Statistica Sinica **32** (2022), 1-25
doi:https://doi.org/10.5705/ss.202020.0084

ROBUST SMOOTHED CANONICAL CORRELATION ANALYSIS FOR FUNCTIONAL DATA

Graciela Boente[1,3] and Nadia L. Kudraszow[2,3]

[1] *Universidad de Buenos Aires,* [2] *Universidad Nacional de La Plata and* [3] *CONICET, Argentina*

Abstract: We provide robust estimators for the first canonical correlation and directions of random elements on Hilbert separable spaces by using robust association and scale measures, combined with basis expansions and/or penalizations as a regularization tool. Under regularity conditions, the resulting estimators are consistent. The finite-sample performance of our proposal is illustrated by means of a simulation study that shows that, as expected, the robust method outperforms the existing classical procedure when the data are contaminated. A real data example is also presented.

Key words and phrases: Canonical correlation analysis, functional data, robust estimation, smoothing techniques.

1. Introduction

In recent years, data collected in the form of functions or curves have received considerable attention in fields such as chemometrics, economics, environmental studies, image recognition, spectroscopy, and many others. These data are known in the literature as functional data; see Ramsay and Silverman (2005) for a complete overview. As is well known, functional data are intrinsically infinite-dimensional, and this structure is a source of information. Therefore, even when recorded at a finite grid of points, functional observations should be considered as random elements of some functional space rather than as multivariate observations. In this manner, some of the theoretical and numerical challenges posed by the high dimensionality may be solved. This framework led to the extension of some classical multivariate analysis concepts, such as dimension-reduction techniques, to the context of functional data, usually using some regularization tool.

This paper focuses on canonical correlation analysis, where data consist of pairs of random curves. The aim of this analysis is to identify and quantify the relation between the observed functions. Under a Gaussian model,

Corresponding author: Nadia L. Kudraszow, Departamento de Matemática, Universidad Nacional de La Plata, CMaLP-CONICET, (1900) La Plata, Argentina. E-mail: nkudraszow@mate.unlp.edu.ar.