Statistique pour données fonctionnelles.
Chapitre 4. Statistique descriptive et exploratoire pour données fonctionnelles

Gaëlle Chagny
CNRS, Labo. de Maths. R. Salem, Univ. Rouen,

Université Paris Dauphine – Executive Master Statistique et Big data, 2020

Plan

Plan

Statistique descriptive pour données fonctionnelles - Objectifs

- **Cadre multivarié.**
  - Observations : $X_1, \ldots, X_n \subset \mathbb{R}^d$, $X_i = {}^t(X_{i,1}, \ldots, X_{i,d})$.
  - Mesures résumées
    - Moyennes empiriques :

$$X_{.j} = n^{-1} \sum_{i=1}^{n} X_{i,j}, \quad j = 1, \ldots, d.$$

    - Matrice de variance-covariance :

$$\Sigma_n = \left(n^{-1} \sum_{i=1}^{n} (X_{i,j} - X_{.j})(X_{i,j'} - X_{.j'}) \right)_{j,j'}.$$

- **Cadre fonctionnel.**
  - Observations : $X_1, \ldots, X_n \subset L^2(T)$, $X_i = \{X_i(t),\, t \in T\}$.
  - Mesures résumées : comment étendre les notions précédentes ?

Plan

Moyenne et variance (1)

Observations lissées. $X_1, \ldots, X_n \subset L^2(T)$, $X_i = \{X_i(t),\ t \in T\}$

Définition

- ***fonction moyenne empirique :***

$$\begin{array}{rcl} \bar{X}_n : T & \longrightarrow & \mathbb{R}, \\ t & \longmapsto & \bar{X}_n(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i(t). \end{array}$$

- ***fonction variance empirique :***

$$\begin{array}{rcl} Var_n : T & \longrightarrow & \mathbb{R}, \\ t & \longmapsto & Var_n(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left(X_i(t) - \bar{X}_n(t)\right)^2. \end{array}$$

On définit également l'écart-type comme étant la racine carrée de cette fonction variance.

Moyenne et variance (2) - Code R

Fonctions du package `fda`

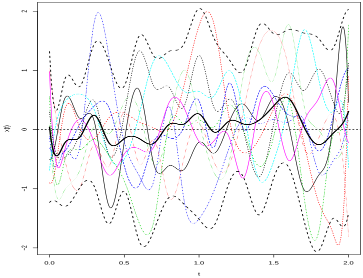
1. `mean.fd`
   - Argument : objet fonctionnel (éventuellement multidimensionnel) comportant plusieurs courbes de la classe `fd`
   - Sortie : objet de la classe `fd` contenant la (les) fonction(s) moyenne de l'objet (ou des objets) en paramètres.

   appliquée à un objet fonctionnel comportant plusieurs courbes (éventuellement multidimensionnel) de la classe `fd`, elle renvoie un objet de la classe `fd`.
2. `sd.fd` ou `std.fd` : de manière similaire à la fonction précédente, retourne les fonctions écart-type des données en paramètres.

Moyenne et variance (3) - Exemple

Exemple. $X_i = \sum_{j=1}^{N} \xi_{i,j}\psi_j$, $i = 1, \dots, 10$, avec

- $\xi_{i,j}$ *i.i.d.* $\mathcal{N}(0,1)$
- $(\psi_j)_j$ base de splines cubiques à 21 noeuds équirépartis sur $[0;2]$.



— — $t \mapsto X_i(t)$

— $t \mapsto \bar{X}_n(t)$,

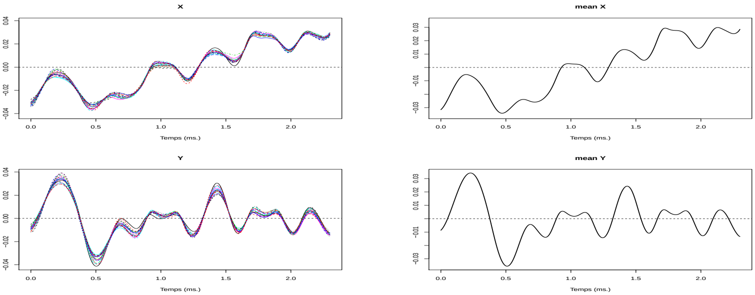
... $t \mapsto \bar{X}_n(t) \pm 2\sqrt{\mathrm{Var}_n(t)}$

Code R.

```
extr=c(0,2)
pts_rupture = seq(0,2,length.out=21)
ordre = 4
nb_fctBase =length(pts_rupture)+ordre-2
fct_base = create.bspline.basis(extr,
     norder=ordre ,breaks=pts_rupture)
coeff_ech = matrix(rnorm(10*nb_fctBase),
                    nbre_fctBase,10)
objetfd2 = fd(coeff_ech,fct_base)

sfd2 = std.fd(objetfd2)
mfd2 = mean(objetfd2)

plot(objetfd2,xlab="t",ylab="x(t)")
lines(mfd2,lwd=3)
lines(mfd2+2*sfd2,lwd=2,lty=2)
lines(mfd2-2*sfd2,lwd=2,lty=2)
```

Moyenne et variance (4) - Exemple

Exemple. Données `Handwrit`



Données Fonctions moyennes

Code R.

```
tps_handwrit  <- seq(0, 2.3, len=1401)
pts_rupture_handwrit = seq(0,2.3,length.out=51)
Fct_base_handwrit = create.bspline.basis(c(0,2.3),norder=6,breaks=pts_rupture_handwrit)
Donnees_lissees_handwrit = smooth.basis(tps_handwrit,handwrit,Fct_base_handwrit)
par(mfrow=c(2,1))
plot(Donnees_lissees_handwrit$fd,xlab='Temps (ms.)')
moy_handwrit = mean(Donnees_lissees_handwrit$fd)
par(mfrow=c(2,1))
plot(moy_handwrit,lwd=2,xlab='Temps (ms.)')
```

Plan

Covariance et corrélation (1)

Observations lissées. $X_1, \ldots, X_n \subset L^2(T)$, $X_i = \{X_i(t),\ t \in T\}$

Définition

- ***fonction de covariance empirique :***

$$\begin{aligned} C_n \ : \ T^2 \quad &\longrightarrow \quad \mathbb{R}, \\ (t_1, t_2) \quad &\longmapsto \quad C_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \big(X_i(t_1) - \bar{X}_n(t_1)\big)\big(X_i(t_2) - \bar{X}_n(t_2)\big). \end{aligned}$$

- ***fonction de corrélation :***

$$\begin{aligned} Corr_n \ : \ T^2 \quad &\longrightarrow \quad \mathbb{R}, \\ (t_1, t_2) \quad &\longmapsto \quad Corr_n(t_1, t_2) = \frac{C_n(t_1, t_2)}{\sqrt{Var_n(t_1) Var_n(t_2)}}. \end{aligned}$$

- ***opérateur de covariance empirique*** *($(X_1, \ldots, X_n)$ supposé centré)*

$$\begin{aligned} \Gamma_n \ : \ L^2(T) \quad &\longrightarrow \quad L^2(T) \\ f \quad &\longmapsto \quad \frac{1}{n} \sum_{i=1}^{n} \langle f, X_i \rangle X_i. \end{aligned}$$

Covariance et corrélation (2)

Observations lissées. $(X_1, Y_1), \ldots, (X_n, Y_n) \subset L^2(T)$, $(X_i, Y_i) = \{(X_i(t), Y_i(t))\ t \in T\}$

Définition

- ***fonction de covariance croisée empirique :***

$$\begin{array}{rcl} Cov_{n,X,Y} \ : \ T^2 & \longrightarrow & \mathbb{R}, \\ (t_1, t_2) & \longmapsto & Cov_{n,X,Y}(t_1, t_2) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \Big(X_i(t_1) - \bar{X}_n(t_1)\Big)\Big(Y_i(t_2) - \bar{Y}_n(t_2)\Big). \end{array}$$

- ***fonction de corrélation croisée :***

$$\begin{array}{rcl} Corr_{n,X,Y} \ : \ T^2 & \longrightarrow & \mathbb{R}, \\ (t_1, t_2) & \longmapsto & Corr_{n,X,Y}(t_1, t_2) = \dfrac{Cov_{n,X,Y}(t_1, t_2)}{\sqrt{Var_X(t_1)\, Var_Y(t_2)}}. \end{array}$$

Covariance et corrélation (3) - Code R.

Fonctions du package `fda`

1. `var.fd` : retourne un ou plusieurs objets de classe `bifd`, objets fonctionnels à deux variables (2 indices de temps par exemple).

   - **cas 1 :** appliquée à $t \mapsto X_i(t)$, $i = 1, \ldots, n$, renvoie $(t_1, t_2) \mapsto C_n(t_1, t_2)$ sous forme d'un objet fonctionnel.
   - **cas 2 :** appliquée à $t \mapsto (X_i(t), Y_i(t))$, $i = 1, \ldots, n$, renvoie $(t_1, t_2) \mapsto C_{n,X}(t_1, t_2)$, $(t_1, t_2) \mapsto \mathrm{Cov}_{n,X,Y}(t_1, t_2)$ et $(t_1, t_2) \mapsto C_{n,Y}(t_1, t_2)$.

2. `cor.fd` : calcul de la corrélation empirique entre un ou deux objets fonctionnels, sous forme d'une matrice.

   - **cas 1 :** appliquée à une grille $\{t_j,\, j = 1, \ldots, d\}$ (vecteur), et $t \mapsto X_i(t)$, $i = 1, \ldots, n$, renvoie la matrice $C = (\mathrm{Corr}_{n,X}(t_j, t_k))_{1 \leq j,k \leq d}$.
   - **cas 2 :** appliquée à une grille $\{t_j,\, j = 1, \ldots, d\}$ (vecteur), $t \mapsto X_i(t)$, $i = 1, \ldots, n$, une seconde grille $\{t'_k,\, k = 1, \ldots, d'\}$ et $t \mapsto Y_i(t)$, $i = 1, \ldots, n$, renvoie la matrice $C = (\mathrm{Corr}_{n,X,Y}(t_j, t'_k))_{1 \leq j \leq d, 1 \leq k \leq d'}$.

Covariance et corrélation (4) - Code R.

Représentation de courbes en 3D :

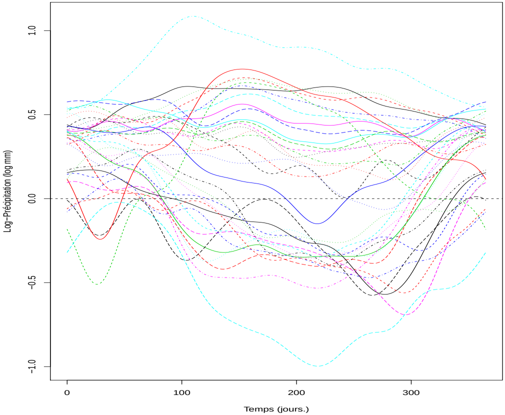
1. `eval.bifd` du package `fda` :
   - Arguments
     - une première grille $\{t_j, j = 1, \ldots, d\}$ (vecteur),
     - une seconde grille $\{t'_k, k = 1, \ldots, d'\}$,
     - un (ou plusieurs) objet(s) fonctionnel(s) de la classe `bifd`.
   - Sortie : un tableau contenant les évaluations de ces objets aux points (t_j, t'_k).

2. Outils de tracé d'une surface $(t_1, t_2) \mapsto z(t_1, t_2)$, à partir d'une grille de discrétisation pour t_1, une pour t_2, et la matrice d'évaluation de z en (t_1, t_2).

   - `persp` : tracé 3D de la surface
   - `contour` : tracé plan des lignes de niveau.
   - `filled.contour` : tracé plan des lignes de niveau colorées (la couleur dépendant du niveau).
   - `levelplot` du package `lattice` : tracé plan des lignes de niveau et couleurs entre elles.

Covariance et corrélation (5) - Exemple

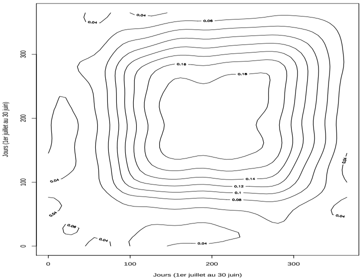
Exemple 1. Données `CanadianWeather`, log-précipitations.



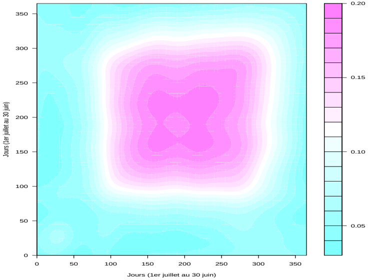
Covariance et corrélation (6) - Exemple

Exemple 1. Données `CanadianWeather`, log-précipitations.

`contour`



`filled.contour`



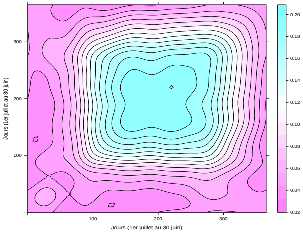
Code R.

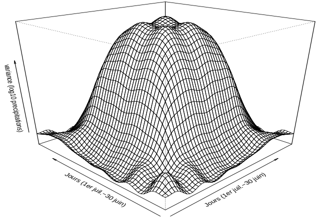
```
logprec_var <- var.fd(Donnees_lissees_Logprecip)
jours_5 <-seq(0,365,5)
logprec_var_mat<-eval.bifd(jours_5,jours_5,logprec_var)

contour(jours_5,jours_5,logprec_var_mat,
                 xlab='Jours (1er juillet au 30 juin)',ylab='Jours (1er juillet au 30 juin)')
filled.contour(jours_5,jours_5,logprec_var_mat,
                 xlab='Jours (1er juillet au 30 juin)',ylab='Jours (1er juillet au 30 juin)'
                 ,lwd=2)
```

Covariance et corrélation (7) - Exemple

Exemple 1. Données `CanadianWeather`, log-précipitations.

`levelplot` `persp`



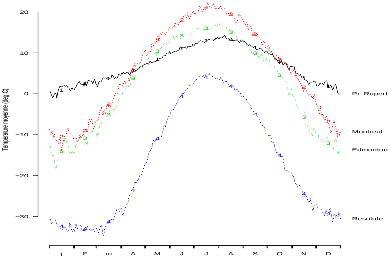


Code R.

```
persp(jours_5, jours_5, logprec_var_mat,xlab='Jours (1er juil.-30 juin)',
           ylab='Jours (1er juil.-30 juin)', zlab="variance (log10 precipitations)",
           theta=-45, phi=25, r=3)

levelplot(row.values=jours_5,column.values=jours_5,
          x=logprec_var_mat,contour=TRUE,xlab='Jours (1er juillet au 30 juin)',
          ylab='Jours (1er juillet au 30 juin)')
```

Covariance et corrélation (8) - Exemple

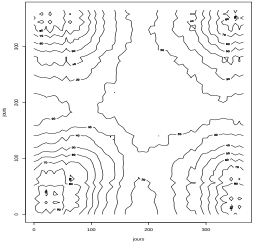
Exemple 2. Données `CanadianWeather`, températures



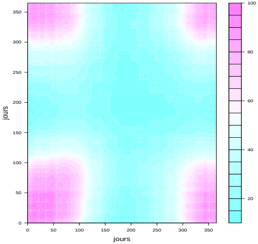
Covariance et corrélation (9) - Exemple

Exemple 2. Données `CanadianWeather`, températures

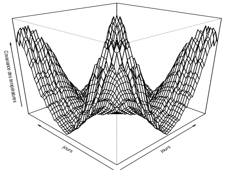
`contour`



`filled.contour`



`persp`

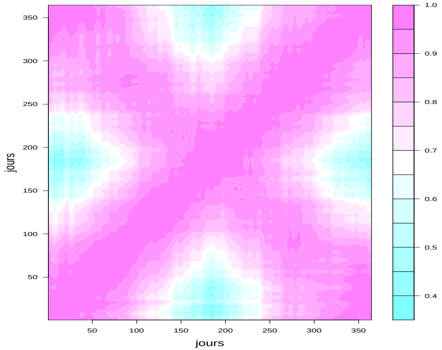


Code R.

```
tvar = var.fd(Donnees_lissees_temp)
temp_var_sem = eval.bifd(weeks,weeks,tvar)

contour(weeks,weeks,temp_var_sem,xlab="jours", ylab="jours")
filled.contour(weeks,weeks,temp_var_sem,
                  cex.lab=1.5,cex.axis=1.5,xlab="jours", ylab="jours")
persp(weeks,weeks,temp_var_sem,xlab="jours", ylab="jours",
                  zlab="Covariance des températures", theta=-45, phi=25, r=3)
```

Covariance et corrélation (10) - Exemple

Exemple 2. Données `CanadianWeather`, températures



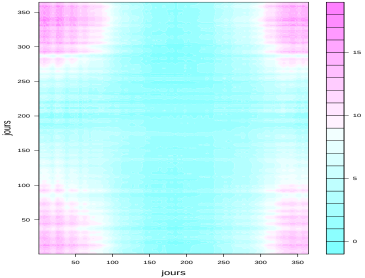
Code R.

```
jours = 1:365 - 0.5
tcor = cor.fd(jours,Donnees_lissees_temp)

filled.contour(jours,jours,tcor,cex.lab=1.5,cex.axis=1.5,xlab="jours", ylab="jours")
```

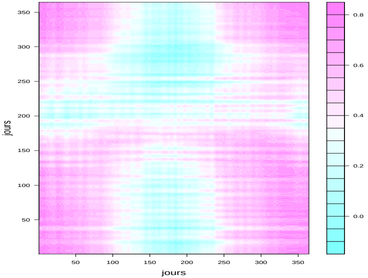
Covariance et corrélation (11) - Exemple

Exemple 3. Données `CanadianWeather`, températures et précipitations

Covariance croisée



Corrélation croisée



Code R.

```
temp_prec_var = var.fd(Donnees_lissees_temp,Donnees_lissees_precip)
temp_prec_var_jours = eval.bifd(jours,jours,temp_prec_var)
filled.contour(jours,jours,temp_prec_var_jours,cex.lab=1.5,
                        cex.axis=1.5,xlab="jours", ylab="jours")

temp_prec_cor_jours = cor.fd(jours,Donnees_lissees_temp,
                  jours,Donnees_lissees_precip)
filled.contour(jours,jours,temp_prec_cor_jours,
               cex.lab=1.5,cex.axis=1.5,xlab="jours", ylab="jours")
```

Plan

Introduction

- **ACP** : Analyse en Composantes Principales.

- **Objectifs** : outil de représentation des données et réduction de la dimension

  produire une synthèse visuelle / la meilleure représentation possible de données multivariées.

- **Type de méthode**
  - méthode factorielle (exploiter les aspects géométriques)
  - statistique exploratoire

- **ACP fonctionnelle** : extension au cadre fonctionnel de l'ACP multivariée.

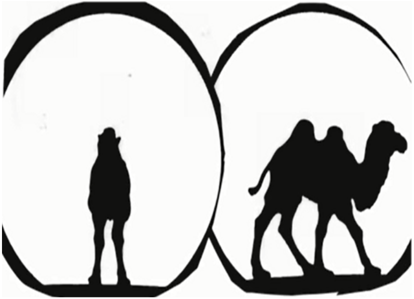
Plan

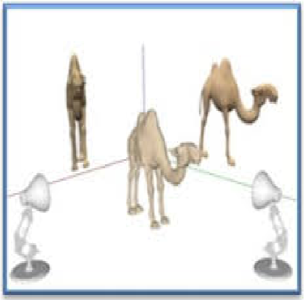
Rappels ACP multivariée (1)

De quelle image 3D celle-ci est-elle la représentation simplifiée ?

Rappels ACP multivariée (2)

Toutes les représentations simplifiées ne se valent pas !





Principe de l'ACP : réduire la dimension d'un nuage de points pour en obtenir une représentation plus simple tout en conservant le plus possible de variabilité.

Rappels ACP multivariée (3) - Point de vue théorique

- **Point de départ.** $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$, $d \geq 3$ vecteur aléatoire.

- **Problématique.** recherche des sous-espaces vectoriels de $\mathbb{R}^d$ qui résument le mieux l'information, pour représenter X dans un espace de dimension $< d$.

- **Méthode.**
  1. Diagonalisation en base orthonormée de la matrice de covariance Σ de X.
  2. Choix des sous-espaces engendrés par les vecteurs propres de Σ.

Rappels ACP multivariée (4) - Point de vue pratique

- **Point de départ.**
  - Observations $X_i = {}^t(X_{i,1}, \ldots, X_{i,d}) \in \mathbb{R}^d$, pour $i \in \{1, \ldots, n\}$

$$X_c = \begin{pmatrix} X_{1,1} - X_{.1} & & \cdots & X_{1,d} - X_{.d} \\ X_{2,1} - X_{.1} & X_{2,2} - X_{.2} & \cdots & X_{2,d} - X_{.d} \\ & & \ddots & \\ \vdots & & \ddots & \vdots \\ & & \ddots & \\ X_{n,1} - X_{.1} & & \cdots & X_{n,d} - X_{.d} \end{pmatrix}.$$

  - Matrice de covariance empirique

$$\Sigma = \left(\frac{1}{n} \sum_{i=1}^{n} (X_{i,j} - X_{.j})(X_{i,j'} - X_{.j'}) \right)_{1 \leq j,j' \leq d} = \frac{{}^t X_c X_c}{n}$$

- **Inertie du nuage de points**

$$I = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_{i,j} - X_{.j})^2 = Tr(\Sigma)$$

- **Décomposition de l'inertie totale.** pour tout sous-espace S de $\mathbb{R}^d$,

$$I = \frac{1}{n} \sum^{n} \|\Pi_S X_i - X_i\|^2 + \frac{1}{n} \sum^{n} \|\Pi_S X_i\|^2 = I_S + I_{S^\perp}.$$

Rappels ACP multivariée (4) - Point de vue pratique

- **Principe de l'ACP.** Recherche séquentielle (par récurrence) des axes de projections minimisant la déformation du nuage de départ.
    1. Recherche de Δ_1 tel que $I_{\Delta_1^\perp}$ est maximale.
       Solution : Δ_1 dirigé par a_1, vecteur propre normé associé à la plus grande valeur propre λ_1 de Σ.
    2. Recherche de Δ_2, orthogonal à Δ_1 et tel que $I_{\Delta_2^\perp}$ est maximale.
       Solution : Δ_2 dirigé par a_2, vecteur propre normé associé à la seconde plus grande valeur propre λ_2 de Σ.
    3. ...
- **Construction de** $S_k = \Delta_1 + \cdots + \Delta_k = \text{Vect}(a_1, \ldots, a_k)$, avec

$$\begin{array}{ccccccc} \Delta_1 & \perp & \Delta_2 & \perp & \ldots & \perp & \Delta_d \\ \lambda_1 & \geq & \lambda_2 & \geq & \cdots & \geq & \lambda_d \\ I_{\Delta_1^\perp} & \geq & I_{\Delta_2^\perp} & \geq & \cdots & \geq & I_{\Delta_d^\perp} \end{array}.$$

- **Représentation** des individus X_i dans la nouvelle base $\{a_1, \ldots, a_k\}$.

Plan

ACP fonctionnelle - Théorie (1)

	ACP multivariée	ACP fonctionnelle
Données	$X \in \mathbb{R}^d$ $X = {}^t(X_1, \ldots, X_d)$	$X \in L^2(T)$ $X = \{X(t), t \in T\}$
Moyenne	vecteur de moyenne $\mathbb{E}[X] = {}^t(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$	courbe de la moyenne $\mathbb{E}[X] = \{\mathbb{E}[X(t)], t \in T\}$
Covariance	matrice $\Sigma_X = \mathrm{Cov}(X, X)$	fonction de covariance C_X ou opérateur Γ_X $C_X(s, t) = \mathrm{Cov}(X(s), X(t))$

ACP fonctionnelle - Théorie (2)

- **Point de départ.** $X = \{X(t),\ t \in T\} \in L^2(T)$

- **Objectifs.** construction d'une base hilbertienne $(\psi_j)_{j \in \mathbb{N}\setminus\{0\}}$ de $L^2(T)$ telle que l'espace $S_k = \text{Vect}\{\psi_1, \dots, \psi_k\}$ minimise la distance L^2 entre X et sa projection orthogonale $\Pi_{S_k} X$.

- **Construction.** Recherche séquentielle (par récurrence)
  1. Recherche de $\psi_1 \in L^2(T)$, de norme $\|\psi_1\|^2 = \int_T \psi_1^2(t)dt = 1$, telle que
  $$\psi_1 \in \arg \min_{f \in L^2(T)} \mathbb{E}[\|X - \langle X, f\rangle f\|].$$
  2. Recherche de $\psi_2 \in L^2(T)$, de norme $\|\psi_2\|^2 = 1$ et $\langle \psi_2, \psi_1 \rangle = 0$, telle que
  $$\psi_2 \in \arg \min_{f \in L^2(T)} \mathbb{E}[\|X - \langle X, \psi_1\rangle \psi_1 - \langle X, f\rangle f\|].$$
  3. $\cdots$

  k+1. Si $\psi_1, \dots, \psi_k$ déjà construits, recherche de $\psi_{k+1} \in L^2(T)$, de norme $\|\psi_{k+1}\|^2 = 1$ et $\langle \psi_{k+1}, \psi_j \rangle = 0$ pour tout $j \leq k$, telle que
  $$\psi_{k+1} \in \arg \min_{f \in L^2(T)} \mathbb{E}[\|X - \Pi_{S_k} X - \langle X, f\rangle f\|].$$

ACP fonctionnelle - Théorie (3)

Proposition

La famille $(\psi_k)_k$ est constituée de fonctions propres de l'opérateur de covariance Γ_X associées aux valeurs propres $(\lambda_k)_{k\geq 1}$ rangées par ordre décroissant :

$$\Gamma_X \psi_j = \lambda_j \psi_j \text{ ou encore } \int_T C_X(s,\cdot)\psi_j(s)ds = \lambda_j \psi_j(\cdot).$$

Ainsi, la base de l'ACP fonctionnelle est, au signe près, la base de Karhunen-Loève de X.

ACP fonctionnelle - Théorie (4)

	ACP multivariée	ACP fonctionnelle
Données	$X \in \mathbb{R}^d$ $X = {}^t(X_1, \ldots, X_d)$	$X \in L^2(T)$ $X = \{X(t), t \in T\}$
Base de l'ACP	vecteurs propres $(a_j)_{j=1,\ldots,d}$ de la matrice de covariance	fonctions propres $(\psi_j)_{j\geq 1}$ de l'opérateur de covariance
Représentation ACP	$X = \mathbb{E}[X] + \sum_{j=1}^{d} \xi_j a_j$	$X(t) = \mathbb{E}[X](t) + \sum_{j=1}^{d} \xi_j \psi_j(t)$

Plan

ACP fonctionnelle - Point de vue pratique

- **Point de départ.**
  - Observations. $X_i = \{X_i(t),\ t \in T\} \in L^2(T)$, pour $i \in \{1, \ldots, n\}$
  - Opérateur de covariance empirique

$$\Gamma_n f(s) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, f\rangle X_i = \int_T C_n(s,t)f(t)dt, \quad C_n(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t).$$

- **Propriétés.** Γ_n est autoadjoint compact donc diagonalisable, de valeurs propres $(\hat{\lambda}_j)_{j\geq 1}$ et de vecteurs propres $(\widehat{\psi}_j)_{j\geq 1}$

- **Estimation de la base de Karhunen-Loève.**
  - $\hat{\lambda}_j$ estime λ_j.
  - $\widehat{\psi}_j$ estime ψ_j.

Plan

ACP fonctionnelle - Code R.

Fonction pca.fd du package `fda`

1. **Arguments principaux.**
   - `fdobj` : objet de la classe `fd` (données fonctionnelles après lissage), éventuellement multivarié,
   - `nharm` : nombre de fonctions propres à calculer (aussi appelées harmoniques),
   - `centerfns` : booléen, si la valeur est `TRUE`, centrage des données avant le calcul de l'ACP.

2. **Sortie.** retourne un objet de la classe `pca.fd`, dont les entrées principales sont
   - `harmonics` : objet fonctionnel (classe `fd`) contenant les fonctions propres $(\widehat{\psi_j})_{j\geq 1}$
   - `values` : ensemble complet des valeurs propres de l'opérateur de covariance empirique $(\hat{\lambda}_j)_{j\geq 1}$
   - `scores` : matrice des scores ie. des coordonnées des individus (courbes) dans la nouvelle base, ie. selon chacune des fonctions propres,
   - `varprop` : vecteur donnant la proportion de variance expliquée par chaque fonction propre,
   - `meanfd` : objet fonctionnel (classe `fd`) contenant la fonction moyenne.

ACP fonctionnelle - Représentation graphique

1. Tracé des premières fonctions propres $(\widehat{\psi_j})_{j\geq 1}$ (ou composantes principales), de manière à obtenir une proportion de variance expliquée supérieure ou égale à 90%.

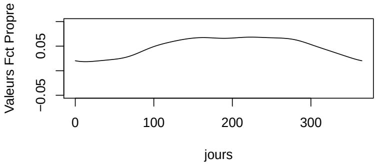
2. Représentation, dans un plan, des coordonnées des courbes de l'échantillon de départ selon la première et la seconde composantes principales (ou la seconde et la troisième, ...).

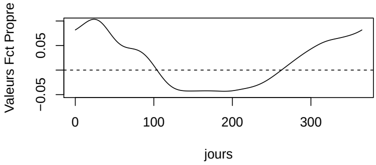
3. Représentation des composantes principales comme perturbation de la fonction moyenne.
   $\longrightarrow$ `plot.pca.fd`.

ACP fonctionnelle - Exemple (1)

Exemple. Données CanadianWeather, log-précipitations.

- $\widehat{\psi}_1$: 87.4% de la variabilité expliquée ($\hat{\lambda}_1 = 39.5$),
- $\widehat{\psi}_2$: 8.6% de la variabilité expliquée ($\hat{\lambda}_2 = 3.9$).



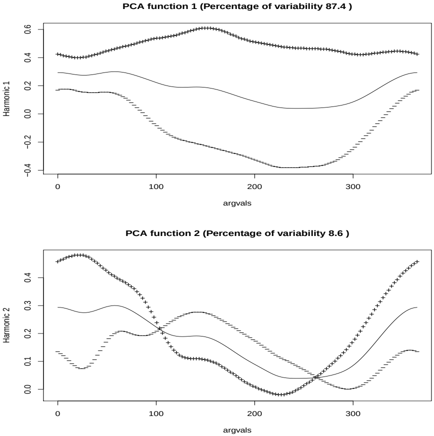


Code R.

```
Logprecip_ACP=pca.fd(Donnees_lissees_Logprecip,
           nharm=2)
print(Logprecip_ACP$values)
print(Logprecip_ACP$varprop)

par(mfrow=c(2,1))
plot(Logprecip_ACP$harmonics[1],xlab='jours',
           ylab="Valeurs Fct Propre",
           ylim=c(-0.05,0.1),main="PC1")
plot(Logprecip_ACP$harmonics[2],xlab='jours',
           ylab="Valeurs Fct Propre",
           ylim=c(-0.05,0.1),main="PC2")
```

ACP fonctionnelle - Exemple (2)

Exemple. Données CanadianWeather, log-précipitations.

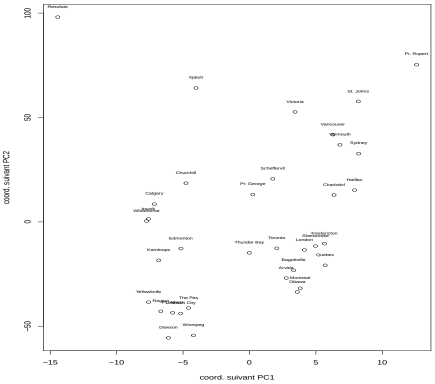


Code R.

```
par(mfrow=c(2,1))
plot.pca.fd(Logprecip_ACP)
```

ACP fonctionnelle - Exemple (3)

Exemple. Données CanadianWeather, log-précipitations.



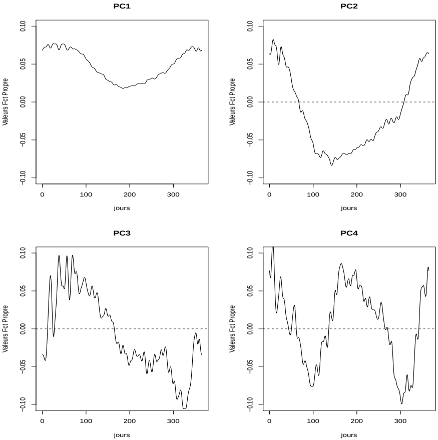
Code R.

```
par(mfrow=c(1,1))
plot(Logprecip_ACP$scores[,1],Temp_ACP$scores[,2],xlab="coord. suivant PC1",
        ylab="coord. suivant PC2")
text(Logprecip_ACP$scores[,1],Temp_ACP$scores[,2]+5,CanadianWeather$place,cex=0.6)
```

ACP fonctionnelle - Exemple (4)

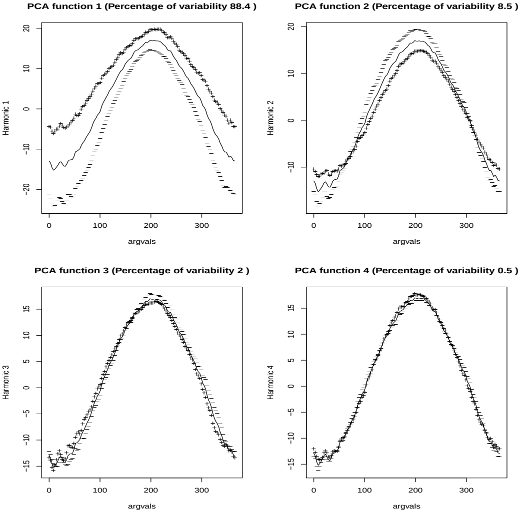
Exemple. Données CanadianWeather, températures.

Fonction propre	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_4$	$\cdots$
Variabilité expliquée	88.4%	8.5%	2%	0.5%	$\cdots$



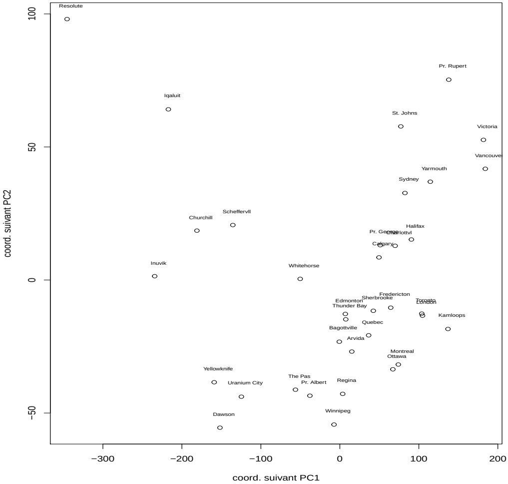
ACP fonctionnelle - Exemple (5)

Exemple. Données CanadianWeather, températures.



ACP fonctionnelle - Exemple (6)

Exemple. Données CanadianWeather, températures.



ACP fonctionnelle - Pour aller plus loin

- **ACP fonctionnelle pour données fonctionnelles multivariées**
- **Transformation de la base de l'ACP** par rotation (stratégie VARIMAX).