

Synthèse des Scores ACM à partir de Données Catégorielles Fonctionnelles

Moustapha Sarr

31/03/2025

1 Introduction

Dans le cadre de la **FMCA** (Functional Multiple Correspondence Analysis) appliquée aux données fonctionnelles catégorielles, on cherche à représenter chaque trajectoire $X_i(t)$ (qui prend des valeurs dans un ensemble fini $E = \{e_1, e_2, \dots, e_K\}$) par une projection dans un espace réduit. On obtient ainsi, pour chaque individu i , un vecteur de scores

$$z_i = (z_{i1}, z_{i2}, \dots, z_{iq}) \in \mathbb{R}^q$$

qui résume sa trajectoire. La FMCA détermine pour chaque état $e \in E$ une fonction $a^e(t)$ qui représente la valeur numérique optimale de l'état e à l'instant t . Pour obtenir les scores individuels de chaque individu, on remplace chaque état de sa trajectoire par son code numérique optimal, puis on intègre (ou on somme) ces valeurs sur le temps..

2 Étapes principales

2.1 Données disponibles

- **Données fonctionnelles catégorielles** : chaque individu i possède une trajectoire temporelle

$$X_i(t), \quad t \in [0, T], \quad \text{avec } X_i(t) \in E = \{e_1, e_2, \dots, e_K\}$$

Cela signifie que l'état de l'individu i évolue dans le temps, en prenant des valeurs dans un ensemble fini d'états catégoriels E .

2.2 Transformation en variables binaires

Pour chaque catégorie $x \in E$, on définit la variable indicatrice binaire associée:

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

Cela permet de convertir les données catégorielles en une représentation numérique exploitable pour l'analyse en composante principale (ACP).

2.3 Calcul des probabilités

- Probabilité marginale : $p^x(t) = P(X_t = x)$.
- Probabilité conjointe : $p^{x,y}(t, s) = P(X_t = x, X_s = y)$.

2.4 Hypothèses

- **Hypothèse H_1** : Le processus X est continu en probabilité

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

Ce qui signifie que de petits changements dans le temps entraînent rarement des changements brusques d'état.

- **Hypothèse H_2** : Pour chaque instant $t \in [0, T]$ (à l'exception possible d'un nombre fini de temps discrets), chaque état a une probabilité strictement positive d'apparaître :

$$p^x(t) \neq 0, \forall x \in E, \forall t \in [0, T].$$

Ce qui signifie que pour presque tout instant t dans l'intervalle $[0, T]$, sauf éventuellement pour quelques instants isolés, chaque état x de l'ensemble E a une probabilité strictement positive d'être observé.

2.5 L'opérateur d'espérance conditionnelle

Nous cherchons ici à extraire l'information dominante contenue dans des données fonctionnelles catégorielles via une décomposition en composantes principales. Pour cela, on définit des opérateurs et des critères qui permettent d'obtenir l'encodage optimal des trajectoires catégorielles.

Nous définissons l'opérateur d'espérance conditionnelle associé à X_t , par :

$$E_t : L^2(\Omega) \rightarrow L(X_t), \quad z \in L^2(\Omega), \quad z \mapsto E_t(z) = \sum_{x \in E} \mathbb{E}(z \mid X_t = x) \mathbf{1}_t^x$$

E_t est le projecteur orthogonal sur l'espace linéaire engendré par $\{\mathbf{1}_t^x, x \in E, t \in [0, T]\}$.

- $L^2(\Omega)$: L'ensemble des variables aléatoires réelles avec une variance finie.
- $L(X_t)$: L'espace linéaire engendré par les indicateurs $\mathbf{1}_t^x$

3 Propriétés de l'opérateur de projection conditionnelle E_t

- L'opérateur E_t est **autoadjoint**, ce qui signifie que pour toutes variables aléatoires $z_1, z_2 \in L^2(\Omega)$:

$$\langle E_t(z_1), z_2 \rangle = \langle z_1, E_t(z_2) \rangle$$

Le produit scalaire est conservé par permutation.

- L'opérateur E_t est **idempotent**, c'est-à-dire que :

$$E_t(E_t(z)) = E_t(z)$$

Cela signifie qu'une fois la projection effectuée, la refaire ne modifie plus le résultat. C'est la propriété fondamentale d'un projecteur.

- Préservation de l'espérance :

$$\mathbb{E}[E_t(z)] = \mathbb{E}[z]$$

Cette propriété découle de la définition de l'espérance conditionnelle : prendre l'espérance de la projection revient à prendre l'espérance originale.

Pour évaluer combien de la variance de z est expliquée par sa projection sur l'espace des indicatrices, on définit :

$$\eta^2(z; X_t) = \frac{\text{Var}(E_t(z))}{\text{Var}(z)}$$

- C'est un coefficient de corrélation entre la variable aléatoire z qu'on cherche à construire et la variable catégorielle X_t .
- $\text{Var}(E_t(z))$: Variance expliquée.
- $\text{Var}(z)$: Variance totale.

Si on a plusieurs instants t_1, t_2, \dots, t_p , alors on cherche un z qui est globalement bien expliquée par toutes ces variables catégorielles. On fait :

$$\text{maximiser } \sum_{i=1}^p \eta^2(z; X_t)$$

C'est à dire trouver une variable aléatoire z qui résume au mieux l'information partagée par plusieurs variables catégorielles. Cette variable devient la première composante principale.

Extension au cas fonctionnel

Dans le cas fonctionnel, on ne se contente plus de quelques instants, considère tous les instants $t \in [0, T]$. Alors on maximise :

$$\int_0^T \eta^2(z; X_t) dt \tag{1}$$

C'est à dire trouver une variable aléatoire z qui est au maximum corrélée avec l'ensemble du processus X_t sur toute la durée d'observation.

3.1 Problème aux valeurs propres

Sous les hypothèses H_1 et H_2 , z qui maximise (1) est associé à la plus grande valeur propre du problème aux valeurs propres stochastique :

$$\int_0^T E_t(z) dt = \lambda z \tag{2}$$

L'opérateur $Q = \int_0^T$ est positif, hermitien et compact. Cela signifie que :

- Ses valeurs propres sont bien définies et positives.
- L'ensemble de ses vecteurs propres $\{z_i\}_{i \geq 1}$ est comptable.
- On peut ordonner ses valeurs propres par importance :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq 0$$

Les variables $\{z_i\}_{i \geq 1}$ sont appelées les composantes principales du processus X . Elles sont centrées et non corrélées.

3.2 Fonctions d'encodages optimales

L'idée est d'exprimer z sous une forme alternative en introduisant une fonction ψ_t telle que:

$$\psi_t = \frac{1}{\lambda} E_t(z), \forall t \in [0, T]$$

Expression de z en fonction de ψ_t

En intégrant la définition de ψ_t , on retrouve :

$$z = \int_0^T \psi_t dt. \quad (3)$$

3.3 Nouvelle formulation sous forme d'un problème aux valeurs propres

$$\int_0^T K(t, s) \psi_s ds = \lambda \psi_t, \forall t \in [0, T] \quad (4)$$

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

Normalisation pour rendre les solutions uniques

Lorsqu'on résout un problème aux valeurs propres, pour obtenir des solutions uniques, il faut imposer une contrainte de normalisation sur les vecteurs (ou fonctions) propres.

$$\int_0^T \text{Var}(\psi_t) dt = \int_0^T \mathbb{E}(\psi_t^2) dt = 1$$

3.4 Relation entre la variance de z et λ

A partir de la contrainte précédente et de l'équation (3), on obtient

$$\text{Var}(z) = \mathbb{E}(z^2) = \lambda$$

La variance de la première composante principale est égale à la plus grande valeur propre λ .

3.5 Définition des fonctions d'encodage optimal

$$\psi_t = \sum_{x \in E} a^x(t) 1_t^x$$

Les coefficients $\{a^x\}_{x \in S}$ sont des fonctions déterministes, appelées fonctions d'encodage optimal, car elle permettent de transformer un état x en une valeur numérique optimisé (ces coefficients déterminent l'importance des états à un instant t donné.)

En remplaçant ψ_t dans l'équation (4), on obtient

$$\int_0^T \sum_{y \in E} p^{x,y}(t, s) a^y(s) ds = \lambda a^x(t) p^x(t), \quad \forall t \in [0, T], \forall x \in E \quad (5)$$

Nouvelle contrainte de normalisation

$$\int_0^T \sum_{x \in E} [a^x(t)]^2 p^x(t) dt = 1$$

Cette contrainte assure que l'ensemble des fonctions d'encodage optimales pondéré par $p^x(t)$ est normalisé. Cela évite des solutions triviales multipliées par une constante arbitraire

3.6 Expression des composantes principales z_i en fonction des $a^x(t)$

Finalement, on peut exprimer les composantes principales z_i comme :

$$z_i = \int_0^T \sum_{x \in E} a_i^x(t) 1_t^x dt, \quad \forall i \geq 1$$

Ceci indique que pour la i -ème composante principale, le score de chaque individu est obtenu en intégrant sur le temps la somme, pour tous les états, des contributions $a_i^x(t)$ pondérées par l'indicateur qui sélectionne l'état réel observé. Ce score résume ainsi la trajectoire catégorielle sous la i -ème dimension optimisée.

Expansion des indicatrices 1_t^x en termes des composantes principales

Les fonctions d'encodage optimales $\{a_i^x(t), i \geq 1, x \in E\}$ (issu de la résolution du problème aux valeurs propres) forment une base d'un sous-espace de $\mathcal{L}^2([0, T])$ associé aux variables X_t . La propriété fondamentale des espaces de Hilbert nous dit alors que toute fonction mesurable par rapport à X_t , en particulier, la fonction indicatrice 1_t^x , peut être exprimée comme une combinaison linéaire (une série) de cette base. Ce qu'indique l'expansion suivante :

$$1_t^x = \sum_{i \geq 1} z_i a_i^x(t) \frac{1}{p^x(t)}, \quad \forall x \in E$$

C'est l'analogie de la décomposition de **Karhunen-Loève**, mais appliqués à des données catégorielles.

Décomposition de la probabilité jointe par application du théorème de Mercer à l'équation (5)

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

Probabilité marginale $p^x(t)$

Si on pose $x = y$ et $t = s$, on obtient

$$p^x(t) = \left(\sum_{i \geq 1} [a_i^x(t)]^2 \right)^{-1}$$

Cela relie directement les probabilités marginales aux fonctions propres.

3.7 Réduction de dimension

L'idée est d'approximer X en utilisant un nombre limité de composantes principales :

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

Au lieu d'utiliser toutes les composantes principales, on peut se limiter aux q premières pour une bonne approximation.

Ainsi, les q premières composantes principales,

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

Application du théorème de Mercer

On part du problème aux valeurs propres formulé pour la fonction d'encodage optimale, qui, après certaines manipulations (voir équations (11) et (12)), conduit à une équation intégrale du type :

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

Ici,

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

représente la probabilité conjointe.

Le théorème de Mercer s'applique à des noyaux continus, symétriques et positifs. On peut construire un noyau « normalisé » en divisant par les marges. Par exemple, on définit le noyau

$$K_{x,y}(t, s) = \frac{p_{x,y}(t, s)}{p_x(t) p_y(s)},$$

qui est bien défini sous l'hypothèse H2 (ce qui garantit que $p_x(t) \neq 0$ pour presque tout t). D'après Mercer, ce noyau peut être décomposé sous la forme :

$$K_{x,y}(t, s) = \sum_{i \geq 1} \lambda_i \psi_i^x(t) \psi_i^y(s),$$

où les fonctions propres $\psi_i^x(t)$ (selon chaque état x) forment une base orthonormale dans l'espace $L^2([0, T], p^x(t) dt)$.

En ramenant cette décomposition à la forme non normalisée, on a

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

Alors, en multipliant chaque côté par $p_x(t) p_y(s)$, nous obtenons :

$$\boxed{p_{x,y}(t, s) = p_x(t) p_y(s) \sum_{i \geq 1} \lambda_i a_i^x(t) a_i^y(s),}$$

où nous avons identifié les fonctions propres $\psi_i^x(t)$ aux fonctions d'encodage optimales $a_i^x(t)$.