Statistica Sinica **35** (2025), 1-23
doi:https://doi.org/10.5705/ss.202021.0250

SUFFICIENT VARIABLE SCREENING FOR ULTRAHIGH-DIMENSIONAL RIGHT CENSORED DATA VIA INDEPENDENCE MEASURES

Baoying Yang*[1], Qingcong Yuan[2,3] and Xiangrong Yin[4]

[1]*Southwest Jiaotong University,* [2]*Miami University,*
[3]*Sanofi US and* [4]*University of Kentucky*

Abstract: We develop two sufficient variable screening procedures utilizing the newly proposed censored distance correlation measures for ultrahigh-dimensional right censored data. Compared to many existing methods, our procedures more effectively detect active predictors that are marginally independent of the response. They are also model-free and robust against model misspecification. Through simulations and real data analysis, we demonstrate the distinct advantages of our proposed procedures over existing variable screening methods.

Key words and phrases: Distance correlation, feature screening, independence measure.

1. Introduction

Variable screening has become increasingly important in various research fields. The renaissance of variable screening began with the sure independent screening (SIS) method of Fan and Lv (2008), which is based on the marginal Pearson correlation and is specifically tailored for linear regressions with Gaussian predictors and responses. Following the pioneering work of SIS, many methods have been proposed over the last two decades, using either model-specific or model-free approaches for ultrahigh-dimensional data. See Fan and Song (2010), Liu, Li and Wu (2014), Chang, Tang and Wu (2013), Zhu et al. (2011), Li, Zhong and Zhu (2012), Mai and Zou (2015), and Shao and Zhang (2014), among many others.

In the context of ultrahigh-dimensional survival data analysis, the response is the time to an event that is often subject to censoring. Censoring brings more difficulties and challenges for the feature screening of ultrahigh-dimensional data. To address these complexities, a variety of ultrahigh-dimensional screening techniques designed to survival outcomes have flourished. For example, Fan, Feng and Wu (2010) investigated the SIS method for the Cox proportional hazards model by ranking variables according to their respective univariate partial log-likelihoods (ISIS_C). He, Wang and Hong (2013) developed a quantile-adaptive

*Corresponding author. E-mail: yangbaoying@swjtu.edu.cn