{Erreur non Gaussienne}} + \underbrace{\frac{\epsilon}{\sqrt{n}} \eta_\ell}_{\text{Erreur Gaussienne classique}}, \text{ avec } \eta_\ell \sim_{i.i.d.} N_{\mathbb{C}}(0, 1)$$

**Estimation non adaptative :** On estime  $\theta_\ell$  par

$$\hat{\theta}_\ell = \frac{\tilde{c}_\ell}{\gamma_\ell} = \theta_\ell + \frac{\xi_\ell}{\gamma_\ell} + \frac{\epsilon}{\sqrt{n}} \frac{\eta_\ell}{\gamma_\ell}.$$

Connaissant la nature de la décroissance des coefficients de Fourier de  $f$ , on utilise un seuillage haute fréquence :

$$\hat{f}_{n,M}(x) = \sum_{|\ell| \leq M} \hat{\theta}_\ell e^{-i2\pi\ell x}$$

## Approche par déconvolution (4)

Précisément, la boule de Sobolev de rayon  $A$  est donnée par :

$$H_s(A) = \left\{ f \in L^2([0, 1]) ; \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s}) |\theta_\ell|^2 \leq A, \right\} \text{ avec } A > 0, s > 0$$

**Proposition (BG, A.O.S. '10)**

Si  $M = M_{n,s} \sim n^{\frac{1}{2s+2\nu+1}}$ , alors  $\sup_{f \in H_s(A)} \mathcal{R}(\hat{f}_{n,M_{n,s}}, f) = \mathcal{O}(n^{-\frac{2s}{2s+2\nu+1}})$

**Problème :**

- ▶  $\hat{f}_{n,M_{n,s}}$  dépend de la régularité  $s$  (estimateur non adaptatif)
- ▶ On peut utiliser des ondelettes de Meyer pour obtenir un estimateur "adaptatif" presque aussi bon.

**Question :** L'apparition de ce problème de déconvolution inverse est-il inévitable? (Rôle de  $g$  incontournable)

## Vitesse Minimax

**Réponse : Oui !**

**Theorem (BG, A.O.S. '10)**

*Si  $s > \nu + 1/2$  and  $\nu > 1/2$ , alors il existe une constante  $C(A, s) > 0$  qui dépend uniquement de  $A, s$  telle que*

$$\lim_{n \rightarrow +\infty} n^{\frac{2s}{2s+2\nu+1}} \mathcal{R}_n(H_s(A)) \geq C(A, s)$$

**Outil :** Pour démontrer ce résultat, il s'agit de définir un rapport de vraisemblance entre deux hypothèses par le biais de la formule de Girsanov

- ▶ Hypothèse  $H_f$  :  $Y_m = \tau_m \cdot f + \epsilon W_m$ ,  $m = 1, \dots, n$
- ▶ Hypothèse  $H_{\tilde{f}}$  :  $Y_m(x) = \tau_m \cdot \tilde{f} + \epsilon W_m$ ,  $m = 1, \dots, n$

Puis déduire le rapport de vraisemblance entre deux modèles :

$$\Lambda(P_f, P_{\tilde{f}})(Y) = \frac{\int_0^1 e^{\langle f_{\theta}^{-\alpha}, dY \rangle - \|f_{\theta}\|^2/2} g(\alpha) d\alpha}{\int_0^1 e^{\langle \tilde{f}_{\theta'}^{-\alpha}, dY \rangle - \|\tilde{f}_{\theta'}\|^2/2} g(\alpha) d\alpha}$$

Et enfin finir en utilisant des techniques classiques de cube d'Assouad ou Lemme de Fano (Calculs très pénibles).



## Exemple numérique (déconvolution)

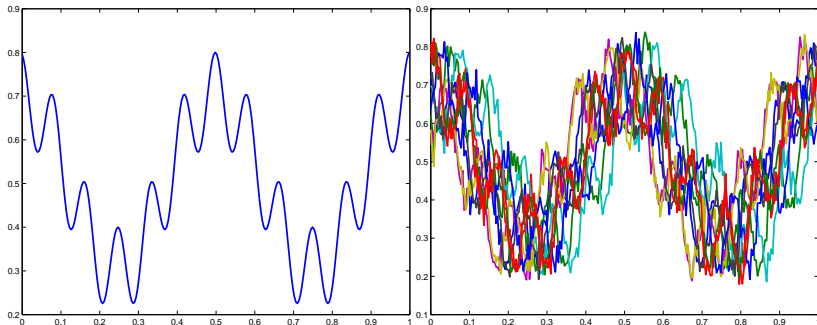
$G = \mathbb{S}^1$  - **group des translations**

**Données** : courbes décalées aléatoirement  $f \in L^2_{per}([0, 1])$  + bruit blanc gaussien

$Y_j(t) = f(t - \theta_j) + W_j(t)$ , pour  $t \in [0, 1]$ , et  $\theta_j$  i.i.d. loi de Laplace,

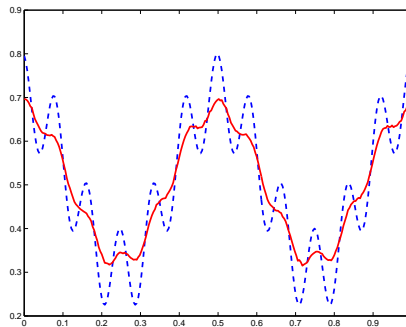
$$g(x) = \frac{1}{\sqrt{2\sigma}} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right), \gamma_\ell = \frac{1}{1+2\sigma^2\pi^2\ell^2}, \nu = 2$$

Courbe  $f$  / Échantillon de 10 courbes parmi  $n = 200$

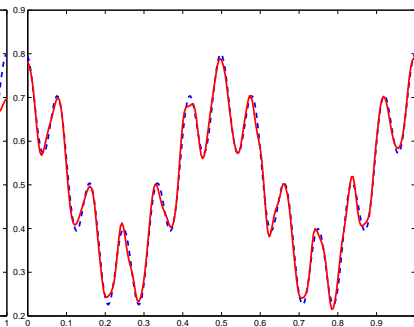


# Exemple numérique (déconvolution)

## Moyenne euclidienne

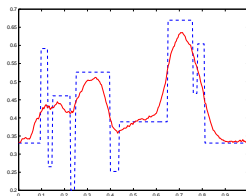


## Déconvolution avec $g$ connue

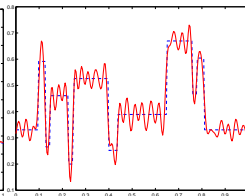


# Exemple numérique (déconvolution)

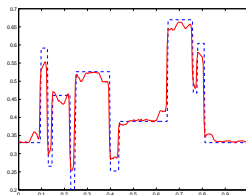
**Moyenne euclidienne**



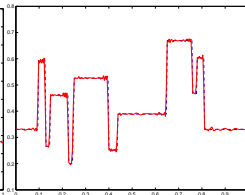
**Déconvolution avec  $g$  connue**



**Moyenne Procruste**



**Moyenne de Fréchet**



## Bilan sur l'approche déconvolution

**Cadre :** Dans la situation où le modèle s'écrit

$$Y_j = \tau_j \cdot f + W_j, \text{ avec } (\tau_j) i.i.d. \sim g \text{ et } (W_j)_j \text{ bruit blanc gaussien,}$$

et si les  $\tau$  sont à valeurs dans un groupe de Lie compact (semi-simple, etc.), les résultats précédents restent vrais. (Cadre un peu plus général de transformations).

**Conclusion :**

- ▶ La régularité de la loi de déformation est fondamentale et dégrade la qualité d'estimation.
- ▶ La reconstruction est optimale d'un point de vue fréquentiste.
- ▶ On peut étendre à des bruits un peu différents ces résultats.
- ▶ On pourrait étudier le cas particulier de la reconstruction par transformée de Radon déformée par des transformations géométriques "rigides".

**Problème :** L'approche peut ne pas être réaliste en pratique car la densité  $g$  des déformations shifts est inconnue dans certaines applications. La moyenne de Fréchet permet de s'absoudre de cette connaissance mais il n'existe pas (encore) d'outils mathématiques efficaces permettant une étude satisfaisante du problème (!)

**Question :** Peut-on reconstruire  $f$  dans le cas où  $g$  est inconnue ?

## Approche Bayésienne : ultime réponse ?

**Cadre :** AAT J.R.S.S.(B) '07, AKT Bernoulli '07 proposent une approche bayésienne pour l'estimation dans le modèle

$$Y = g \cdot f_0 + \epsilon W,$$

où  $g \sim h_0$  avec  $f$  et  $h_0$  inconnus appartenant à une famille paramétrique.

**Méthode :** Poser une loi *a priori*  $\pi_1$  sur  $f$  et  $\pi_2$  sur  $h$ , puis utiliser la loi *a posteriori* pour fabriquer des estimations de  $f_0$  et  $h_0$ .

$$\pi := \pi_1 \otimes \pi_2 \text{ et } \pi_n := \pi_1 \otimes \pi_2 \left[ \cdot \mid Y_1^n \right]$$

**Question 1 :** Lorsque  $n \mapsto +\infty$ , a-t-on  $\pi_n \mapsto \delta_{f_0, g_0}$  dans la situation paramétrique ?

**Question 2 :** Peut-on sortir du cadre paramétrique et quelles en sont les conséquences ?

**Question 3 :** Comment procéder au calcul de l'*a posteriori*, et des estimateurs bayésiens qui en dépendent ?

## Approche Bayésienne : rappels

- ▶  $f$  et  $g$  sont paramétrées par  $\theta \in \Theta \subset \mathbb{R}^d$ , espace des paramètres.
- ▶ On suppose que les observations  $(Y_j)_{j=1..n}$  sont issues d'un  $\theta_0$  inconnu, ici composé de  $\theta_0 = (\theta_f^0, \theta_g^0)$  et l'estimation *a posteriori* est :

$$\hat{\theta}_n = \mathbb{E}[u | Y_1, \dots, Y_n]$$

### ▶ Theorem (Le Cam, Schwartz (1965))

Si  $\pi$  est un *a priori* tel que

$$\forall \epsilon > 0 \quad \pi(B_{KL}(\theta_0, \epsilon)) > 0,$$

alors

$$\pi_n(\theta | KL(P_\theta, P_{\theta_0}) > \epsilon_n) \longrightarrow 0 \text{ lorsque } n \longrightarrow +\infty.$$

**Attention :** C'est un résultat de consistance sur les lois de probabilités et non sur les paramètres eux-mêmes.

### ▶ Theorem (Has'minskii & Ibragimov, 1979)

Si  $\pi$  est un *a priori* absolument continu par rapport à la mesure de Lebesgue sur les paramètres  $\Theta$  et de densité  $\pi$  continue et  $\pi(\theta_0) > 0$ , alors sous deux conditions de séparabilités au sens de Hellinger des lois  $P_\theta$ , on a

$$n^\alpha(\hat{\theta}_n - \theta_0) \longrightarrow \theta_0 \text{ où } \alpha \text{ dépend de la séparabilité.}$$

## Approche Bayésienne : ultime réponse ?

AAT J.R.S.S.(B) '07, AKT Bernoulli '07 proposent une approche bayésienne pour l'estimation dans le modèle

$$Y = g \cdot f_0 + \epsilon W,$$

où  $g \sim h_0$  avec  $f$  et  $h_0$  inconnus appartenant à une famille paramétrique.

**Question 1** : Lorsque  $n \mapsto +\infty$ , a-t-on  $\pi_n \mapsto \delta_{f_0, g_0}$  dans la situation paramétrique ?

**Méthode** :  $\pi_1$  et  $\pi_2$  lois uniformes (par exemple) sur  $\Theta_1 \otimes \Theta_2$  avec  $\pi = \pi_1 \otimes \pi_2$ . On a alors le résultat suivant

**Theorem (Consistance de l'estimation bayésienne, preprint '12)**

*Si le modèle est identifiable (l'application  $\theta \mapsto \mathbb{P}_\theta$  est injective) et différentiable en moyenne quadratique, alors*

$$\pi_n(\theta | KL(\mathbb{P}_\theta, \mathbb{P}_{\theta_0}) > Cn^{-1/2}) \mapsto 0 \text{ lorsque } n \mapsto +\infty.$$

## Approche Bayésienne : ultime réponse ?

**Question 2 :** Peut-on sortir du cadre paramétrique et quelles en sont les conséquences ?

**Cadre :** Retour au modèle de shifts 1D

$$Y(t) = f_0(t - \tau) + \epsilon W_t, \forall t \in [0, 1]$$

où  $\tau \sim g_0$  avec  $f_0$  et  $g_0$  inconnues appartiennent à une famille non-paramétrique.

### Proposition (Identifiabilité, preprint '12)

Si  $f$  est supposée telle que  $c_1(f) > 0$  et  $g$  satisfait l'hypothèse de problème inverse

$$C^{-1} \ell^{-\nu} < |c_\ell(g)| < C \ell^{-\nu},$$

alors le modèle est identifiable.

### Comportement de l'a posteriori :(preprint,'12)

$f \in H_s(A)$ ,  $g$  mesure de probabilités dans  $H_\nu(B)$ , si l'a priori est construit en utilisant une distribution de type Laplace sur le nombre de coefficients de Fourier de  $f$ , puis avec une loi gaussienne sur chaque coefficient "allumé" (de même sur  $g$ ), convergence de l'a posteriori non paramétrique.

**Outils :** Récents travaux de Ghosal, Gosh, van der Vaart, AOS '00,'07 sur l'estimation bayésienne non paramétrique : description métrique de  $H_s(A)$  et géométrie des lois de mélanges.



## Approche Bayésienne : ultime réponse ?

- ▶ **Question 3** : Comment procéder au calcul de l'*a posteriori*, et des estimateurs bayésiens qui en dépendent ?
- ▶ Travaux AAT, AKT en paramétrique avec des algorithmes stochastiques.
- ▶ Extension possible en non paramétrique.

## Introduction

Motivations

Modèle statistique déformable

## Méthodes d'estimation de forme moyenne

Données

Définition d'une distance au travers de l'action de groupe

Estimation par recalage + moyenne

Estimation par Moyenne de Fréchet

## Résultats de statistiques mathématiques

$M$  estimation

Sur les procédés de recalage

Approche par déconvolution

Estimation Bayésienne

## Conclusion

## Conclusion

- ▶ Les modèles déformables sont assez délicats à traiter statistiquement.
- ▶ Il est important de supprimer le bruit des observations avant la mise en oeuvre d'une procédure de recalage.
- ▶ Le problème est de nature problème inverse avec opérateur inconnu ou partiellement inconnu.
- ▶ La régularité de l'opération de déformation joue un rôle important dans les vitesses d'estimation atteignables.
- ▶ Peu de résultats connus sur les moyennes de Fréchet dans le contexte non paramétrique.
- ▶ Vitesses inconnues en statistiques bayésiennes (travail en cours).

# Bibliographie

- [1] Allasonnière, S. & Kuhn, E. & Trouvé, A. (2010) Construction of Bayesian deformable models via stochastic approximation algorithm : A convergence study. *Bernoulli Journal*, 16 (3), p.641-678.
- [2] Bhattacharya, R. & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds (i). *Annals of statistics*, 31(1) :1–29.
- [3] Bhattacharya, R. & Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds (ii). *Annals of statistics*, 33 :1225–1259.
- [4] Bigot, J. & Gadat, S. (2010) A deconvolution approach to estimation of a common shape in a shifted curves model, *Annals of Statistics*, 38(4), 2422-2464.
- [5] Bigot, J. & Charlier B. (2011) On the consistency of Fréchet means in deformable models for curve and image analysis, *Electronic Journal of Statistics*, 5, 1054-1089.
- [6] Bigot, J., Christophe C. & Gadat, S. (2012) Random action of compact Lie groups and minimax estimation of a mean pattern, *IEEE Transactions on Information Theory*, 58(5), to be published.
- [7] Gadat, S. (2012) Bayesian consistency in the non parametric randomly shifted curves model, preprint.