Statistica Sinica Preprint No: SS-2023-0290

Title	Two Kernel-based Feature Screening Procedures for High-dimensional Response Data
Manuscript ID	SS-2023-0290
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0290
Complete List of Authors	Yuke Shi, Na Li, Qizhai Li, Dongdong Pan and Jinjuan Wang
Corresponding Authors	Jinjuan Wang
E-mails	wangjinjuan@bit.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0290

Statistica Sinica

Two Kernel-based Feature Screening Procedures for High-dimensional Response Data

Yuke Shi[1,2], Na Li[3], Qizhai Li[1,2], Dongdong Pan[4], and Jinjuan Wang[5,*]

[1] *LSC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences.*

[2] *School of Mathematical Sciences, University of Chinese Academy of Sciences.*

[3] *School of Applied Science, Beijing Information Science and Technology University.*

[4] *School of Mathematics and Statistic, Yunnan University.*

[5] *School of Mathematics and Statistics, Beijing Institute of Technology.*

[*] *Corresponding author. E-mail: wangjinjuan@bit.edu.cn.*

Abstract:

We consider feature screening for high-dimensional response data without and with the existence of confounding factors. First, we introduce kernel covariance and kernel correlation for high-dimensional associaiton analysis, and further propose partial kernel covariance and partial kernel correlation that can handle situations with confounding factors. Then, based on the kernel correlation and partial kernel correlation, we propose two feature screening procedures. Both screening procedures possess sure screening property and ranking consistency property, and are complementary to each other by respectively dealing with situations without and with the existence of confounding factors. The proposed procedures make no assumptions on model, and are suitable for high-dimensional response variable and non-Euclidean data. Extensive simulation results and a real data analysis demonstrate the satisfying performances and advantages of the proposed procedures over existing methods.

Key words and phrases: Feature screening, high-dimensional response variable, kernel correlation, partial kernel correlation, confounding factors.

1. Introduction

High-dimensional responses are frequently encountered in feature screening analysis nowadays. For example, in both gene expression experiments and genetic pleiotropic association studies that often search for disease-associated genetic variants among hundreds of thousands of variants, the responses are usually high-dimensional, which are expression levels of tens of thousands of genes in the former analysis and are hundreds of human complex traits in the latter study, respectively (Gratten and Visscher, 2016; Zhang et al., 2017; Watanabe et al., 2019).

A good strategy to complete such searching is feature screening accompanied with association analysis. There exist many variable screening procedures dealing with diverse situations. For example, feature screening procedures for linear model (Fan and Lv, 2008), generalized linear model (Fan and Song, 2010), non-parametric additive model (Fan, Feng and Song, 2011), varying-coefficient model (Fan, Ma and Dai, 2014), semiparametric single-index model (Li, Zhong and Zhu, 2012), quantile regression model (He, Wang and Hong, 2013), and categorical data analysis (Cui, Li and Zhong, 2015; Xie et al., 2020) have been developed, among others. But these procedures can handle only one response at a time, and thus need to be repeatedly conducted in situations of multiple responses and high-dimensional responses. This repetition can cause substantial power loss, attributable not only to the requirements for multiple hypothesis testing and correction but also to its failure to employ on the association information among response variables.

To take the association information into consideration, some methods that can handle multiple responses have been proposed. For example, Li, Zhong and Zhu (2012) introduced the DC-SIS procedure for multiple responses based on distance correlation (Székely et al., 2007). But DC-SIS has been found to perform poorly for heavy-tailed data (Mai and Zou, 2015). Shao and Zhang (2014) proposed an adaptable approach to screen out variables that have limited influence on particular facets of the conditional distribution, such as conditional quantiles, based on the concept of martingale difference correlation (MDC). He, Zhou and Zou (2021) developed an mRCC procedure utilizing rank canonical correlation coefficients, which is specifically tailored for multiple responses.

Though the methods described above can handle multiple responses, they are not adequate for high-dimensional responses. So feature screening procedures that can deal with high-dimensional responses are much needed. Among techniques dealing with high-dimensional data, kernel-based methods have proven to be powerful since they can efficiently capture dependence among random variables by mapping them into a felicitous reproducing kernel Hilbert space (RKHS).

Therefore, we investigate and propose new feature screening methods from the perspective of kernel functions for high-dimensional response data. We introduce the kernel correlation (kCor) coefficient, and propose a feature screening procedure named KC-SIS, to screen the vital predictors that are potentially related to the high-dimensional responses of interest. Additionally, to deal with situations where confounding factors exist, we further propose the partial kernel correlation (pkCor) coefficient, and develop a new pkCor-based feature screening procedure named PKC-SIS.

Both proposed screening procedures exhibit the following merits. First, KC-SIS and PKC-SIS can handle not only ultrahigh-dimensional predictors but also high-dimensional responses. Second, KC-SIS and PKC-SIS possess sure screening property and ranking consistency property, which guarantee that potentially response-related features are obtained with high probability. Third, KC-SIS and PKC-SIS are non-parametric screening procedures that do not specify any parameter models, making them robust to data outliers and model misspecification. Fourth, the implementation of KC-SIS and PKC-SIS is simple and convenient for practical application. Last but not least, both KC-SIS and PKC-SIS exhibit applicability across various data types, including continuous, discrete, Euclidean, and non-Euclidean data, provided a kernel function is accessible.

The rest of the article is organized as follows. Section 2 provides the necessary preliminaries for kernel covariance and kernel correlation, and introduces the partial kernel covariance and partial kernel correlation coefficients. In Section 3, two screening procedures, KC-SIS and PKC-SIS, are proposed with their theoretical properties developed. Extensive simulations are conducted in Section 4 to demonstrate the performances of the proposed procedures. And an application of PKC-SIS to a heterogeneous stock mice dataset is conducted to show its practical application in Section 5. Section 6 demonstrates conclusion remarks.

2. Kernel-based Coefficients for Association Analysis

In this section, we provide a brief review of the kernel covariance (kCov) and the kernel correlation (kCor), which are used to capture the independence of high-dimensional random variables. Subsequently, we introduce the partial kernel covariance (pkCov) and the partial kernel correlation (pkCor), which take into account the presence of additional confounding factors. Throughout the article, let $|a|$ be the absolute value of $a \in \mathbb{R}$, $\mathbb{R}^+ = [0, \infty)$, and $||b|| = \sqrt{b^\top b}$ be the Euclidean norm of $b \in \mathbb{R}_l$, where $\mathbb{R}_l$ is the l-dimensional real number space, l is a positive integer and the superscript $^\top$ indicates the transpose of a vector or a matrix.

2.1 Kernel covariance and kernel correlation

Let U and V be two random variables taking values in two separable metric spaces $\mathbb{U}$ and $\mathbb{V}$, and $\mathcal{H}_{\mathbb{U}}$ and $\mathcal{H}_{\mathbb{V}}$ be the corresponding RKHSs defined by two measurable kernels k_ϕ and k_ψ, respectively, with $E\big\{k_\phi(U, U)\big\} < \infty$ and $E\big\{(k_\psi(V, V)\big\} < \infty$. The kCov of U and V is defined as the square root of $\mathcal{A}^2_{k_\phi,k_\psi}(U,V)$ which equals to

$$\begin{aligned}\mathcal{A}^2_{k_\phi,k_\psi}(U,V) &= E_{U,V}E_{\widetilde{U},\widetilde{V}}\{k_\phi(U,\widetilde{U})k_\psi(V,\widetilde{V})\}\\ &\quad + E_U E_{\widetilde{U}}\{k_\phi(U,\widetilde{U})\}E_V E_{\widetilde{V}}\{k_\psi(V,\widetilde{V})\}\\ &\quad -2E_{U,V}\big[E_{\widetilde{U}}\{k_\phi(U,\widetilde{U})\}E_{\widehat{V}}\{k_\psi(V,\widehat{V})\}\big],\end{aligned}$$

where $(\widetilde{U},\widetilde{V})$ and $(\widehat{U},\widehat{V})$ are independent copies of (U,V) (Sejdinovic et al., 2013).

If k_ϕ and k_ψ are characteristic kernels, the corresponding kCov is a measure of independence, i.e., $\mathcal{A}^2_{k_\phi,k_\psi}(U,V) = 0$ if and only if U and V are independent (Lyons, 2013). As shown in Fukumizu et al. (2008), the Gaussian kernel $k_{\mathtt{G}}(a,\tilde{a}) = \exp\big(-\frac{1}{2\sigma^2}||a-\tilde{a}||^2\big)$ $(\sigma>0)$, Laplacian kernel $k_{\mathtt{L}}(a,\tilde{a}) = \exp\big(-\lambda\sum_{i=1}^{l}|a_i-\tilde{a}_i|\big)$ $(\lambda>0)$, and another two kernels with $k_{\mathtt{I}}(b,\tilde{b}) = \prod_{i=1}^{l}\big(b_i+\tilde{b}_i+\alpha\big)^{-1}$ $(\alpha>0)$ and $k_{\mathtt{II}}(b,\tilde{b}) = \exp\big\{-\beta\sum_{i=1}^{l}(b_i+\tilde{b}_i)^{1/2}\big\}$ $(\beta>0)$ are all characteristic kernels, where $a=(a_1,\ldots,a_l)^\top\in\mathbb{R}_l$, $\tilde{a}=(\tilde{a}_1,\ldots,\tilde{a}_l)^\top\in\mathbb{R}_l$, $b=(b_1,\ldots,b_l)^\top\in\mathbb{R}_l^+$, $\tilde{b}=(\tilde{b}_1,\ldots,\tilde{b}_l)^\top\in\mathbb{R}_l^+$, with $\mathbb{R}_l^+$ being the l-dimensional positive real number space.

The kernel variance is defined as $\mathcal{A}^2_{k_\phi}(U) = \mathcal{A}^2_{k_\phi,k_\phi}(U,U)$, and the squared kCor is

$$\mathcal{R}^2_{k_\phi,k_\psi}(U,V) = \begin{cases}\dfrac{\mathcal{A}^2_{k_\phi,k_\psi}(U,V)}{\sqrt{\mathcal{A}^2_{k_\phi}(U)\mathcal{A}^2_{k_\psi}(V)}}, & \mathcal{A}^2_{k_\phi}(U)\mathcal{A}^2_{k_\psi}(V)>0,\\ 0, & \mathcal{A}^2_{k_\phi}(U)\mathcal{A}^2_{k_\psi}(V)=0,\end{cases}$$

Then based on the Cauchy-Schwarz inequality, we have $0 \leqslant \mathcal{R}^2_{k_\phi,k_\psi}(U,V) \leqslant 1$.

Comparable to the distance correlation introduced by Szekely (2007), which derives from Euclidean distance computations, kCor provides an readily computable sample analog. Moreover, kCor demonstrates a heightened versatility in its applicability, particularly when quantifying independence among random variables that traverse complex topological domains.

Denote n independent observations of (U,V) as $(\boldsymbol{u}_i, \boldsymbol{v}_i)$ with $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ip})^\top$ and $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{im})^\top$, $i = 1, \ldots, n$. Calculate two kernel matrices $G_{k_\phi} = (g_{k_\phi,il})_{n\times n} = \big\{k_\phi(\boldsymbol{u}_i, \boldsymbol{u}_l)\big\}_{n\times n}$ and $S_{k_\psi} = (s_{k_\psi,il})_{n\times n} = \big\{k_\psi(\boldsymbol{v}_i, \boldsymbol{v}_l)\big\}_{n\times n}$, $i,j = 1, \ldots, n$. Following the definition in Székely and Rizzo (2014), the U-centered version of $\tilde{G}_{k_\phi} = (\tilde{g}_{k_\phi,il})_{n\times n}$ can be obtained through

$$\tilde{g}_{k_\phi,il} = \begin{cases} g_{k_\phi,il} - \frac{1}{n-2}\sum\limits_{\substack{l=1\\ l\neq i}}^{n} g_{k_\phi,il} - \frac{1}{n-2}\sum\limits_{\substack{i=1\\ i\neq l}}^{n} g_{k_\phi,il} + \frac{2}{(n-1)(n-2)}\sum\limits_{1\leqslant i<l\leqslant n} g_{k_\phi,il}, & i \neq l, \\ 0, & i = l. \end{cases}$$

And the U-centered kernel matrix $\tilde{S}_{k_\psi} = (\tilde{s}_{k_\phi,il})_{n\times n}$ can be calculated similarly. Then a good estimator of kCov is given by

$$\widehat{\mathcal{A}}^2_{k_\phi,k_\psi}(U,V) = \frac{2}{n(n-3)}\sum_{1\leqslant i<l\leqslant n} \tilde{g}_{k_\phi,il}\tilde{s}_{k_\psi,il}.$$

and the squared sample kCor can be naturally defined by

$$\widehat{\mathcal{R}}^2_{k_\phi,k_\psi}(U,V) = \begin{cases} \dfrac{\widehat{\mathcal{A}}^2_{k_\phi,k_\psi}(U,V)}{\sqrt{\widehat{\mathcal{A}}^2_{k_\phi}(U)\widehat{\mathcal{A}}^2_{k_\psi}(V)}}, & \widehat{\mathcal{A}}^2_{k_\phi}(U)\widehat{\mathcal{A}}^2_{,k_\psi}(V) > 0, \\ 0, & \widehat{\mathcal{A}}^2_{k_\phi}(U)\widehat{\mathcal{A}}^2_{k_\psi}(V) = 0, \end{cases}$$

where $\widehat{\mathcal{A}}^2_{k_\phi}(U) = \widehat{\mathcal{A}}^2_{k_\phi}(U,U)$ is the sample kernal variance. Similar to the properties of distance covariance and distance correlation proposed in Székely and Rizzo (2014), we have the following proposition.

Proposition 1. Assume $E\big\{k_\phi(U,U)\big\} < \infty$ and $E\big\{k_\psi(V,V)\big\} < \infty$, then the estimators have the following properties:

$$\lim_{n\to\infty} \widehat{\mathcal{A}}^2_{k_\phi,k_\psi}(U,V) = \mathcal{A}^2_{k_\phi,k_\psi}(U,V),\ \lim_{n\to\infty} \widehat{\mathcal{R}}^2_{k_\phi,k_\psi}(U,V) = \mathcal{R}^2_{k_\phi,k_\psi}(U,V) \text{ almost surely;} \tag{2.1}$$

$$0 \leqslant \widehat{\mathcal{R}}^2_{k_\phi,k_\psi}(U,V) \leqslant 1. \tag{2.2}$$

2.2 Partial kernel variance and partial kernel correlation

Now we consider measuring the independence between U and V with the existence of confounding variable W. Suppose the random variable W takes values in the separable metric spaces $\mathbb{W}$, and the measurable positive definite kernels k_η of the corresponding RKHS $\mathcal{H}_\mathbb{W}$ satisfies $E\big\{k_\eta(W,W)\big\} < \infty$. Inspired by the idea of partial distance covariance proposed by Székely and Rizzo (2014) which is initially rooted in Euclidean distance, we propose partial kernel covariance (pkCov) that applies to diverse kernel functions, to encompass more flexibility. The extended notion pkCov is defined as follows:

$$\mathcal{B}_{k_\phi,k_\psi;k_\eta}(U,V;W)=\begin{cases}\mathcal{A}^2_{k_\phi,k_\psi}(U,V)-\dfrac{\mathcal{A}^2_{k_\phi,k_\eta}(U,W)\mathcal{A}^2_{k_\psi,k_\eta}(V,W)}{\mathcal{A}^2_{k_\eta}(W)}, & \mathcal{A}^2_{k_\eta}(W)>0,\\ \mathcal{A}^2_{k_\phi,k_\psi}(U,V), & \mathcal{A}^2_{k_\eta}(W)=0.\end{cases}$$

And the corresponding pkCor defined as

$$\mathcal{Q}_{k_\phi,k_\psi;k_\eta}(U,V;W)=$$
$$\begin{cases}\dfrac{\mathcal{R}^2_{k_\phi,k_\psi}(U,V)-\mathcal{R}^2_{k_\phi,k_\eta}(U,W)\mathcal{R}^2_{k_\psi,k_\eta}(V,W)}{\sqrt{1-\mathcal{R}^4_{k_\phi,k_\eta}(U,W)}\sqrt{1-\mathcal{R}^4_{k_\psi,k_\eta}(V,W)}}, & \mathcal{R}^2_{k_\phi,k_\eta}(U,W)\neq 1 \text{ and } \mathcal{R}^2_{k_\psi,k_\eta}(V,W)\neq 1,\\ 0, & \mathcal{R}^2_{k_\psi,k_\eta}(U,W)=1 \text{ or } \mathcal{R}^2_{k_\psi,k_\eta}(V,W)=1.\end{cases}$$

The pkCor conceptually equals to kCor when U (or V) and W are independent, which will be validated later in the simulation section.

The sample pkCov and pkCor can be defined as follows. Denote n independent observations of W as $\boldsymbol{w}_i=(w_{i1},\ldots,w_{iq})^\top$, $i=1,\ldots,n$, and the corresponding kernel matrix as $S_{k_\eta}=(s_{k_\eta,il})_{n\times n}=\big\{k_\eta(\boldsymbol{w}_i,\boldsymbol{w}_l)\big\}_{n\times n}$. Then the sample pkCov is defined by

$$\widehat{\mathcal{B}}_{k_\phi,k_\psi;k_\eta}(U,V;W)=\begin{cases}\widehat{\mathcal{A}}^2_{k_\phi,k_\psi}(U,V)-\dfrac{\widehat{\mathcal{A}}^2_{k_\phi,k_\eta}(U,W)\widehat{\mathcal{A}}^2_{k_\psi,k_\eta}(V,W)}{\widehat{\mathcal{A}}^2_{k_\eta}(W)}, & \widehat{\mathcal{A}}^2_{k_\eta}(W)>0,\\ \widehat{\mathcal{A}}^2_{k_\phi,k_\psi}(U,V), & \widehat{\mathcal{A}}^2_{k_\eta}(W)=0,\end{cases}$$

and the sample pkCor is

$$\widehat{\mathcal{Q}}_{k_\phi,k_\psi;k_\eta}(U,V;W) =$$

$$\begin{cases} \dfrac{\widehat{\mathcal{R}}^2_{k_\phi,k_\psi}(U,V) - \widehat{\mathcal{R}}^2_{k_\psi,k_\eta}(U,W)\widehat{\mathcal{R}}^2_{k_\psi,k_\eta}(V,W)}{\sqrt{1-\widehat{\mathcal{R}}^4_{k_\phi,k_\eta}(U,W)}\sqrt{1-\widehat{\mathcal{R}}^4_{k_\psi,k_\eta}(V,W)}}, & \widehat{\mathcal{R}}^2_{k_\psi,k_\eta}(U,W) \neq 1 \text{ and } \widehat{\mathcal{R}}^2_{k_\psi,k_\eta}(V,W) \neq 1, \\ 0, & \widehat{\mathcal{R}}^2_{k_\psi,k_\eta}(U,W) = 1 \text{ or } \widehat{\mathcal{R}^2}_{k_\psi,k_\eta}(V,W) = 1. \end{cases}$$

The following proposition demonstrates the consistency of these estimators.

Proposition 2. Suppose $E(k_\phi(U,U)) < \infty$, $E(k_\psi(V,V)) < \infty$, and $E(k_\eta(W,W) < \infty$, then the estimators have the following properties:

$$\lim_{n\to\infty} \widehat{\mathcal{B}}^2_{k_\phi,k_\psi}(U,V) = \mathcal{B}^2_{k_\phi,k_\psi}(U,V) and \lim_{n\to\infty} \widehat{\mathcal{Q}}^2_{k_\phi,k_\psi}(U,V) = \mathcal{Q}^2_{k_\phi,k_\psi}(U,V) \text{ almost surely}; \tag{2.3}$$

$$0 \leqslant \widehat{\mathcal{Q}}^2_{k_\phi,k_\psi}(U,V) \leqslant 1. \tag{2.4}$$

3. Sure Independence Screening Procedures

In this section, a model-free marginal screening procedure based on kCor named KC-SIS and a partial marginal screening procedure based on pkCor named PKC-SIS are proposed to screen out irrelevant variables effectively.

3.1 Screening procedure based on kCor

Let $X = (X_1, \ldots, X_p)^\top$ be a p-dimensional predictor with support $\mathcal{X}$ and $Y = (Y_1, \ldots, Y_m)^\top$ be an m-dimensional response variable with support $\mathcal{Y}$.

We consider the problem of feature screening when both p and m diverge with sample size n, where the dimension of predictors p can far exceed n. Define the conditional distribution function of Y given X as $F(Y|X)$. And denote the index set for active predictors of the true active model as

$$\mathcal{G} = \big[j : F(Y|X) \text{ functionally depends on } X_j, j = 1, \dots, p\big\},$$

whose cardinality is $s = |\mathcal{G}|$, and denote $\bar{\mathcal{G}} = \{1, 2, \cdots, p\}/\mathcal{G}$ to be the index set for inactive predictors. To screen out inactive predictors, for $j = 1, \cdots, p$, denote

$$R_j = \mathcal{R}^2_{k_\phi, k_\psi}(X_j, Y)$$

for any given kernels k_ϕ and k_ψ. And their estimators can be calculated as

$$\widehat{R}_j = \widehat{\mathcal{R}}^2_{k_\phi, k_\psi}(X_j, Y), j = 1, \cdots, p.$$

Then predictors with corresponding estimators $\widehat{R}_j \geq c_*$ can be treated as active predictors, where $c_* = cn^{-\tau}$ is a preassigned threshold value defined later in the Assumption 2. Since c_* is unknown, we propose the following KC-SIS procedure to screen out inactive predictors with a given constant L:

(1) Select two suitable kernels k_ϕ and k_ψ, and an model size integer L which is suggested to be $L = n/\log(n)$ in literature;

(2) Use the observations to calculate $\widehat{R}_1, \ldots, \widehat{R}_p$;

(3) Choose X_j with the top L largest values among $\widehat{R}_1, \ldots, \widehat{R}_p$.

To investigate the statistical properties of KC-SIS, we make the following three assumptions.

Assumption 1. There exist positive constants $t' > 0$ and $\varpi > 0$ such that for all $0 < t \leqslant t'$,

$$\sup_p \max_{1 \leqslant j \leqslant p} E\Big[\exp\big\{k_\phi^{2\varpi}(X_j, X_j)t\big\}\Big] < +\infty$$

and

$$\sup_m E\Big[\exp\big\{k_\psi^{2\varpi}(Y, Y)t\big\}\Big] < +\infty.$$

Assumption 2. There exist positive constants $c > 0$ and $0 < \tau < 1/2$ such that for any $j = 1, \cdots, p$,

$$\min_{j \in \mathcal{G}} R_j \geqslant 2cn^{-\tau}.$$

Assumption 3. There exists positive constant $c' > 0$ such that

$$\liminf_{p \to \infty} \inf_m \big[\min_{j \in \mathcal{G}} R_j - \max_{j' \in \bar{\mathcal{G}}} R_{j'}\big\} \geqslant c'.$$

Remark 1. Assumption 1 is explicitly satisfied when utilizing bounded characteristic kernels, such as the Gaussian kernel k_{G} and Laplacian kernel

k_{L}. In comparison to assumptions underlying DC-SIS in Li, Zhong and Zhu (2012), the condition posited in Assumption 1 is less stringent. Therefore, the KC-SIS method can effectively handle heavy-tailed variables by compressing them into finite values via these bounded characteristic kernels. According to Assumption 2, the minimum association signal of active features cannot be too small, which is widely used in variable screening studies (Fan and Lv, 2008; Li, Zhong and Zhu, 2012). Assumption 3 indicates that the marginal utility R_j can distinguish active predictors from inactive ones immaculately. And it is a comparatively weaker condition than the partial orthogonality condition which assumes $R_j > 0$ for $j \in \mathcal{G}$ and $R_j = 0$ for $j \in \bar{\mathcal{G}}$ (Huang, Horowitz and Ma, 2008).

The following theorem presents the sure screening property and ranking consistency property of KC-SIS, whose proofs are given in the supplementary material.

Theorem 1. *(1) Under Assumption 1, for any $\tau > 0$, $\gamma > 0$ and $1 - 2\gamma - 2\tau > 0$, there exist positive constants c_1, c_2, c_3, and c_4 such that*

$$P\Big(\max_{1\leqslant j\leqslant p}\big|\widehat{R}_j - R_j\big| \geqslant cn^{-\tau}\Big) \leqslant c_1 p \exp\big(-c_2 n^{1-2\gamma-2\tau}\big) + c_3 p \exp\big(-c_4 n^{\gamma\varpi}\big),$$

when n is sufficiently large.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0290

(2) Under Assumptions 1,2 and 3,

$$P(\mathcal{G} \subseteq \widehat{\mathcal{G}}) \geqslant 1 - sc_1 \exp\big(-c_2 n^{1-2\gamma-2\tau}\big) - sc_3 \exp\big(-c_4 n^{\gamma\varpi}\big).$$

where $\widehat{\mathcal{G}} = \big\{j : \widehat{R}_j \geq cn^{-\tau}, j = 1, \ldots, p\big\}$ *and* $s = |\mathcal{G}|$.

(3) [Ranking Consistency Property] Under Assumptions 1,2 and 3, for $\log(p) = o(n^{\nu})$ *with* $0 < \nu < \min\{1 - 2\gamma - 2\tau, \gamma\varpi\}$,

$$\liminf_{n\to\infty}\{\min_{j\in\mathcal{G}} \widehat{R}_j - \max_{j'\in\bar{\mathcal{G}}} \widehat{R}_{j'}\} \geqslant 0, a.s.$$

Remark 2. If we choose $\gamma = (1-2\tau)/(\varpi+2)$, the result (2) in Theorem 1 can be rewritten as $P(\mathcal{G} \subseteq \widehat{\mathcal{G}}) \leq c_5 p \exp\big\{-c_6 n^{\frac{\varpi}{\varpi+2}(1-2\tau)}\big\}$ for some constants c_5 and c_6. It shows that KC-SIS can handle non-polynomial dimensionality of order $\log p = o\big\{n^{\frac{\varpi}{\varpi+2}(1-2\tau)}\big\}$.

3.2 Screening procedure based on pkCor

Next, we propose the sure independent screening procedure based on pkCor. With the existence of confounding variables, define the index set for the true active predictors as

$$\mathcal{G}_z = \big\{j : \mathcal{Q}^2_{k_\phi,k_\psi;k_\eta}(X_j, Y; Z) \neq 0 \text{ for some } Y \in \mathcal{Y}, j = 1, \ldots, p\big\},$$

where size $s_z = |\mathcal{G}_z|$, and $Z = (Z_1, \ldots, Z_q)^\top$ is a q-dimensional random variable with support $\mathcal{Z}$. Denote $\bar{\mathcal{G}}_z = \{1, \cdots, p\}/\mathcal{G}_z$ as index set for the inactive predictors. Define

$$Q_j = \mathcal{Q}^2_{k_\phi,k_\psi;k_\eta}(X_j, Y; Z), \quad j = 1, \ldots, p.$$

Then predictor X_j can be treated as active predictor when

$$\widehat{Q}_j = \widehat{\mathcal{Q}}^2_{k_\phi,k_\psi;k_\eta}(X_j, Y; Z) \geq c_z n^{-\tau_z}, \quad j = 1, \ldots, p,$$

where $c_z n^{-\tau_z}$ is a preassigned threshold value defined later in Assumption 5. Then we propose the following PKC-SIS procedure with a given constant L to screen out irrelevant predictors:

(1) Select three suitable kernels k_ϕ, k_ψ and k_η, and set the model size as L which is recommended to be $L = n/\log(n)$;

(2) Use the sample to calculate $\widehat{Q}_1, \ldots, \widehat{Q}_p$;

(3) Choose X_j with the top L largest values among $\widehat{Q}_1, \ldots, \widehat{Q}_p$.

To verify the sure screening property, we present the following three regularity assumptions in addition to Assumptions 1-3.

Assumption 4. There exist positive constants $t'' > 0$ and $\varpi_z > 0$ such that for all $0 < t \leqslant t''$,

$$E\Big[\exp\big\{k_\eta^{2\varpi_z}(Z, Z)t\big\}\Big] < +\infty.$$

Assumption 5. There exist positive constants $c_z > 0$ and $0 < \tau_z < 1/2$, such that

$$\min_{j \in \mathcal{G}_z} Q_j \geqslant 2c_z n^{-\tau_z}.$$

Assumption 6. There exist a positive constant $c'' > 0$ such that

$$\liminf_{p \to \infty} \inf_{m} \{\min_{j \in \mathcal{G}_z} Q_j - \max_{j' \in \bar{\mathcal{G}}_z} Q_{j'}\} \geqslant c''.$$

Then we present the sure screening property and ranking consistency property of PKC-SIS in the following theorem.

Theorem 2. *(1) Under Assumptions 1 and 4, for any* $\tau_z > 0$, $\varpi^* = \min\{\varpi, \varpi_z\} > 0$, $\gamma_z > 0$ *and* $1 - 2\gamma_z - 2\tau_z > 0$, *there exist positive constants* c_1^*, c_2^*, c_3^*, *and* c_4^* *such that*

$$P\Big(\max_{1 \leqslant j \leqslant p} \big|\widehat{Q}_j - Q_j\big| \geqslant cn^{-\tau_z}\Big) \leqslant c_1^* p \exp(-c_2^* n^{1-2\gamma_z-2\tau_z}) + c_3^* p \exp(-c_4^* n^{\gamma_z \varpi^*}),$$

when n *is sufficiently large.*

(2) Under Assumptions 1, 2, 4, and 5,

$$P(\mathcal{G}_z \subseteq \widehat{\mathcal{G}}_z) \geqslant 1 - s_z c_1^* \exp(-c_2^* n^{1-2\gamma_z-2\tau_z}) - s_z c_3^* \exp(-c_4^* n^{\gamma_z \varpi^*}),$$

where $\widehat{\mathcal{G}}_z = \left\{j \in \{1, \ldots, p\} : \widehat{Q}_j \geq c_z n^{-\tau_z}\right\}$ *and* $s_z = |\mathcal{G}_z|$.

(3) [Ranking Consistency Property] Under Assumptions 1, 3 , 4, and 6, for $\log(p) = o(n^{\nu_z})$ *with* $0 < \nu_z < \min\{1 - 2\gamma_z - 2\tau_z, \gamma_z \varpi^*\}$,

$$\liminf_{n \to \infty} \{\min_{j \in \mathcal{G}_z} \widehat{Q}_j - \max_{j' \in \bar{\mathcal{G}}_z} \widehat{Q}_{j'}\} \geqslant 0, a.s.$$

Remark 3. In PKC-SIS, if we chose $\gamma_z = (1-2\tau_z)/(\varpi_z+2)$, the result (2) in Theorem 2 is equal to $P(\mathcal{G}_z \subseteq \widehat{\mathcal{G}}_z) \leq c_5^* p \exp\left\{-c_6^* n^{\frac{\varpi^*}{\varpi^*+2}(1-2\tau_z)}\right)\right\}$ for some constants c_5^* and c_6^*. It shows that KC-SIS can handle the non-polynomial dimensionality of order $\log p = o\{n^{\frac{\varpi^*}{\varpi^*+2}(1-2\tau_z)}\}$.

3.3 Kernel choice

Since the proposed statistics are kernel-based, the choice of kernel functions may affect their performances. To elucidate the impact of kernel choice, we conduct simulations to compare the efficacy of Gaussian kernel and Laplacian kernel, both commonly utilized in scenarios involving continuous outcomes, within the framework of KC-SIS. Detailed simulation outcomes, available in the supplementary material, demonstrate that these two kernels yield comparable effects on the performance of the novel procedure. This suggests that opting for the Gaussian kernel is a prudent choice in situations where empirical knowledge is lacking. In addition to the Gaussian kernel, alternative kernels can be employed to tailor to diverse data characteristics and incorporate specific empirical knowledge. For example, when analyzing positive real numbers, kernel functions such as k_{I} and k_{II}, as discussed in Section 2.1 may be considered.

It is imperative to note that the application of characteristic kernels is not mandatory and should be approached with careful consideration. For instance, although the linear kernel may not be characteristic, it exhibits optimal performance under a linear model with normality. This arises from the fact that the marginal utility R_j in the linear regression model transforms into the Pearson correlation coefficient between X_j and Y which also equals zero if and only if X_j and Y are independent. Therefore it is possible that the sure screening properties of KC-SIS and PKC-SIS remain valid even when using uncharacteristic kernels under appropriate assumptions.

Since diverse kernels can be employed in the proposed KC-SIS and PKC-SIS procedures and the optimal kernel is always unknown beforehand, it is advisable to use multiple candidate kernels to enhance statistical power and analysis efficiency. Specifically, we can build a composite kernel based on a set of reproducing kernels $\{k_i\}_{i=1}^{M}$ via $\tilde{k} = \sum\limits_{i=1}^{M} \omega_i k_i$ with $\omega_i \geqslant 0$ and $\sum\limits_{i=1}^{M} \omega_i = 1$, and apply this new composite kernel to the statistics construction. Screening procedures based on the composite kernel is more robust against the underlying data structures and association patterns, thus can achieve satisfying performances in diverse situations. A rational strategy to determine the non-negative weight ω_i is to set $\omega_i = 1/\mathrm{tr}(K_i)$, where $\mathrm{tr}(\cdot)$ represents the trace of a matrix (Zhan et al., 2017).

---

4. Simulation Studies

In this section, we conduct extensive simulations to investigate the performances of KC-SIS and PKC-SIS by comparing with the existing methods DC-SIS (Li, Zhong and Zhu, 2012), MDC-SIS (Shao and Zhang, 2014) and the Kendall's τ- based mRCC (He, Zhou and Zou, 2021), in both situations without and with confounding factors.

To accommodate different relationships, diverse scenarios are considered, and 500 repetitions are conducted in each scenario to evaluate the performances. The following three criteria are employed in the comparison, which are commonly adopted in the literature on feature screening:

(1) $\mathcal{S}$: the minimum model size to include all active predictors. We report the 5%, 25%, 50%, 75%, and 95% quantiles of $\mathcal{S}$ out of 500 replications;

(2) $\mathcal{P}_s$: the proportion that an individual active predictor is selected for a given model size L in the 500 replications;

(3) $\mathcal{P}_a$: the proportion that all active predictors are selected for a given model size L in the 500 replications.

Among these criteria, $\mathcal{S}$ quantifies the capability of a screening procedure to detect all active predictors, and methods with smaller quantiles of $\mathcal{S}$

are less conservative, since they select less variables to detect all active predictors. The criteria $\mathcal{P}_s$ and $\mathcal{P}_a$ assess the screening procedure from the perspective of selecting an single predictor and selecting all true predictors, respectively.

4.1 Without the existence of confounding factors

In this subsection, independence testing without the existence of confounding factors are considered. Due to the incompatibility of mRCC with high-dimensional response models, we will exclude its implementation in these scenarios. For ease of presentation, we apply the Gaussian kernel to both X_j and Y in KC-SIS, i.e., $k_{\mathtt{G}}(x_{ij}, x_{lj}) = \exp(-||x_{ij} - x_{lj}||_1^2/\sigma_{x_j}^2)$ and $k_{\mathtt{G}}(\boldsymbol{y}_i, \boldsymbol{y}_l) = \exp(-||\boldsymbol{y}_i - \boldsymbol{y}_l||_m^2/\sigma_y^2)$, $i, l = 1, \ldots, n$, where the shape parameters $\sigma_{x_j}^2$ and σ_y^2 are typically selected as the median of $\{||x_{ij} - x_{lj}||_1^2 : 1 \leqslant i, l \leqslant n\}$ and $\{||\boldsymbol{y}_i - \boldsymbol{y}_l||_m^2 : 1 \leqslant i, l \leqslant n\}$, respectively, $j = 1, \ldots, p$.

Two types of scenarios are considered, where the sample size is set to be $n = 200$, and the dimension $p = 2000$. In the following scenarios, the candidate multivariate predictor $X = (X_1, \ldots, X_p)^\top$ is generated from the normal distribution $N(0_p, \Sigma_p)$ where covariance matrix structure has an autoregressive pattern, i.e., $\Sigma_p = (\sigma_{il})_{p\times p} = (\xi^{|i-l|})_{p\times p}$ with $\xi = 0.8$, and the error term ϵ_j is independently generated from a normal distribution

$N(0,1)$ for $j=1,\ldots,m$. Three choices of L are regularly considered in the screening procedure (Li, Zhong and Zhu, 2012; Shao and Zhang, 2014), i.e., $L_1=[n/\log n]$, $L_2=2[n/\log n]$ and $L_3=3[n/\log n]$, where $[a]$ denotes the integer part of a. The specific settings of each scenario are as follows.

- *Scenario 1.* Consider the following two heteroscedastic high-dimensional response models, where the dimension of response is set to be $m=300$.

  (1.a) Each response is generated as

  $$Y_j=\beta_{1j}^*X_1+\beta_{2j}^*X_2X_3+\beta_{3j}^*X_4X_5+\exp(X_{20}+X_{21}+X_{22})\epsilon_j,$$

  for $j=1,\ldots,300$, where the parameters β_{lj}^* are independently generated from a uniform distribution $U(0,1)$ for $l=1,2,3$ and $j=1,\ldots,300$.

  (1.b) Each response is generated as

  $$\begin{aligned}Y_j= &\ \alpha_1^*\beta_{1j}^*X_1+\alpha_2^*\beta_{2j}^*X_2+\alpha_3^*\beta_{3j}^*X_3+\alpha_4^*\beta_{4j}^*X_4+\alpha_5^*\beta_{5j}^*X_5\\ &+\exp(X_{20}+X_{21}+X_{22})\epsilon_j,\end{aligned}$$

  for $j=1,\ldots,300$, where the parameter $(\alpha_1^*,\alpha_2^*,\alpha_3^*,\alpha_4^*,\alpha_5^*)=(1,0.8,0.6,0.4,0.2)$ and β_{lj}^*s are independently generated from a uniform distribution $U(0,1)$ for $l=1,\ldots,5$ and $j=1,\ldots,300$.

- *Scenario 2.* Consider the following multivariate response models generated from Li, Zhong and Zhu (2012). In this setting, response $Y = (Y_1, \dots, Y_m)^\top$ is generated from the normal distribution $N(0_m, \Sigma_{Y|X})$, where $\Sigma_{Y|X} = (\sigma_{X,il})_{m\times m}$ with $\sigma_{X,il} = \sigma_X^{|i-l|}$ and $\sigma_X = \sin(0.8X_1 + 0.6X_2 + 0.4X_3 + 0.2X_4)$, a more complex autoregressive associaton pattern. Three different dimensions $m = 20, 50, 100$ are considered, and the corresponding scenarios are denoted as Scenarios (2.a), (2.b) and (2.c), respectively.

We report the values of $\mathcal{S}$ under Scenario 1 in Table 1. It can be seen that all the quantiles for $\mathcal{S}$ of the proposed KC-SIS are smaller than their counterparts of DC-SIS and MDC-SIS. This indicates that the novel KC-SIS is less conservative than DC-SIS and MDC-SIS, thus is more efficient. Additionally, we provide visual representations of $\mathcal{P}_s$ and $\mathcal{P}_a$ in Figure 1. It can be seen that KC-SIS achieves a conspicuous success in identifying all active variables in heteroscedastic Models (1.a) and (1.b), whereas DC-SIS and MDC-SIS merely capture association signals of variables X_{20}, X_{21} and X_{22}. For example, in Model (1.b), all eight active predictors are selected by KC-SIS with the model size L_2, while the selected rates of $(X_1, X_2, X_3, X_4, X_5)$ are all lower than 69% by DC-SIS, and all the chosen proportions of $(X_{20}, X_{21}$ and $X_{22})$ are lower than 5% by MDC-SIS.

The simulation results of $\mathcal{S}$, $\mathcal{P}_s$ and $\mathcal{P}_a$ under Scenario 2 are illustrated in Table 2 and Figure 2. Among all the methods considered, KC-SIS demonstrates the most favorable performance, characterized by the smallest quantiles, as depicted in Table 2. This observation underscores its superior efficiency, attributable to its minimal conservativeness. Table 2 indicates that KC-SIS performs significantly better than the rest three competitors in variable selection regardless of the dimension of Y. In Figure 2, it indicates KC-SIS demonstrates satisfactory performance in identifying all active variables. However, both DC-SIS and MDC-SIS exhibit notable deficiencies, particularly in the selection of X_4. It appears that mRCC is substantially inferior to the other three methods across all these three models. For instance, in Model (2.a), 71% of X_4 is selected by KC-SIS for the model size L_1, as opposed to 51% by DC-SIS, 53% by MDC-SIS and 6% by mRCC.

4.2 With the existence of confounding factors

In scenarios that accommodate additional confounding factors, the finite sample performance of PKC-SIS is investigated by comparing with those of KC-SIS, DC-SIS and mRCC. For simplicity, the Gaussian kernels are applied to X_j, Y and Z, respectively, in PKC-SIS. Being the same as scenarios without confounding factors, in this subsection, the sample size is set as $n = 200$ and the dimension is $p = 2000$. In the following two scenarios, data $X = (X_1, \ldots, X_p)^\top$ is generated from the normal distribution $N(0_p, \Sigma_p)$ where covariance matrix structure has an autoregressive pattern, i.e., $\Sigma_p = (\sigma_{il})_{p\times p} = (\xi^{|i-l|})_{p\times p}$ with $\xi = 0.8$, and the error term ϵ_j is independently generated from a normal distribution $N(0, 1)$ for $j = 1, \ldots, m$. The same three choices of L as before are considered in the screening procedure. The specific settings for the considered two scenarios are as follows.

- *Scenario 3.* Consider the following two heteroscedastic high-dimensional response model. The jth predictor X_j^* is generated by letting $Z \sim N(1, 1)$ and $X_j^* = X_j + Z$, $j = 1, \ldots, p$, and the jth response is

  $$Y_j = 3X_1^* + 3\sin(X_{11}^*) + 3|Z|^3 + \exp(X_{11}^* + X_{12}^*)\epsilon_j, \quad j = 1, \ldots, 300. \tag{4.1}$$

  $$Y_j = 3X_1^* + 3X_1^*X_{11}^* + 3|Z|^3 + \exp(X_{11}^* + X_{12}^*)\epsilon_j, \quad j = 1, \ldots, 300. \tag{4.2}$$

  Model (3.a) has in a linear term X_1^* and a nonlinear term $\sin(X_{11}^*)$, and Model (3.b) embodies an interaction term $X_1^*X_{11}^*$.

- *Scenario 4.* Consider multivariate responses. First generate unobserved multivariate variables $Y^* = (Y_1^*, \ldots, Y_m^*)^\top$ with $m = 5$ from a normal distribution $N(0_m, \Sigma_{Y|X})$, where the covariance matrix is

$$\Sigma_{Y^*|X} = (\sigma_{X,il})_{m\times m} = \left(\sigma_X^{|i-l|}\right)_{m\times m} = \left\{\sin(0.8X_1 + 0.6X_2)^{|i-l|}\right\}_{m\times m}.$$

Then observed responses via the following two models:

$$Y_j = Y_j^* + 0.3Z, \quad j = 1, \ldots, m. \tag{4.3}$$

$$Y_j = Y_j^* + 0.1|Z|^3, \quad j = 1, \ldots, m. \tag{4.4}$$

where $Z \sim N(1, 1)$.

Table 3 and Figure 3 summarize the simulation results of $\mathcal{S}$, $\mathcal{P}_s$ and $\mathcal{P}_a$ under Scenario 3. It is evident from the results that PKC-SIS displays the lowest level of conservativeness, followed by KC-SIS, DC-SIS and MDC-SIS. This observation suggests that while KC-SIS outperforms DC-SIS and MDC-SIS in terms of efficiency, the incorporation of confounding factors, when present, enhances efficiency further. For example, in Model (3.a), the 25% quantile of the minimum model size $\mathcal{S}$ is 14.8 for PKC-SIS, in contrast to 19.0 for KC-SIS, 92.0 for DC-SIS and 1973.0 for MDC-SIS. Due to its low sensitivity in capturing active predictors in additive models with heteroscedastic errors, DC-SIS has less chance to select the predictor X_1 than PKC-SIS and KC-SIS in Model (3.a), whereas MDC-SIS performs the worst in capturing all the predictors across all methods.

The results corresponding to scenario 4 are presented in Table 4 and Figure 3. It can be seen that PKC-SIS performs significantly better than DC-SIS and mRCC, and slightly outperforms KC-SIS in both two models. For example, in Model (4.b), when $\xi = 0.8$, the 95% quantile of the minimum model size $\mathcal{S}$ is 13.3 for PKC-SIS, compared to that of 18.0 for KC-SIS, 743.0 for DC-SIS, 1071.0 for MDC-SIS, and 1445.5 for mRCC. Note that KC-SIS still achieves satisfactory results when the variable Z is independent of X, thereby confirming the theoretical equivalence between PKC-SIS and KC-SIS when the confounding factors are independent of either predictors or response variables.

5. Application to Stock Mice Genome-wide Data

Mice are a key model organism for understanding the gene function in mammals. Valdar et al. (2006) recorded 12,226 SNPs and 101 phenotypes of 667 mice to conduct researches on their association relationships. In this work, we apply the proposed PKC procedure to this dataset to identify the phenotype-predisposing SNPs, which is crucial for understanding the shared etiology among complex phenotypes.

We first perform quality control on both SNPs and phenotyps. Specifically, SNPs with missing values and minor allele frequencies outside the range of 0.05 to 0.95 are removed, resulting in 5,796 SNPs remain to be further analyzed. And phenotypes with missing rate larger than 0.15 are deleted, and the remaining missing data is imputed via mean imputation, leading to 97 phenotypes for subsequential analysis.

In genetic association analysis, population stratification is a significant confounding factor that can lead to numerous false positive findings. To address this issue, we adopt the R package *AssocTests* (Wang, Zhang and Li, 2020) to select the top 10 significant eigenvalues as the population stratification confounding factors. This approach is in line with many previous studies such as those in Price et al. (2006) and Li and Yu (2008).

After quality control on data and adjustment on confounding factors, we apply PKC-SIS, DC-SIS and mRCC to screen out irrelevant predictors in this mice dataset, where the model size is $L = [n/\log n]$ with n being the sample size. Besides, we choose the top four SNPs from each method to be representatives of the significant associations found in this applicaiton. PKC-SIS select rs13482952, rs3705058, CEL-17_31069801, and rs3023110, among which SNPs rs3657760 and rs13480933 have been announced to be associated with the phenotypes in Valdar et al. (2006). For DC-SIS, the top four ranked SNPs are rs3685424, rs3659789, rs6168647, and rs3717220. And mRCC identifies rs6222023, rs13478330, UT_5_74.043446, and rs6163111 as

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0290

the top four ranked SNPs. However, none of these eight SNPs are reported in Valdar et al. (2006). This suggests the advantages of the proposed PKC-SIS over other existing methods.

6. Concluding Remarks

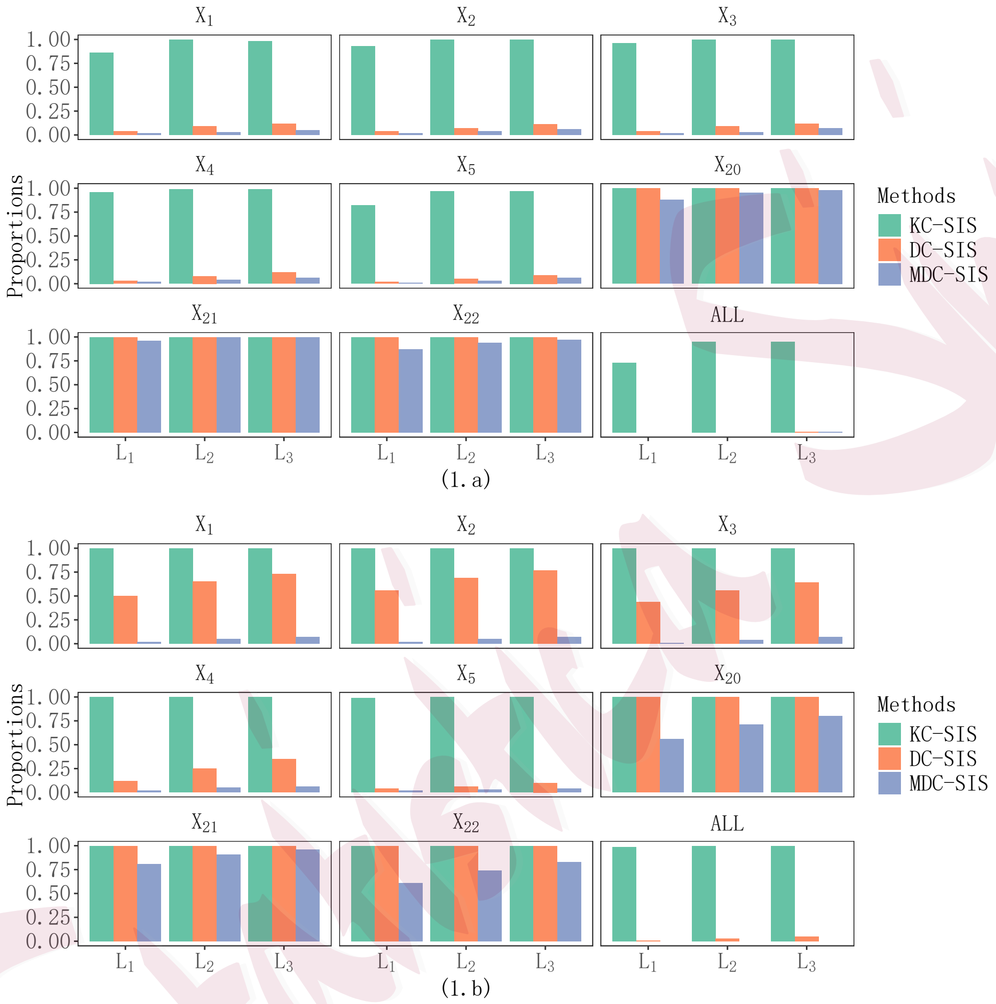
We consider feature screening procedures for high-dimensional response data, which is an important issue but receive less attention in literature. To fill this gap, we propose two feature screening procedures, the kernel correlation-based KC-SIS and partial kernel correlation-based PKC-SIS, with KC-SIS suitable for situations without confounding factors and PKC-SIS suitable for those with the existence of confounding factors. Both screening procedures are non-parametric, with no need on model restrictions, which makes them robust against outliers and model misspecification. And they possess sure screening property and ranking consistency property that guanrantee their utility in practice. Additionally, both KC-SIS and PKC-SIS are applicable to a variety of data types, including continuous and discrete data, Euclidean and non-Euclidean data, and light-tailed and hevay-tailed data.

Supplementary Materials

The proofs of Theorems 1 and 2, as well as additional numerical simulations on one-dimensional response variable models, two distinct kernels, and other popular machine learning approaches, can be found in the Supplementary Material.

Acknowledgements

We would like to express our sincere gratitude to Qunqiang Feng for his invaluable contributions to this article. Jinjuan Wang has been supported by National Natural Science Foundation of China (NSFC) (Grant No. 12101047) and Beijing Institute of Technology Research Fund Program for Young Scholars. Qizhai Li has been supported by National Natural Science Foundation of China (NSFC) (Grant No. 12325110) and CAS Project for Young Scientists in Basic Research (Grant No. YSBR-034).



(1. a)

(1. b)

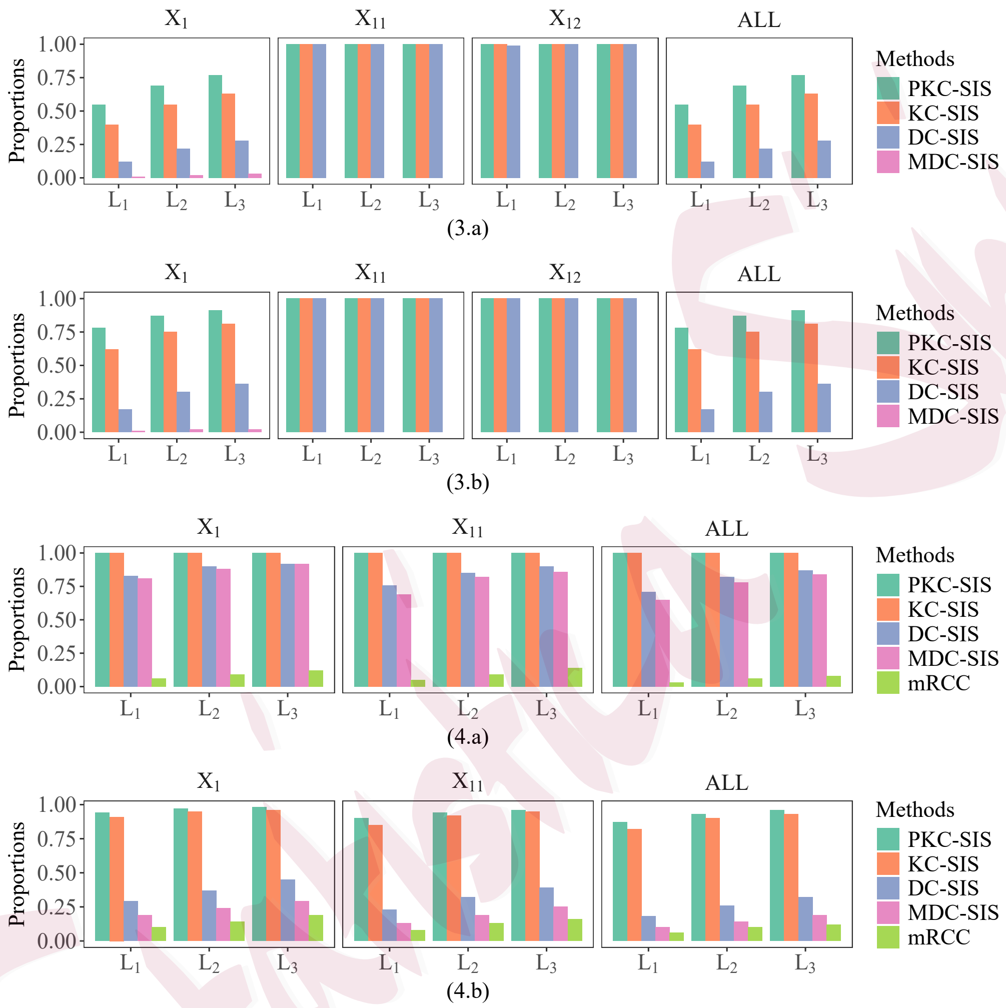
Figure 1: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ in Scenario 1 with $L_1 = [n/\log n]$, $L_2 = 2[n/\log n]$, $L_3 = 3[n/\log n]$, $n = 200$, and $p = 2000$.

(2. a)

(2. b)

(2. c)

Figure 2: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ in Scenario 2 with $L_1 = [n/\log n]$, $L_2 = 2[n/\log n]$, $L_3 = 3[n/\log n]$, $n = 200$, and $p = 2000$.



(3.a)

(3.b)

(4.a)

(4.b)

Figure 3: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ in Scenarios 3 and 4 with $L_1 = [n/\log n]$, $L_2 = 2[n/\log n]$, $L_3 = 3[n/\log n]$, $n = 200$, and $p = 2000$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0290

Table 1: The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size $\mathcal{S}$ out of 500 replications in Scenario 1, $n = 200$ and $p = 2000$.

Model	Method	5%	25%	50%	75%	95%
(1.a)	KC-SIS	12.0	14.0	18.0	30.0	92.1
	DC-SIS	23.0	24.0	25.0	938.0	1864.4
	MDC-SIS	809.7	1406.8	1683.0	1866.5	1972.0
(1.b)	KC-SIS	8.0	9.0	11.0	12.0	15.0
	DC-SIS	447.1	1255.5	1587.0	1803.0	1949.1
	MDC-SIS	1269.7	1604.0	1793.0	1903.0	1985.1

Table 2: The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size $\mathcal{S}$ out of 500 replications in Scenario 2 with $\xi = 0.8$, $n = 200$ and $p = 2000$.

Model	Method	5%	25%	50%	75%	95%
(2.a)	KC-SIS	4.0	4.0	10.0	54.0	418.3
	DC-SIS	4.0	9.0	35.0	162.3	702.5
	MDC-SIS	4.0	8.0	33.0	144.5	626.1
	mRCC	86.7	454.3	956.5	1499.8	1871.7
(2.b)	KC-SIS	4.0	5.0	18.5	96.3	500.4
	DC-SIS	4.0	13.8	69.5	251.3	843.4
	MDC-SIS	4.0	11.8	59.5	208.5	857.8
	mRCC	204.8	666.8	1138.5	1552.3	1902.4
(2.c)	KC-SIS(G)	4.0	8.0	36.0	155.0	796.5
	DC-SIS	5.0	26.0	110.5	432.0	1149.8
	MDC-SIS	5.0	19.8	88.0	369.8	1034.2
	mRCC	362.0	883.8	1318.5	1657.0	1933.1

Table 3: The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size $\mathcal{S}$ out of 500 replications in Scenario 3 with $\xi = 0.8$, $n = 200$ and $p = 2000$.

Model	Method	5%	25%	50%	75%	95%
(3.a)	PKC-SIS	9.0	14.8	30.5	103.5	516.2
	KC-SIS	9.0	19.0	60.0	225.0	810.7
	DC-SIS	19.0	92.0	302.5	715.5	1405.0
	MDC-SIS	1880.9	1973.0	1993.0	1999.0	2000.0
(3.b)	PKC-SIS	7.0	10.0	15.0	33.3	165.1
	KC-SIS	7.0	11.0	23.0	76.0	421.1
	DC-SIS	15.0	60.0	213.5	578.8	1320.3
	MDC-SIS	1887.8	1974.0	1993.0	1999.0	2000.0

Table 4: The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size $\mathcal{S}$ out of 500 replications in Scenario 4 with $n = 200$ and $p = 2000$.

Model	Method	5%	25%	50%	75%	95%
(4.a)	PKC-SIS	2.0	2.0	2.0	2.0	3.1
	KC-SIS	2.0	2.0	2.0	2.0	4.0
	DC-SIS	2.0	3.0	13.0	46.3	260.4
	MDC-SIS	2.0	4.0	17.0	61.8	294.4
	mRCC	65.9	372.5	881.5	1414.3	1854.3
(4.b)	PKC-SIS	2.0	2.0	3.0	13.3	106.0
	KC-SIS	2.0	2.0	4.0	18.0	152.1
	DC-SIS	6.0	72.0	280.5	743.0	1379.4
	MDC-SIS	14.0	159.8	526.0	1071.0	1633.6
	mRCC	76.6	417.5	929.5	1445.5	1898.1

References

Cui, H., Li, R. and Zhong, W. (2015). Model-free feature screening for ultrahigh-dimensional discriminant analysis. *Journal of the American Statistical Association* **110**, 630-641.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh-dimensional feature space. *Journal of the Royal Statistical Society: Series B* **70**, 849-911.

Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality, *The Annals of Statistics* **38**, 3567–3604.

Fan, J., Feng, Y. and Song, R. (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association* **106**, 544–557.

Fan, J., Ma, Y. and Dai, W. (2014). Nonparametric independence screening in sparse ultrahigh-dimensional varying coefficient models. *Journal of the American Statistical Association* **109**, 1270–1284.

Fukumizu, K., Gretton, A., Schölkopf, B., and Sriperumbudur, B. K.

(2008). Characteristic kernels on groups and semigroups. *Advances in Neural Information Processing Systems* **21**, 473–480.

Gratten, J. and Visscher, P. M. (2016). Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Medicine* **8**, 1-3.

He, D., Zhou, Y. and Zou, H. (2021). On sure screening with multiple responses. *Statistica Sinica* **31**, 1749-1777.

He, X., Wang, L. and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics* **41**, 342–369.

Huang, J., Horowitz, J. L. and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics* **36**, 587-613.

Li, Q. and Yu, K.(2008). Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genetic Epidemiology* **32**, 215-226.

Li, R., Zhong, W. and Zhu, L. (2012) Feature Screening via Distance Correlation Learning. *Journal of the American Statistical Association* **107**, 1129-1139.

Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability* **41**, 3284-3305.

Mai, Q. and Zou, H. (2015). The fused Kolmogorov filter: A nonparametric model-free screening method. *The Annals of Statistics* **43**, 1471-1497.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* 2263-2291.

Shao, X. and Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association* **109**, 1302-1318.

Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *The Annals of Statistics* **42**, 2382-2412.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769-2794.

Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M.S., Rawlins, J.N.P., Mott, R., and Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics* **38**, 879-887.

Wang, L., Zhang, W. and Li, Q. (2020). Assoctests : an R package for genetic association studies. *Journal of Statistical Software* **94**, 1-26.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D.(2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348.

Xie, J., Lin, Y., Yan, X. and Tang, N. (2020). Category-adaptive variable screening for ultra-high-dimensional heterogeneous categorical data. *Journal of the American Statistical Association* **115**, 747-760.

Zhan, X., Plantinga, A., Zhao, N., and Wu, M. C. (2017). A fast small-sample kernel independence test for microbiome community-level association analysis. *Biometrics* **73**, 1453-1463.

Zhang, W., Yang, L., Tang, L. L., Liu, A., Mills, J. L., Sun, Y., and Li, Q. (2017). GATE: an efficient procedure in study of pleiotropic genetic associations. *BMC Genomics* **18**, 1-15.