

STATISTICAL INFERENCE FOR FUNCTIONAL DATA OVER MULTI-DIMENSIONAL DOMAIN

Qirui Hu and Lijian Yang*

Shanghai University of Finance and Economics and Tsinghua University

Abstract: This work develops inference tools for the mean function of functional data over a multi-dimensional domain. A two-step mean estimator based on tensor product spline estimates of individual trajectories is shown oracally efficient, i.e., it is asymptotically indistinguishable from the infeasible estimator using unobservable trajectories. Consistent estimates of covariance function as well as exact quantile of the limiting maximal deviation are obtained by innovative use of results on sharp comparison of Gaussian extreme distributions and quantiles, leading to asymptotic coverage and order $n^{-1/2}$ uniformly adaptive width of data-driven simultaneous confidence regions (SCRs). Also formulated are one-sided SCRs that can be used for testing against uniform upper and lower bound of the mean function. Extensive Monte Carlo experiments corroborate the theory, and a satellite ocean dataset collected by the Copernicus Marine Environment Monitoring Service illustrates how the proposed SCR is used.

Key words and phrases: Exact quantile, functional data, Gaussian approximation, simultaneous confidence region, tensor product spline.

1. Introduction

Functional data analysis (FDA) has been an important area of statistics research for over three decades. Functional data consist of observations of stochastic processes such as Electrocardiogram (ECG), Electroencephalogram (EEG), and human growth; see Ramsay and Dalzell (1991), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Hsing and Eubank (2015). The goal of FDA is to secure crucial information about the stochastic process.

There is already a rich collection of works on statistical inference of mean, covariance functions and related structures of functional data, such as Degras (2011), Cao, Yang and Todem (2012), Gu et al. (2014), Zheng, Yang and Härdle (2014), Cao et al. (2016), Wang et al. (2020a), Huang, Zheng and Yang (2022), Li and Yang (2023), Zhong and Yang (2023), Cai and Hu (2024, 2025, 2026), and Hu and Li (2025). This body of work systematically develops various types of simultaneous confidence region (SCR), which is for an unknown function what a confidence interval is for an unknown parameter. Nonparametric SCRs are powerful tools for making global and uniform inference on unknown functions,

*Corresponding author. E-mail: yanglijian@tsinghua.edu.cn