Synthèse : Récupération des scores ACM à partir de données catégorielles fonctionnelles

1 Données d'entrée

On observe un processus catégoriel :

$$X = \{X_t : t \in [0, T]\}, \quad X_t \in S = \{s_1, \dots, s_K\}$$

Chaque trajectoire représente l'évolution d'un individu dans le temps entre différents états s_k.

2 Encodage indicateur (binarisation)

On transforme chaque trajectoire en vecteurs indicateurs :

$$1_t^x = \begin{cases} 1 & \text{si } X_t = x \\ 0 & \text{sinon} \end{cases}$$

On obtient K trajectoires binaires (une par état).

3 Estimation des probabilités

— Probabilité marginale :

$$p_x(t) = \mathbb{P}(X_t = x)$$

— Probabilité conjointe :

$$p_{x,y}(t, s) = \mathbb{P}(X_t = x, X_s = y)$$

— Noyau de covariance fonctionnelle :

$$K(t, s) = E_t E_s$$

4 Problème spectral

On résout le problème aux valeurs propres suivant :

$$\int_0^T \sum_{y \in S} p^{x,y}(t, s)\, a^y(s)\, ds = \lambda\, a^x(t)\, p^x(t), \quad \forall t \in [0, T], \forall x \in S$$

Ce système donne :
— Les valeurs propres λ_i
— Les fonctions propres $a_i^x(t)$ (encodages optimaux)

5 Construction des composantes principales

— Fonction encodée :

$$x_i(t) = \sum_{x \in S} a_i^x(t) \cdot 1_t^x, \quad \forall t \in [0, T]$$

— Score ACM (composante principale) :

$$z_i = \int_0^T x_i(t)\, dt$$

6 Propriétés des scores ACM

— $\mathbb{E}(z_i) = 0$
— $\mathrm{Cov}(z_i, z_j) = 0$ si $i \neq j$

Approximation de la trajectoire :

$$1_t^x \approx \sum_{i=1}^{q} z_i\, a_i^x(t)\, p^x(t), \quad \forall x \in S$$

Ainsi, on obtient les q premières composantes principales :

$$\{z_1, z_2, \ldots, z_q\}, \quad q \geq 1$$