Statistica Sinica **35** (2025), 1-19
doi:https://doi.org/10.5705/ss.202022.0087

THE IMPORTANCE OF BEING A BAND: FINITE-SAMPLE EXACT DISTRIBUTION-FREE PREDICTION SETS FOR FUNCTIONAL DATA

Jacopo Diquigiovanni*[1], Matteo Fontana[2,3] and Simone Vantini[2]

[1] *University of Padova,* [2] *Politecnico di Milano and* [3] *Royal Holloway, University of London*

Abstract: Functional Data Analysis represents a field of growing interest in statistics. Despite several studies have been proposed leading to fundamental results, the problem of obtaining valid and efficient prediction sets has not been thoroughly covered. Indeed, the great majority of methods currently in the literature rely on strong distributional assumptions (e.g., Gaussianity), dimension reduction techniques and/or asymptotic arguments. We propose a new nonparametric approach in the field of Conformal Prediction, based on a new family of nonconformity measures inducing conformal predictors able to create closed-form finite-sample valid or exact prediction sets for functional data under very minimal distributional assumptions. In addition, our proposal ensures that the prediction sets obtained are bands, an essential feature in the functional setting that allows the visualization and interpretation of such sets. The procedure is also fast, scalable, does not rely on functional dimension reduction techniques and allows the user to select different nonconformity measures depending on the problem at hand always obtaining valid bands. Within this family of measures, we propose also a specific measure leading to prediction bands asymptotically no less efficient than those with constant width.

Key words and phrases: Conformal Prediction, distribution-free prediction, exact prediction set, functional data, prediction band, uncertainty quantification.

1. Introduction

One of the main roles of statistics in our new, data-rich world is to provide scientists, businesspeople, and policymakers with tools able to deal with an increasing amount of data of increasing complexity. Automated sensor arrays and measuring systems now provide huge quantities of high-frequency and high-dimensional data about all sorts of social or physical phenomena.

Among the most popular toolboxes that have the capacity to deal with this kind of complex data, one can find Functional Data Analysis (FDA, Ramsay and Silverman, 2005). FDA is an ebullient field of statistics which aim is to develop theory and methods to deal with data sets made of functions defined over a domain, either uni- or multidimensional, and usually characterized by some

*Corresponding author. E-mail: jacopodiquigiovanni93@gmail.com