UNIFIED TESTS FOR NONPARAMETRIC FUNCTIONS IN RKHS WITH KERNEL SELECTION AND REGULARIZATION

Tao He[1], Ping-Shou Zhong[2], Yuehua Cui[3] and Vidyadhar Mandrekar[3]

[1]*San Francisco State University,* [2]*University of Illinois at Chicago and* [3]*Michigan State University*

Abstract: This study develops a unified test procedure for nonparametric functions in a reproducing kernel Hilbert space of high-dimensional or functional covariates. The test procedure is simple, computationally efficient, and practical because we do not need to distinguish high-dimensional or functional covariates. We derive the asymptotic distributions of the proposed test statistic under the null and a series of local alternative hypotheses. The asymptotic distributions depend on the decay rate of the eigenvalues of the kernel function. This decay rate is determined by the kernel function and the types of covariates. We also develop a novel kernel selection procedure to maximize the power of the proposed test by maximizing the signal-to-noise ratio. The proposed kernel selection procedure is shown to be consistent in selecting the kernels that maximize the power function. Moreover, a test with a regularized kernel is constructed to further improve the power. It is shown that the proposed test nearly achieves the power of an oracle test if the regularization parameter is properly chosen. Extensive simulation studies evaluate the finite-sample performance of the proposed method. Finally, we apply the proposed method to a Yorkshire gilt data set to identify pathways that are associated with the triiodothyronine level. The proposed methods are included in an R package "KerUTest."

Key words and phrases: Gene set analysis, kernel function, nonparametric regression, reproducing kernel Hilbert space.

1. Introduction

High-dimensional or functional data arise in a wide range of areas, including biology, imaging, and climate. In genetic studies, millions of single nucleotide polymorphisms (SNPs) can be measured simultaneously using high-throughput technologies. The identification of genes that are associated with certain traits, such as blood pressure and grain yield, is becoming increasingly important in health and agriculture sciences. Although the traditional methods focus on a

Corresponding author: Ping-Shou Zhong, Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, USA. E-mail: pszhong@uic.edu.