Topics in Functional Data Analysis

Commentary on the Habilitation Thesis
Masaryk University

David Kraus

August 2021

This habilitation thesis presents recent developments of statistical methodology driven by applications that I contributed to. This work belongs to functional data analysis, which is an active area of statistics that deals with data sets consisting of data units that are observations of mathematical functions such as curves, surfaces or images.

This text comprises five research articles containing my and my co-authors' contributions to the field accompanied by an introductory section which summarizes the contents of each paper. The papers included in the appendix of the thesis are:

(A) Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, 105(490):670–682.

(B) Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99(4):813–832.

(C) Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):777–801.

(D) Kraus, D. and Stefanucci, M. (2019). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika*, 106(1):161–180.

(E) Kraus, D. (2019). Inferential procedures for partially observed functional data. *Journal of Multivariate Analysis*, 173:583–603

These papers have all been published in high ranking journals in the field of methodological statistics. Two papers (C, E) are single-authored, the other three are collaborative with equal contribution of each co-author. The papers have been published with peer-reviewed supplements, which are included as well.

The core of each paper is the development of statistical methods to address problems that arise in applications in areas such as biology and medicine. This methodological

development is accompanied by the development of the relevant theory, computational methods and numerical studies. Motivated by a research problem in molecular biology, Paper A focuses on inference on the second-order structure of random processes. It proposes and analyzes two-sample tests for comparing covariance operators of functional data. Paper B addresses this objective in the setting of functional data contaminated by atypical observations. It introduces dispersion operators and develops resistant two-sample tests. Paper C deals with functional data that are incompletely observed, with motivation coming from a public health project. In the setting of functional fragments, it develops estimators of basic characteristics and proposes methods for reconstructing functions from their partial observation. Paper D focuses on optimal binary classification of functional data. It proposes regularized linear classifiers and develops them in the partial observation setting, which is motivated by a medical data set. Paper E addresses the problem of mean and covariance inference with missing functional data. It provides confidence intervals and hypothesis tests, along with theory and algorithms.