A functional spatial autoregressive model using signatures

Camille Frévent

1Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France.

We are interested here in modelling the relationship between a real-valued random variable $Y$ and a functional covariate $\{X(t), t \in T\}$ observed in $N$ spatial locations. A traditional approach is to assume that $X$ belongs to $L^2(T)$, the space of square-integrable functions on $T$, and to consider the following model:

$$Y_i = \rho^* \sum_{j=1}^{N} v_{ij,N} Y_j + \int_T X_i(t) \theta^*(t) \, dt + \varepsilon_i, \quad i = 1, \ldots, N, \; N = 1, 2, \ldots$$

where the spatial dependency structure between the spatial units is described by the spatial weights matrix $V_N = (v_{ij,N})_{1 \leq i,j \leq N}$, the autoregressive parameter $\rho^*$ is in a compact space $R$ and $\theta^* \in L^2(T)$.

Fermanian (2022) recently investigated the use of signatures in the context of a non-spatial linear regression model with functional covariates. They present the advantages of being applicable to a wide range of processes that are not necessary square-integrable processes.

Concept of signatures Let $T$ be a compact interval and $X : T \to \mathbb{R}^p$ be a $p$-dimensional continuous function, $p \geq 2$. Let $(e_i)_{i=1}^p$ be the canonical orthonormal basis of $\mathbb{R}^p$. Then the signature of $X$ can be written as

$$Sig(X) = 1 + \sum_{d=1}^{\infty} \sum_{i_1, \ldots, i_d} S_{i_1, \ldots, i_d}(X)e_{i_1} \otimes \cdots \otimes e_{i_d}$$

where $S_{i_1, \ldots, i_d}(X) = \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} \cdots dX^{(i_1)}(t_1) \cdots dX^{(i_d)}(t_d)$.

Model We consider the following signatures-based spatial autoregressive model:

$$Y_i = \rho^* \sum_{j=1}^{N} v_{ij,N} Y_j + \alpha^* + (\theta^*, Sig(X_i)) + \varepsilon_i$$

(1)

where the parameter $\theta^*$ is assumed to be written as $\theta^* = 1 + \sum_{d=1}^{\infty} \sum_{i_1, \ldots, i_d} \beta_{i_1, \ldots, i_d}^* e_{i_1} \otimes \cdots \otimes e_{i_d}$.

The disturbances $\varepsilon_i$ are assumed to be independent and identically distributed random variables such that $E(\varepsilon_i) = 0$, $E(\varepsilon_i^2) = \sigma^2$. They are also independent of $X$.

Then, one can rewrite (1) as

$$Y_i = \rho^* \sum_{j=1}^{N} v_{ij,N} Y_j + \alpha^* + 1 + \sum_{d=1}^{\infty} \sum_{i_1, \ldots, i_d} \beta_{i_1, \ldots, i_d}^* S_{i_1, \ldots, i_d}(X_i) + \varepsilon_i$$

(2)

However, this model cannot be maximized without addressing the difficulty produced by the infinite dimension of the signatures $Sig(X_i)$ (and thus the infinite number of coefficients $\beta_{i_1, \ldots, i_d}^*$).

Thus we proposed two estimation methods that overcome this challenge, respectively based on a penalized spatial regression and a PCA.

References


*Corresponding author: camille.frevent@univ-lille.fr