



# Statistical inference using regularized M-estimation in the reproducing kernel Hilbert space for handling missing data

Hengfang Wang<sup>1</sup> · Jae Kwang Kim<sup>2</sup>

Received: 11 November 2022 / Revised: 5 February 2023 / Accepted: 20 February 2023 /

Published online: 27 April 2023

© The Institute of Statistical Mathematics, Tokyo 2023

## Abstract

Imputation is a popular technique for handling missing data. We address a nonparametric imputation using the regularized M-estimation techniques in the reproducing kernel Hilbert space. Specifically, we first use kernel ridge regression to develop imputation for handling item nonresponse. Although this nonparametric approach is potentially promising for imputation, its statistical properties are not investigated in the literature. Under some conditions on the order of the tuning parameter, we first establish the root- $n$  consistency of the kernel ridge regression imputation estimator and show that it achieves the lower bound of the semiparametric asymptotic variance. A nonparametric propensity score estimator using the reproducing kernel Hilbert space is also developed by the linear expression of the projection estimator. We show that the resulting propensity score estimator is asymptotically equivalent to the kernel ridge regression imputation estimator. Results from a limited simulation study are also presented to confirm our theory. The proposed method is applied to analyze air pollution data measured in Beijing, China.

**Keywords** Imputation · Kernel ridge regression · Missing at random · Propensity score

---

✉ Jae Kwang Kim  
jkim@iastate.edu

<sup>1</sup> School of Mathematics and Statistics, Fujian Normal University, No. 8 Xuefu South Road, Shangjie, Fuzhou 350117, Fujian, China

<sup>2</sup> Department of Statistics, Iowa State University, 2438 Osborn Dr., Ames, IA 50011, USA