Shape Invariant Model: problème inverse et méthodes bayésiennes

S. Gadat

Institut de Mathématiques de Toulouse
Université Paul Sabatier

Séminaire du GREMAQ, 23 Octobre 2012

I - 1 Motivations

Questions pour le statisticien : Comparer des objets qui partagent des caractéristiques de formes 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 $Y_m, m = 1, \dots, n$ i.i.d. La moyenne empirique est

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

Puis, l'ACP consiste à diagonaliser la covariance empirique

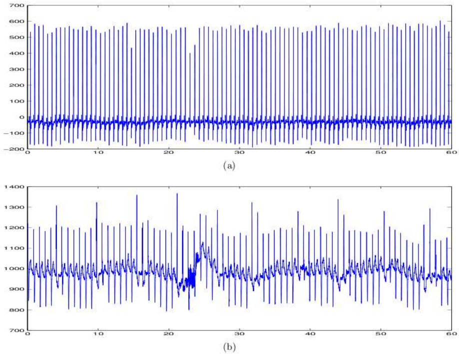
$$S = \frac{1}{n}\sum_{m=1}^{n}(Y_m - \bar{Y}^n)(Y_m - \bar{Y}^n)',$$

et choisir les premiers v. propres comme modes de variations principaux.

I - 1 Motivations : Applications en biologie

Problématique : On s'intéresse à des données qui partagent des caractéristiques de forme commune.

ECG de deux patients Normal / Arythmie



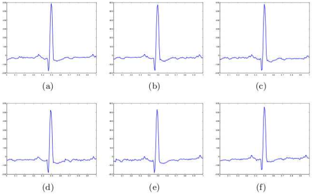
(a)

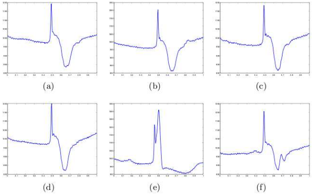
(b)

Source : *J. Bigot 2012, Preprint, Fréchet means of curves for signal averaging and application to ECG data analysis*

I - 1 Motivations : Applications en biologie

Données ECG ”Zoomées”

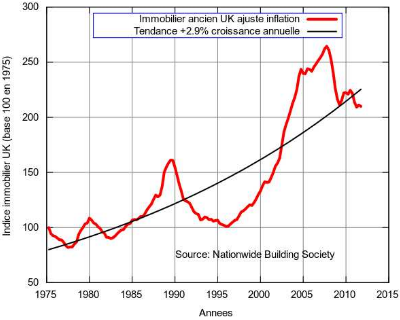


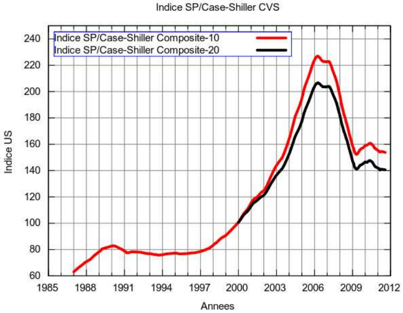


Gauche : Cycles Normaux / Droite : Cycles Arythmiques
Source : *J. Bigot 2012, Preprint, Fréchet means of curves for signal averaging and application to ECG data analysis*
Chacune des classes de signaux semble posséder une forme caractéristique qui la distingue de l’autre classe.

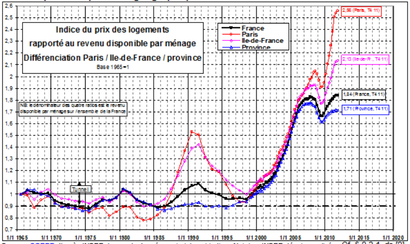
I - 1 Motivations : Applications en économie

Bulles spéculatives





2.4. Comparaison par zone géographique

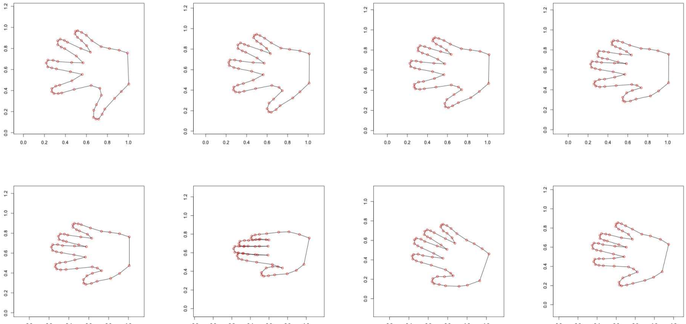


Source : CGEDD d'après INSEE, bases de données notariales et indices Notaires-INSEE désaisonnalisés. Cf. § 2.3.1. de [2].

I - 1 Motivations : Applications en traitement de l'image

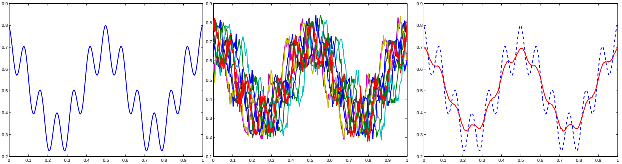
Données : Images de $N \times N$ pixels, i.e. $\mathcal{H} = \mathbb{R}^{N \times N}$

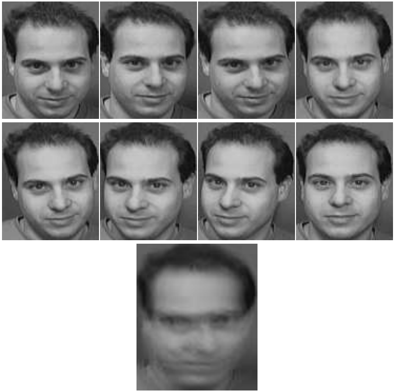
Données : formes planes de mains pour $k = 13$ landmarks (points rouges)



I - 1 Exemples de moyenne empirique

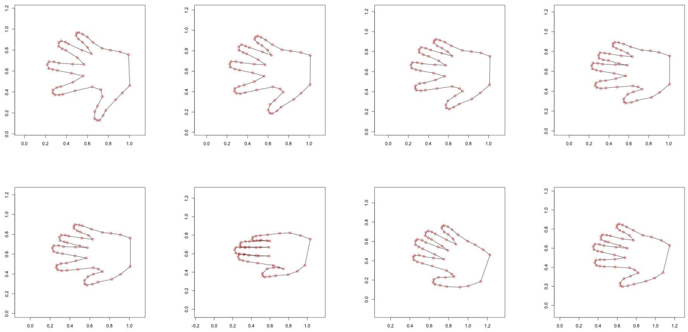
Courbes / Images





I - 1 Un exemple pour l'analyse de formes

Données : formes planes de mains pour $k = 13$ landmarks (points rouges)



Problème : Approches standard non adaptées pour les données Y_m déformés puisque $\bar{Y}^n$ n'est pas une bonne estimation de la forme moyenne.
Objectifs idéaux : Proposer un estimateur "consistant" de la forme moyenne d'un ensemble de données et estimer les modes de variations.

Dans cet exposé : Décrire ce qui est possible ou non d'un point de vue statistique, dans des cadres plus ou moins limitatifs pour la forme moyenne.

I - 2 Modèle Statistique

Idée : les n signaux sont modélisés comme la déformation d'une forme commune par l'action d'un groupe de transformation.

Modèle déformable : l'observation $Y_m : \Omega \to \mathbb{R},\ m = 1, \ldots, n$ vérifie

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

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

Dans cet exposé : Comment estimer $\mathbf{f}$ et $\phi_\mathbf{m}$ lorsque $n \to +\infty$?

On supposera par la suite qu'il n'y a pas de variabilité en intensité ($Z = 0$).

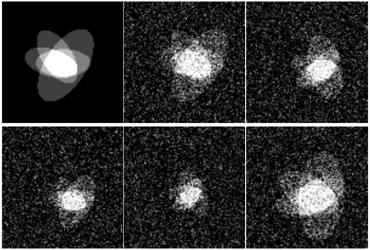
I - 2 Différents modèles de déformations

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”



Déformations élastiques : Un autre monde...

I - 3 Question posée ?

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

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

- $f : [0,1] \to \mathbb{R}$ est 1-périodique
- les τ_m sont des shifts aléatoires i.i.d. de loi de densité g
- les W_m sont des bruits additifs centrés indépendants des shifts

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

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

La suite de cet exposé se cantonera à des estimations dans le modèle de shift aléatoire lorsque $n \longmapsto +\infty$ pour f, $(\tau_m)_{m=1\dots n}$ et g.

II - 1 Idée naïve : Estimation par recalage + moyenne

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

$$\bar{Y}^n = \frac{1}{n}\sum_{j=1}^{n} Y_j \cdot \hat{g}_j^{-1} = \frac{1}{n}\sum_{j=1}^{n} Y_j(x + \hat{\tau}_j).$$

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

II - 2 Définition de la distance adaptée

Quelle distance mettre sur les objets observés ?

- **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...

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

- **Distance associée au Groupe de déformation :** si $G = [0, 1]$ agit sur $f \in L^2_{per}([0, 1])$ au travers de $g_\theta \cdot f(t) = f(t - \theta)$, on définit d_G :

$$d_G^2(Y, h) := \inf_{g \in G} d_E^2(Y, g \cdot h) = \inf_{\tau \in [0,1]} \int_0^1 |Y(x) - h(x - \tau)|^2 dx$$

La moyenne au sens de Fréchet pour d_G est $\tilde{Y}^n \in \arg\min_{h \in \mathcal{H}} \sum_j d_G^2(Y_j, h)$.

II - 3 Estimation par Moyenne de Fréchet

Moyenne de Fréchet empirique $\tilde{Y}^n \in \arg\min_{y \in \mathcal{M}} \frac{1}{n} \sum_{m=1}^n d_G^2(y, Y_m)$.
Moyenne de Fréchet théorique $Y^F := \arg\min_{y \in \mathcal{M}} \mathbb{E}_{W,g} d_G^2(y, Y_{W,g})$

- Consistance et convergence de $\tilde{Y}^n$ vers Y^F lorsque $n \to +\infty$ démontrée dans le cas où $\mathcal{M}$ est une variété Riemannienne de **dimension finie**.
- **Problème :** Résultats non adaptés à l'étude des moyennes de Fréchet pour les courbes ou images paramétrées (dimension "infinie").

Calcul de la moyenne de Fréchet Empirique

Étape 1 - Registration/warping des données

$$(\hat{\tau}_1, \ldots, \hat{\tau}_n) = \underset{(\tau_1, \ldots, \tau_n) \in [0,1]^n}{\arg\min} \left\{ \frac{1}{n} \sum_{j=1}^n \int_\Omega \left| Y_j(t - \tau_j) - \frac{1}{n} \sum_{m=1}^n Y_m(t - \tau_m) \right|^2 dt \right.$$

Étape 2 - Alignement et moyennisation des données

$$\hat{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j(. - \hat{\tau}_j)$$

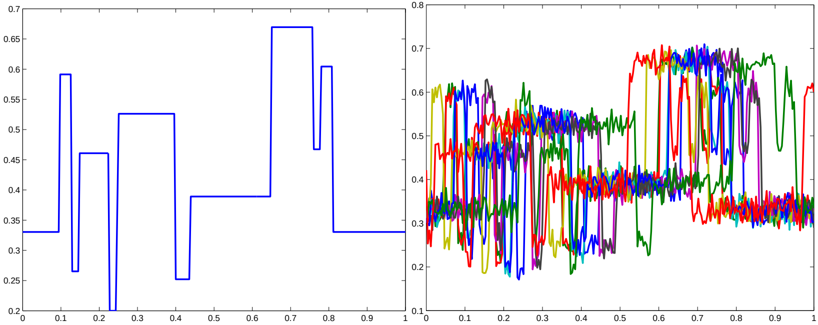
II - 4 Exemple de moyenne de courbes

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

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

$j = 1, \dots, n$

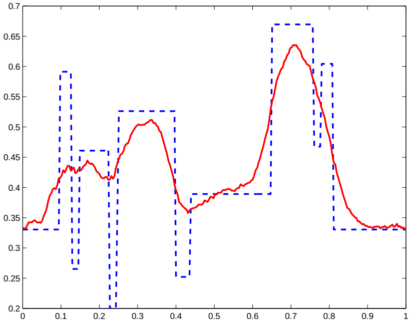
Courbe f / Échantillon de 10 parmi $n = 200$



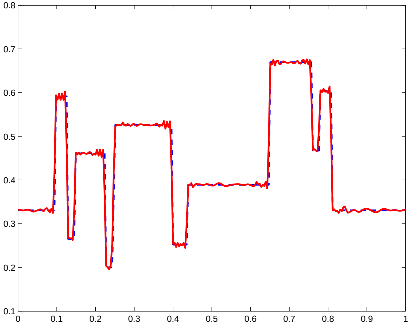
BG, AOS, '10

II - 4 Exemple de moyenne de courbes

Moyenne euclidienne



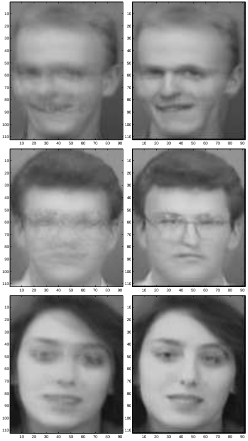
Moyenne de Fréchet



BG, AOS, ’10

II - 4 Exemple de moyenne d'images par moyenne Procruste

Moyenne euclidienne **Moyenne Procruste**



BGL, JMIV, '09

II - 5 Fiabilité des processus de recalages (1)

Cadre : $Y_j(t) = f(t - \tau_j^*) + \epsilon W_j(t)$, bruit blanc et τ_j^* $i.i.d. \sim g$.
Le problème est clairement non identifiable en l'état, mais si on suppose

- (H_g) **:** g à support compact dans $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$, 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 en f est identifiable. Par moyenne de Fréchet, les décalages $(\hat{\tau}_j)_{j=1\ldots n}$ tels que $\sum_j \hat{\tau}_j = 0$ vérifient

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

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 peut se produire qu'uniquement lorsque ϵ tend aussi vers 0 (niveau de bruit évanescent, ou augmentation de la résolution).

II - 5 Fiabilité des processus de recalages (2)

Cadre : $Y_j(t) = f(t - \tau_j^*) + \epsilon W_j(t)$, bruit blanc et τ_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** $(\hat{\tau}_1, \ldots, \hat{\tau}_n)$ *des vrais paramètres de recalage* $(\tau_1^*, \ldots, \tau_n^*)$, *on a*

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

Ultime bug :

Theorem (BGKM '12,BC'11)

Sous (H_f) *et* (H_g), **quels que soient des estimateurs** $(\hat{\tau}_1, \ldots, \hat{\tau}_n)$ *des vrais paramètres de recalage* $(\tau_1^*, \ldots, \tau_n^*)$, *on a*

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

II - 5 Fiabilité des processus de recalages (3)

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

- L'estimation des recalages en vue d'obtenir une estimation de f n'est possible que si $\epsilon \longrightarrow 0$.
- Ce genre de résultats a déjà été obtenu par BLV, PTRF '10 lorsqu'on augmente la fréquence d'échantillonage de la courbes par exemple.
- Lorsque f est moins régulière, la borne inférieure devient plus permissive.
- Les démonstrations exploitent astucieusement l'inégalité de van Tree (inégalité de Cramer-Rao "bayésienne").
- Cela condamne implicitement la réussite des méthodes dites de recalage (Procruste, Fréchet, . . .)
- Il est important dans ce genre de modèle de supprimer le bruit des données avant la phase de recalage.

Approche par déconvolution (1)

Cadre : $Y_j(t) = f(t - \tau_j) + \epsilon W_j(t)$, bruit blanc et τ_j $i.i.d. \sim g$.
On suppose par ailleurs que g est connue.
Objectif : estimer f ainsi qu'une borne inf et sup (lorsque $n \to +\infty$) du risque minimax dans la situation des courbes shiftées

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

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

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

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

Approche par déconvolution (2)

Déconvolution ? L'espérance ponctuelle de chaque courbe :

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

La moyenne des n courbes vérifie

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

Cas de déconvolution classique avec erreur Gaussienne :

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

Vitesse de convergence minimax : si $\gamma_\ell = \int_0^1 e^{-i2\pi\ell\tau}g(\tau)d\tau$ et si

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

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

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f) \sim C_A n^{-\frac{2s}{2s+2\nu+1}}$$ (au lieu de $n^{-\frac{2s}{2s+1}}$ 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

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

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 θ_ℓ 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 :

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

III - 2 Lien avec un problème inverse (1)

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)

III - 2 Lien avec un problème inverse (2)

Influence de la régularité de g inétivable ? Oui !

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

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

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

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

Puis déduire le rapport de vraisemblance entre deux modèles :

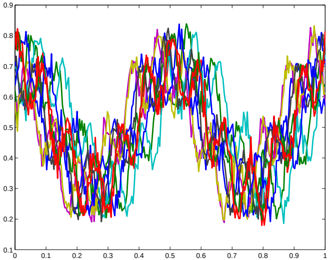
$$\Lambda(P_f, P_{\tilde{f}})(Y) = \frac{\int_0^1 e^{\langle f^{-\alpha}, dY \rangle - \|f\|^2/2})g(\alpha)d\alpha}{\int_0^1 e^{\langle \tilde{f}^{-\alpha}, dY \rangle - \|\tilde{f}\|^2/2})g(\alpha)d\alpha}$$

Et enfin finir en utilisant des techniques classiques de cube d'Assouad.

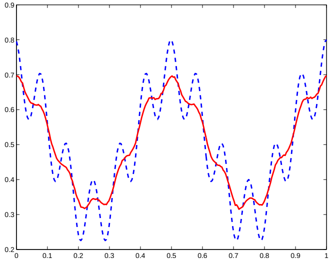
III - 3 Exemple numérique (déconvolution)

$G = \mathbb{S}^1$ **- group des translations Données :** courbes décalées

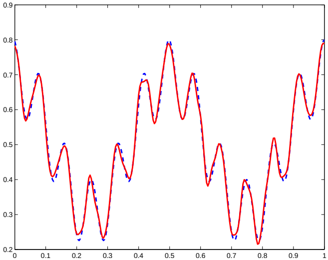
$f \in L^2_{per}([0,1])$ + bruit blanc gaussien, $g(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right)$



Moyenne euclidienne

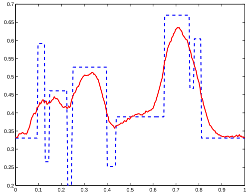


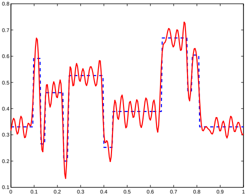
Déconvolution avec g connue



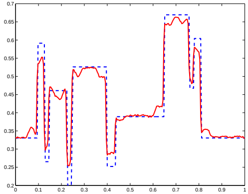
III - 3 Exemple numérique (déconvolution)

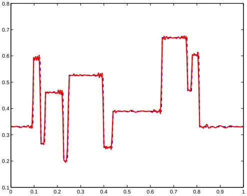
Moyenne euclidienne **Déconvolution avec g connue**





Moyenne Procruste **Moyenne de Fréchet**





III - 4 Bilan sur l'approche déconvolution

Cadre : $Y_j(t) = f(t - \tau_j) + \epsilon W_j(t)$, bruit blanc et τ_j $i.i.d. \sim g$.
Conclusion :

- La régularité de la loi de déformation g est fondamentale et dégrade la qualité d'estimation. Vitesse en $(\log(n)/n)^{-2s/(2s+2\nu+1)}$.
- La reconstruction est asymp. optimale d'un point de vue fréquentiste.
- On peut étendre à des bruits un peu différents ces résultats.

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 (Bigot & Gendre '12 avec $\epsilon \longrightarrow 0$)
- **Question** : Peut-on reconstruire f dans le cas où g est inconnue ?

III - 5 Approche Bayésienne : ultime réponse ?

Cadre :

$$Y_j(x) = f^0(x - \tau_j) + \epsilon W_j(x),$$

où (τ_j) $i.i.d. \sim g^0$ et f^0 et g^0 inconnus appartenant à une famille paramétrique / non paramétrique.

Méthode : Poser une loi *a priori* π_1 sur f^0 et π_2 sur g^0, puis utiliser la loi *a posteriori* pour fabriquer des estimations de f^0 et g^0.

$$\pi := (\pi_1 \otimes \pi_2) \text{ et } \pi_n := \pi[\quad . \quad |Y_1^n]$$

Question 1 : Lorsque $n \longmapsto +\infty$, a-t-on $\pi_n \longmapsto \delta_{f^0, g^0}$ 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 ?

III - 5 Approche Bayésienne : rappels

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

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

- Theorem (Le Cam, Schwartz (1965))

  Si π est un a priori tel que

$$\forall \epsilon > 0 \qquad \pi\left(B_{KL}(\theta^0, \epsilon)\right) > 0,$$

  alors

$$\pi_n\left(\theta | KL(P_\theta, P_{\theta_0}) > \epsilon_n\right) \longmapsto 0 \textit{ lorsque } n \longmapsto +\infty.$$

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

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

III - 7 Approche Bayésienne : cadre non-paramétrique (1)

Dans le domaine de Fourier : si $\theta^0 = \theta^0(f) := (\theta^0_{-\infty}, \ldots \theta^0_{-1}, \theta^0_0, \theta^0_1, \ldots, \theta^0_{-\infty})$

$$c_\ell \quad = \quad \theta^0_\ell e^{-i2\pi\ell\tau} + z_\ell \text{ où } z_\ell \sim_{i.i.d.} N_{\mathbb{C}}(0,1), \qquad \tau \sim g^0,$$

soit finalement pour la fréquence ℓ le modèle de mélange

$$\mathbb{P}^\ell_{\theta^0, g^0} = \int_0^1 \gamma_{\theta^0_\ell} e^{-i2\pi\ell\tau} dg^0(\tau)$$

Classe d'identifiabilité : $\mathcal{F}_s(A) := \{f \in H_s(A) \text{ et } c_1(f) > 0\}$ et

$$\mathfrak{M}_\nu([0,1]) := \left\{g \in \mathfrak{M}([0,1]) \mid \exists \forall k \in \mathbb{Z} \quad c|k|^{-\nu} < |c_k(g)| < C|k|^{-\nu}\right\}$$

Proposition (Bontemps & G. '12)

Si $(f, g) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])$, alors le modèle est identifiable.

Idée de la preuve :

- Établir que $d_{VT}(\mathbb{P}^1_{f,g}, \mathbb{P}^1_{\tilde{f},\tilde{g}}) > 0 \Rightarrow c_1(f) = c_1(\tilde{f})$.
- En déduire que nécessairement $g = \tilde{g}$ par méthode Hilbertienne (Parseval) ou transformée de Laplace.
- Conclure sur les autres coefficients de f en utilisant les autres marginales de $\mathbb{P}_{f,g}$.

III - 7 Approche Bayésienne : cadre non-paramétrique (2)

Dans le cadre non paramétrique, il est nécessaire de procéder à des arguments de recouvrements de $\mathbb{P}_{f,g}$, $(f,g) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])$.

Theorem (Bontemps & G. '12)

Pour $\epsilon > 0$ *assez petit et si* ℓ_ϵ *est tel que* $\log \frac{1}{\epsilon} \lesssim l_\epsilon$*, et* $A \lesssim \log \frac{1}{\epsilon}$ *alors*

$$\log N(\epsilon, \mathcal{F}_s^{\ell_\epsilon}(A) \times \mathfrak{M}_\nu([0,1]), d_{VT}) \lesssim l_\epsilon^2 \log \frac{1}{\epsilon}.$$

- Résultat un peu rébarbatif . . . mais très important !
- Application de recouvrements explicites pour les lois de mélanges de gaussiennes utilisant la régularité des lois Gaussiennes (GGvdV '00).
- Par rapport au contexte paramétrique en f, et non paramétrique en g, on "perd" un terme ℓ_ϵ : $\ell_\epsilon^2 \log \frac{1}{\epsilon}$ au lieu de $\ell_\epsilon \log \frac{1}{\epsilon}$

Comportement de l'*a posteriori* : En suivant les théorèmes "boite noire" de GGvdV'00, il faut :

- Majorer la complexité du modèle :

$$\log N(\epsilon_n, \mathcal{F}_s^{\ell_n}(A) \times \mathfrak{M}_\nu([0,1]), d_{VT}) \lesssim n\epsilon_n^2$$

- Minorer le poids de l'*a priori* pour les lois proches des $\mathbb{P}_{f^0,g^0}$.

III - 7 Approche Bayésienne : cadre non-paramétrique (3)

Construction de l'*a priori* π :

- On choisit un entier ℓ_{max} suivant une loi vérifiant $p(\ell_{max} = k) \propto e^{-k^2 \log k}$.
- Chaque coefficient "allumé" suit une loi Gaussienne $\mathcal{N}_{\mathbb{C}}(0, 1)$.
- Processus de Dirichlet $D(\alpha)$ comme *a priori* sur g.

Theorem (Consistance Bayésienne, Bontemps & Gadat '12)

Si $f^0 \in \mathcal{F}_s(A)$ *avec* $s \geq 1$*, alors pour* $\epsilon_n = \left(\frac{\log^{5/2} n}{n}\right)^{2s/(2s+2)}$ *:*

$$\pi_n \left\{\mathbb{P}_{f,g} \quad s.t.\, d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n\right\} = 1 + O_p(1)$$

Commentaires :

- On obtient une vitesse polynomiale !
- On paye le $\ell_\epsilon^2 \log \frac{1}{\epsilon}$ en obtenant une vitesse $2s/(2s+2)$ au lieu de $2s/(2s+1)$, il faut estimer f^0 et g^0.

III - 7 Approche Bayésienne : cadre non-paramétrique (4)

Revenir aux objets sur $\mathcal{F}_s(A)$ ou $\mathfrak{M}_\nu([0,1])$?

- Implicitement, cela pose implicitement la question d'identifiabilité.
- On peut 'facilement' déduire une vitesse polynomiale sur le premier coefficient de Fourier θ_1^0 du théorème précédent.
- Pour g, il faut utiliser une minoration du type

$$d_{VT}(\mathbb{P}_{f,g}, \mathbb{P}_{f,g^0}) \geq C(\theta_1^0) \sum_{n \in \mathbb{Z}} |c_n(g - g^0)|^2 \omega_n^2,$$

  où ω_n est à relier à la fonction de Bessel modifiée (densité d'une CTRW). Malheureusement, $\omega_n^2 \sim e^{-n \log n}$!
- Très forte dégénerescence du problème inverse.

Theorem (Consistance Bayésienne, Bontemps & Gadat '12)

Si $(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])$ *inconnus, alors*

$$\Pi_n\left(g : \|g - g^0\|_2^2 > M \log^{-2\nu}(n) | Y_1, \ldots, Y_n\right) \mapsto 0$$

$$\Pi_n\left(f : \|f - f^0\|_2^2 > M (\log n)^{-2s \times \frac{2\nu}{2s+2\nu+1}} |Y_1, \ldots, Y_n\right) \longrightarrow 0$$

III - 7 Approche Bayésienne : cadre non-paramétrique (4)

Optimalité des bornes précédentes ?

- Il s'agit de voir si l'injectivité de l'application $(f, g) \longmapsto \mathbb{P}_{f,g}$ est suffisament "précise".
- Dans la preuve de l'identifiabilité, on constate que les difficultés arrivent lorsque les modules des coeffs. de Fourier sont égaux : $|\theta_\ell^0| = |\theta_\ell|$.
- On construit des fonctions tests $f_j(x) = e^{i2\pi x} + \theta_j e^{i2\pi p x}$ avec $|\theta_j| = |\theta'_j|, \forall j \neq j'$ et g_j telle que $f_j \star g_j \simeq f'_j \star g'_j$.

On peut alors calibrer $p(n)$ et $\theta_j(n)$ et appliquer le lemme de Fano et obtenir

Theorem (Bontemps & Gadat '12)

Si $(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])$, on peut trouver c tel que le risque minimax sur $\mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])$ est minoré par

$$\liminf_{n \longrightarrow +\infty} (\log n)^{2s+2} \inf_{\hat{f} \in \mathcal{F}_s(A)} \sup_{(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])} \|\hat{f} - f\|_2^2 \geq c,$$

et

$$\liminf_{n \longrightarrow +\infty} (\log n)^{2\nu+1} \inf_{\hat{g} \in \mathcal{F}_s(A)} \sup_{(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0,1])} \|\hat{g} - g\|_2^2 \geq c.$$

Conclusion

- 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.
- Optimalité des vitesses en statistiques bayésiennes ?
- Extension à d'autres types de problèmes que du traitement du signal possible.
- Question ouverte pour des algorithmes efficaces de simulation de loi *a posteriori*...

Bibliographie

[1] Allassonnière, 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.

[4] Bigot, J., Gadat, S. , Klein, T. & Marteau, C. (2012) Intensity estimation of non-homogeneous Poisson processes from shifted trajectories, Preprint.

[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] Bigot, J. & Gendre, X. (2012) Minimax properties of Fréchet means of discretely sampled curves, Preprint.

[8] Bontemps, D & Gadat, S. (2012) Bayesian consistency in the non parametric randomly shifted curves model, preprint.