Statistics and Its Interface Volume 15 (2022) 155

Discussion on "Estimation of Hilbertian varying coefficient models"

Xiongtao Dai

The authors Lee, Park, Kong, and Kim are to be congratulated on their comprehensive work. It is encouraging that the varying coefficient models and the smooth backfitting algorithm can be successfully extended to Hilbertian responses. The estimators enjoy nice theoretical properties akin to those in the Euclidean space.

While the challenge of infinite-dimensional responses has been solved, a related question is how to deal with non-linear responses in varying coefficient models. In the electricity consumption application, the densities are modeled in a linear space after applying a logarithm transformation. However, if the geometry specified for the densities is non-Euclidean, the Hilbertian techniques must be modified. For example, when there are zero values in the densities, the densities are better modeled as square-root densities on the Hilbert sphere [3], which has a non-linear geometry.

More generally, consider response $\mathbf{Y}$ lying on a m-dimensional Riemannian manifold $\mathcal{M}$ endowed with geodesic distance metric $\rho : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$, paired with d-dimensional Euclidean predictors $\mathbf{X} = (X_1, \ldots, X_d)$. Adopting the Fréchet mean [6] as a notion of average, define the conditional Fréchet mean function as $\boldsymbol{\mu} : \mathbb{R}^d \to \mathcal{M}$,

$$\boldsymbol{\mu}(\mathbf{x}) = \underset{\mathbf{p} \in \mathcal{M}}{\arg\min}\, E_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}[\rho^2(\mathbf{p}, \mathbf{Y})],$$

where the expectation is taken w.r.t. the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$.

Let $\phi : U \subset \mathcal{M} \to \mathbb{R}^m$ be a chart where U is an open set covering the conditional Fréchet means of $\mathbf{Y}$ given predictors running through their support. A Riemannian varying coefficient model considers $\phi(\boldsymbol{\mu}(\mathbf{X})) = \eta(\mathbf{X})$, where $\eta : \mathbb{R}^m \to \mathbb{R}^m$ is an index function that depends on the inputs in a varying coefficient fashion. For an error model, assume that

$$\mathbf{Y} = \exp_{\boldsymbol{\mu}(\mathbf{X})}(\boldsymbol{\epsilon}),$$

where $\exp_{\boldsymbol{\mu}(\mathbf{X})} : T_{\boldsymbol{\mu}(\mathbf{X})}\mathcal{M} \to \mathcal{M}$ is the Riemannian exponential map at $\boldsymbol{\mu}(\mathbf{X}) \in \mathcal{M}$, and $\boldsymbol{\epsilon} \in T_{\boldsymbol{\mu}(\mathbf{X})}\mathcal{M}$ is the error on the linear tangent space $T_{\boldsymbol{\mu}(\mathbf{X})}\mathcal{M}$. The exponential map plays the role of addition on a Riemannian manifold. This error structure has been considered on matrix Lie groups [1], Riemannian symmetric spaces [2] which is more general than Lie groups, and general Riemannian manifolds [5, 4], only to cite a few.

Received 28 March 2021

REFERENCES

[1] Chau, J. and von Sachs, R. (2020), "Intrinsic Wavelet Regression for Curves of Hermitian Positive Definite Matrices," *Journal of the American Statistical Association*, 0, 1–14.

[2] Cornea, E., Zhu, H., Kim, P., and Ibrahim, J. G. (2017), "Regression Models on Riemannian Symmetric Spaces," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 463–482. MR3611755

[3] Dai, X. (2020), "Statistical Inference on the Hilbert Sphere with Application to Random Densities," *ArXiv*.

[4] Dai, X., Lin, Z., and Müller, H.-G. (2020), "Modeling Longitudinal Data on Riemannian Manifolds," *Biometrics*, Accepted.

[5] Dai, X. and Müller, H.-G. (2018), "Principal Component Analysis for Functional Data on Riemannian Manifolds and Spheres," *Annals of Statistics*, 46, 3334–3361. MR3852654

[6] Fréchet, M. (1948), "Les Éléments Aléatoires de Nature Quelconque Dans Un Espace Distancié," *Annales de l'Institut Henri Poincaré*, 10, 215–310. MR0027464

Xiongtao Dai
Department of Statistics, Iowa State University, Ames, Iowa 50011
USA
E-mail address: xdai@iastate.edu