Statistica Sinica **36** (2026), 1-25
doi:https://doi.org/10.5705/ss.202022.0346

VARIABLE SCREENING VIA CONDITIONAL MARTINGALE DIFFERENCE DIVERGENCE

Lei Fang*[1], Qingcong Yuan[2], Xiangrong Yin[3] and Chenglong Ye[3]

[1]*Miami University,* [2]*Sanofi and* [3]*University of Kentucky*

Abstract: Variable screening has been a useful research area that deals with ultrahigh-dimensional data. When there exist both marginally and jointly dependent predictors to the response, existing methods such as conditional screening or iterative screening often suffer from instability against the selection of the conditional set or the computational burden, respectively. In this article, we propose a new independence measure, named **c**onditional **m**artingale difference **d**ivergence ($\text{CMD}_{\mathcal{H}}$), that can be treated as either a conditional or a marginal independence measure. Under regularity conditions, we show that the *sure screening property* of $\text{CMD}_{\mathcal{H}}$ holds for both marginally and jointly active variables. Based on this measure, we propose a kernel-based model-free variable screening method, which is efficient, flexible, and stable against high correlation among predictors and heterogeneity of the response. In addition, we provide a data-driven method to select the conditional set. In simulations and real data applications, we demonstrate the superior performance of the proposed method.

Key words and phrases: Conditional screening, dimension reduction, independence measure, reproducing kernel Hilbert space, sure screening property.

1. Introduction

Variable screening has been a research area that deals with ultrahigh-dimensional data, where high-dimensional methods may fail due to the curse of dimensionality, as Fan, Samworth and Wu (2009) suggested. Fan and Lv (2008)'s seminal work suggests to screen the ultrahigh-dimensional data before conducting variable selection. They proposed a sure independent screening (SIS) method for linear models to screen out inactive variables based on Pearson correlation. After that, variable screening receives more attention since it only requires that the selected set of variables covers the set of active variables, which is referred to as the *sure screening property* (Fan and Lv, 2008). Screening methods with this property suffer less from *instability* (Yu, 2013) that is seen in many variable selection methods.

Various model-based screening methods have been developed. For linear regression models, screening methods have been proposed based on different measures, including marginal Pearson correlation (Fan and Lv, 2008), forward

*Corresponding author. E-mail: fangl17@miamioh.edu