Annals of the Institute of Statistical Mathematics (2022) 74:539–557
https://doi.org/10.1007/s10463-021-00806-2

Robust model selection with covariables missing at random

Zhongqi Liang[1] · **Qihua Wang**[1,2] · **Yuting Wei**[3]

Received: 30 November 2020 / Revised: 21 July 2021 / Accepted: 3 August 2021 /
Published online: 25 August 2021

Abstract

Let $f_{Y|X,Z}(y|x,z)$ be the conditional probability function of Y given (X, Z), where Y is the scalar response variable, while (X, Z) is the covariable vector. This paper proposes a robust model selection criterion for $f_{Y|X,Z}(y|x,z)$ with X missing at random. The proposed method is developed based on a set of assumed models for the selection probability function. However, the consistency of model selection by our proposal does not require these models to be correctly specified, while it only requires that the selection probability function is a function of these assumed selective probability functions. Under some conditions, it is proved that the model selection by the proposed method is consistent and the estimator for population parameter vector is consistent and asymptotically normal. A Monte Carlo study was conducted to evaluate the finite-sample performance of our proposal. A real data analysis was used to illustrate the practical application of our proposal.

Keywords Model selection · Inverse probability weight · Model misspecification · Missing at random · Kullback–Leibler divergence · Robust

✉ Qihua Wang
qhwang@amss.ac.cn

Zhongqi Liang
2945155436@qq.com

Yuting Wei
ytwei@mail.ustc.edu.cn

[1] School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, Zhejiang, China

[2] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[3] Department of Statistics and Finance, Universtiy of Science and Technology of China, Hefei 230026, China