

Regression with categorical functional data

Cristian Preda

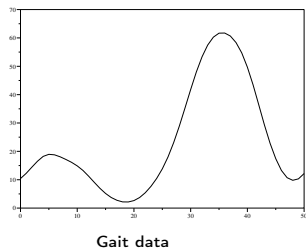
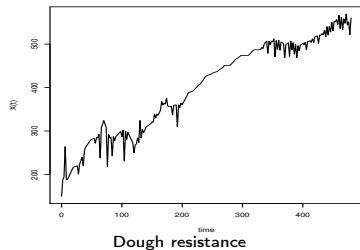
Laboratoire de Mathématiques Paul Painlevé,
Université des Sciences et Technologies de Lille,
Villeneuve d'Ascq, 59655, France

`cristian.preda@polytech-lille.fr`

December 14, 2015
ERCIM 2015, London

Categorical functional data

In the literature, most of the functional data is scalar (univariate/multivariate) :



Model : Stochastic process, $X = \{X_t, t \in \mathcal{T}\}$,

$$X_t \in \mathbb{R}^p, p \geq 1,$$

for some index set \mathcal{T}

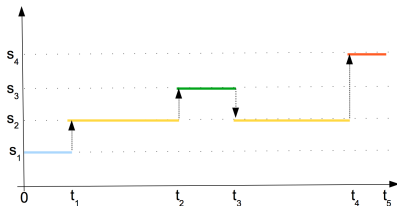
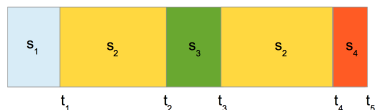
Categorical functional data

Set of states : $\mathbf{S} = \{s_1, s_2, \dots, s_m\}$, $m \geq 2$

$\mathbf{X} = \{X_t : t \in [0, T]\}$, (Ω, \mathcal{A}, P) , $X_t : \Omega \rightarrow \mathbf{S}$.

A path of \mathbf{X} on $[0, T]$ is a sequence of states s_i and times points t_i of transitions from one state to another :

$$\{(s_1, t_1), (s_2, t_2), \dots, (s_f, t_f)\}$$



Categorical functional data

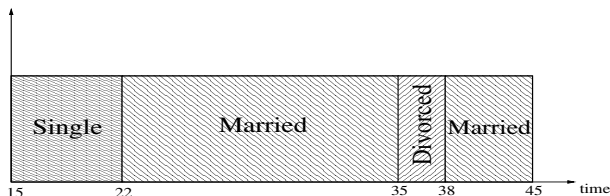
Saporta et Deville (1982,1983) :

- *Analyse de données chronologiques qualitatives : comment analyser des calendriers ?*, Annales de l'INSEE, No. 45, p. 45-104.

- *Correspondence analysis with an extension towards nominal time series*, Journal of Econometrics, 22, p. 169-189.

X = marital status of women from 15 to 45

$$X_t \in \{\text{"Single"}, \text{"Married"}, \text{"Divorced"}, \text{"Widowed"}\}$$



Categorical functional data

Model :

$$\mathbf{S} = \{s_1, s_2, \dots, s_m\}, (\Omega, \mathcal{A}, P),$$

$\mathbf{X} = \{X_t : t > 0\}$, $X_t : \Omega \rightarrow \mathbf{S}$, a continuous-time jump process

$$\mathbf{X} = \vec{\mathbf{1}} = \{\vec{1}_t : t \in [0, T]\},$$

$$1_t^x(\omega) = \begin{cases} 1 & \text{if } X_t(\omega) = x, \\ 0 & \text{otherwise.} \end{cases}$$

Markov jump process and categorical functional data :

Harmonic qualitative analysis (extension of the multiple correspondance analysis towards categorical fd)

- Richard (1988), Heijden et al (1997), Preda (1998).

Length of paths and the absorbing states

- ▶ X can have one or several *absorbing states* :

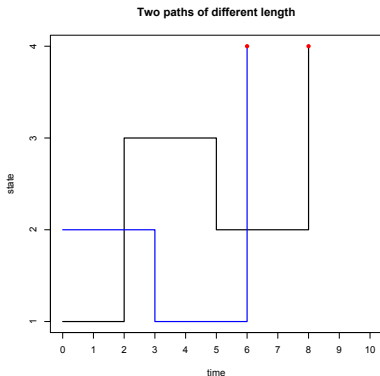
$$\mathbb{P}(X_{t+s} = s | X_t = s) = 1, \forall s > 0$$

Two situations :

X is observed :

- ▶ until an absorbant state is reached

Remark :
functional data with different lengths !



- ▶ on a predefined period $[0, T]$
(classical framework of functional data)

The functional regression model

Let $Y : \Omega \rightarrow \mathbb{R}$ be the response variable.

$$Y = f(X) + \varepsilon,$$

where f is a functional (regression function).

Minimize **risk expectation** :

$$R = \mathbb{E}(\mathcal{C}(X, Y, f)),$$

where \mathcal{C} is a loss function.

Examples.

- Least squares : $\mathcal{C}(x, y, f) = (y - f(x))^2$.
- Logistic : $f(x) = \log\left(\frac{p(x)}{1 - p(x)}\right)$, with $p(x) = P(Y = 1|X = x)$

$$\mathcal{C}(x, y, f) = -yf(x) + \log(1 + e^{f(x)}).$$

Risk minimisation and consistency

Minimize :

$$R[f] = \mathbb{E}(\mathcal{C}(X, Y, f)),$$

over some space \mathcal{F} of functionals.

From data $\{(X_i, Y_i)\}_{i=1}^n$, minimize :

$$R_{emp}[f] = \frac{1}{n} \sum_{i=1}^n \mathcal{C}(X_i, Y_i, f(X_i)).$$

Consistency : dimension reduction, risk regularization.

Dimension reduction : Functional MCA

Hypothesis :

$$\lim_{h \rightarrow 0} \mathbb{P}(X_{t+h} \neq X_t) = 0.$$

$\forall t \in [0, T], x \in \mathbf{S}$

$$\mathbf{1}_t^x(\omega) = \begin{cases} 1 & \text{if } X_t(\omega) = x, \\ 0 & \text{otherwise.} \end{cases}$$

$$p^x(t) = \mathbb{P}(X_t = x),$$

$$p^{x,y}(t, s) = \mathbb{P}(X_t = x \cap X_s = y).$$

FMCA : The principal components

$$z = \int_0^T \sum_{x=1}^m \mathbf{1}_t^x a^x(t) dt.$$

$\{a^x\}_{x=1,\dots,m}$ (optimal encoding states).

$$\lambda a^x(t) = \int_0^T \sum_{y=1}^m \frac{p^{x,y}(t,s)}{p^x(t)} a^y(s) ds,$$

Expansion formula :

$\{\lambda_i\}_{i \geq 1}, \{\bar{a}_i\}_{i \geq 1},$

$\forall s, t \in T, \forall x, y \in \mathbf{S} :$

$$p^{x,y}(t,s) = p^x(t)p^y(s) \sum_{i \geq 1} \lambda_i a_i^x(t) a_i^y(s).$$

Approximating the FMCA

The optimal encodings :

$$\{a^x\}_{x=1,\dots,m}$$

$$\lambda a^x(t) = \int_0^T \sum_{y=1}^m \frac{p^{x,y}(t,s)}{p^x(t)} a^y(s) ds,$$

Approximation into a basis of functions :

$$a : \mathbf{S} \times [0, T] \rightarrow \mathbf{R},$$

Basis functions : $\{\varphi_i\}_{i=1,\dots,p}$,

$$\varphi_i : \mathbf{S} \times T \rightarrow \mathbf{R}, \forall i = 1, \dots, p, p \geq 1$$

The approximation :

$$a^x(t) = \sum_{j=1}^p \alpha_j \varphi_j(x, t)$$

FMCA : optimal encode approximation

Let define the random variables :

$$V_i = \int_0^T \varphi_i(X_t, t) dt \quad \forall i = 1, \dots, p,$$
$$U_{i,j} = \int_0^T \varphi_i(X_t, t) \varphi_j(X_t, t) dt \quad \forall i, j = 1, \dots, p,$$

and the matrices \mathbf{G} and \mathbf{F} by

$$G_{i,j} = \text{Cov}(V_i, V_j),$$

$$F_{i,j} = E(U_{i,j}), \quad \forall i, j = 1, \dots, p.$$

Then,

$$\lambda a^x(t) = \int_0^T \sum_{y=1}^m \frac{p^{x,y}(t, s)}{p^x(t)} a^y(s) ds \iff \mathbf{F}^{-1} \mathbf{G} \alpha = \lambda \alpha.$$

FMCA and regression

$$\mathbf{X} = \{X_t, t \in [0, T]\} : \iff \{z_1, z_2, \dots, z_k, \dots\}$$

(principal components)

FMCA approximation for the linear regression model :

$$\mathbb{E}(Y|\mathbf{X}) = \mathbb{E}(Y|z_1, \dots, z_k) = b_0 + b_1 z_1 + \dots + b_k z_k$$

$$\begin{aligned} &= b_0 + b_1 \int_0^T \sum_{i=1}^m a_1^i(t) \mathbf{1}_t^i dt + \dots + b_k \int_0^T \sum_{i=1}^m a_k^i(t) \mathbf{1}_t^i dt \\ &= b_0 + \int_0^T \sum_{i=1}^m \beta_i(t) \mathbf{1}_t^i dt. \end{aligned}$$

where $\beta_i(t) = \sum_{j=1}^k b_j a_j^i(t)$.

Risk regularization, RKHS and the Representer Theorem

$(\mathcal{F}, \|\cdot\|)$,

$$R_{\text{reg}}[f] = \frac{1}{n} \sum_{i=1}^n \mathcal{C}(X_i, Y_i, f(X_i)) + \lambda \|f\|_{\mathcal{F}}^2,$$

where $\lambda > 0$ is a regularized parameter.

$\mathcal{F} = \mathcal{H}_{\mathbf{K}}$, $\mathcal{H}_{\mathbf{K}}$ is an RKHS with kernel \mathbf{K} .

Representer Theorem. Let $\{(x_i, y_i)\}_{i=1}^n$, $x_i \in H$, $y_i \in \mathbb{R}$, $n > 0$, $n \in \mathbb{N}$, $\lambda > 0$ and $\mathcal{C} : H \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a convex loss function of third argument. Then, the solution to the problem : find $\hat{f} \in \mathcal{H}_{\mathbf{K}}$ to minimize

$$\frac{1}{n} \sum_{i=1}^n \mathcal{C}(x_i, y_i, \hat{f}(x_i)) + \lambda \|\hat{f}\|_{\mathcal{H}_{\mathbf{K}}}^2$$

exists, is unique and admits a representation of the form

$$\hat{f}(x) = \sum_{i=1}^n \alpha_i \mathbf{K}(x, x_i), \quad \forall x \in H,$$

where $\alpha_i \in \mathbb{R}$, $i = 1, \dots, n$.

Kernels for categorical functional data

Let H be the space whose elements are the paths of X .

Let $x, y \in H$, $x = \{x_t, t \in [0, T]\}$, $y = \{y_t, t \in [0, T]\}$.

$$x(t) = (\mathbf{1}_x^1(t), \dots, \mathbf{1}_x^m(t))$$

$$y(t) = (\mathbf{1}_y^1(t), \dots, \mathbf{1}_y^m(t))$$

The inner product :

$$\langle x, y \rangle_H = \int_0^T \sum_{i=1}^m \mathbf{1}_x^i(t) \mathbf{1}_y^i(t) dt.$$

$\langle x, y \rangle_H$ = the length of time within $[0, T]$ where x and y are in the same state.

Notice that, $\|x\|^2 = T$ and $\|x - y\|^2 = 2(T - \langle x, y \rangle_H)$

Kernels for categorical functional data

$K : H \times H \mapsto \mathbb{R}$ a RBF or polynomial kernel,

Gaussian :

$$K(x, y) = e^{-\frac{\|x - y\|_H^2}{2\sigma^2}}, \quad \sigma > 0$$

Polynomial :

$$K(x, y) = (c + \langle x, y \rangle_H)^d, \quad c > 0$$

Numerical application : a simulation example

$$\mathbf{S} = \{1, 2\}, T = [0, 1].$$

$$X(0) = 1, \theta \sim U[0, 1],$$

$$\omega \in \Omega : X(\omega, t) = \begin{cases} 1, & \text{if } t \leq \theta(\omega) \\ 2, & \text{otherwise.} \end{cases}$$

$$\begin{aligned} Y &= \int_0^T e^t \mathbf{1}_t dt + \varepsilon(t), \\ &= e^\theta - 1 + \varepsilon(t), \end{aligned}$$

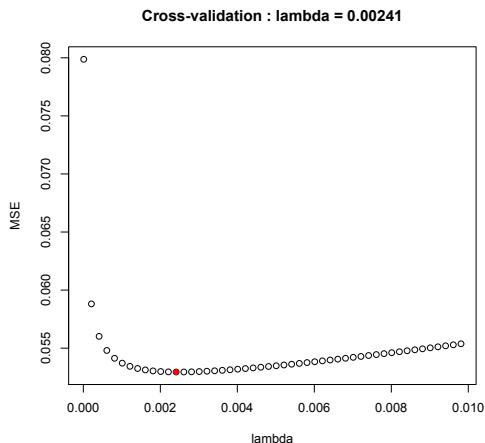
$$\varepsilon(t) \sim \mathcal{N}(0, \sigma_r^2)$$

With $\sigma_r^2 = 0.05$ we have a noise-to signal ratio ($\frac{\sigma_r^2}{\mathbb{V}(Y)}$) of about 80%.

Numerical application : a simulation example

Data : $n = 200$ simulated paths of X .

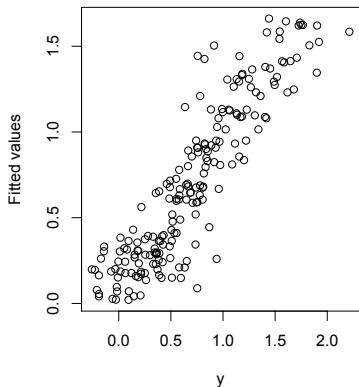
RKHS regression : Gaussian Kernel ($\sigma^2 = 1$).



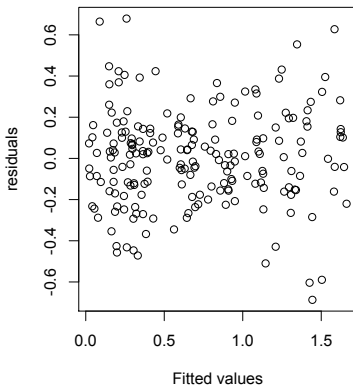
Numerical application : a simulation example

RKHS regression

RKHS regression: Gaussian(1)



RKHS regression : MSE = 0.0529



References

1. B. Scholkopf, A.J. Smola : *Learning with kernels*, The MIT Press (2002)
3. C. Preda : *Regression models for functional data by reproducing kernel Hilbert spaces*, Journal of Statistical Planning and Inference, 137, 829–840 (2007).
3. H. Lian : *Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces*, Canadian Journal of Statistics, 35 (4), 597–606 (2007)