Journal of Multivariate Analysis 149 (2016) 81–91

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On permutation tests for predictor contribution in sufficient dimension reduction

Yuexiao Dong [a], Chaozheng Yang [a], Zhou Yu [b,*]

[a] Department of Statistics, Temple University, Philadelphia, PA, 19122, United States
[b] School of Statistics, East China Normal University, Shanghai, 200241, China

ARTICLE INFO

Article history:
Received 2 April 2015
Available online 22 March 2016

AMS subject classifications:
62F03
62F05

Keywords:
Nonparametric test
Directional regression
Sliced average variance estimation
Sliced inverse regression

ABSTRACT

To test predictor contribution in a model-free fashion, marginal coordinate tests based on sliced inverse regression (SIR) and sliced average variance estimation (SAVE) have been studied in Cook (2004), and Shao et al. (2007) respectively. Estimating the null distributions of the test statistics is a critical step for such tests. We propose a novel permutation test approach to facilitate the marginal coordinate tests, which applies to existing tests based on SIR and SAVE, and can be readily extended to a new marginal coordinate test based on directional regression (Li and Wang, 2007).

1. Introduction

Sufficient dimension reduction [1] is an effective tool in high-dimensional data analysis. Consider a univariate response Y and a p-dimensional predictor $X = (X_1, \ldots, X_p)^\top$. The goal for sufficient dimension reduction is to find linear combinations of X, such that Y is independent of X given these linear combinations. Pioneered by sliced inverse regression (SIR; Li [6]), various methods have been developed in the sufficient dimension reduction literature. Instead of using the intraslice means as in SIR, Cook & Weisberg [3] utilized the intraslice variances and suggested sliced inverse variance estimation (SAVE). Directional regression (DR) is proposed in [7] to combine the information from the intraslice means and the intraslice variances. SIR, SAVE and DR all belong to the family of moment-based sufficient dimension reduction estimators.

For $X \in \mathbb{R}^p$ and subscript $i \in \{1, \ldots, p\}$, denote $X_{-i} \in \mathbb{R}^{p-1}$ as $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)^\top$. One can test the contribution of X_i through the following hypotheses,

$$H_0 : Y \perp\!\!\!\perp X | X_{-i} \text{ versus } H_a : Y \text{ is not independent of } X \text{ given } X_{-i}, \quad (1)$$

where $\perp\!\!\!\perp$ denotes statistical independence. In (1), the null hypothesis implies that Y is independent of X given X_{-i}. If we fail to reject H_0, we conclude that predictor X_i does not have significant contribution to the regression between Y and X. Hypotheses (1) are first proposed as the marginal coordinate hypotheses in the seminal work of Cook [2]. Cook [2] also proposed the corresponding marginal coordinate test based on SIR. A similar test based on SAVE has been discussed in [9]. More recently, the marginal coordinate test based on DR is developed in [12]. For marginal coordinate tests based on SIR, SAVE and DR, a technical difficulty is to develop the distributions of their corresponding test statistics under H_0. The asymptotic distribution for each method-specific test statistic turns out to be a sum of weighted $\chi^2(1)$ distributions, where the exact weights for each method can be found in [2,9,12] respectively.

* Corresponding author.
E-mail address: zyu@stat.ecnu.edu.cn (Z. Yu).

http://dx.doi.org/10.1016/j.jmva.2016.02.019
0047-259X/

As an alternative to derive the null distribution of the aforementioned marginal coordinate tests, we propose a unified permutation test approach in this paper. Our proposal is easy to implement, as it only involves random permutations of the observed predictors while fixing the responses before recalculating the sample test statistics with the permuted samples. It applies to marginal coordinate tests based on SIR, SAVE and DR, and no longer requires calculation of the method-specific weights to determine the asymptotic null distribution. The rest of the paper is organized as follows. We briefly review the sufficient dimension reduction methods SIR, SAVE and DR in Section 2. The permutation approach with SIR is developed in Section 3, and the analogous approaches for SAVE and DR are proposed in Section 4. The developments in Sections 3 and 4 require the normality of predictor X. Extensive numerical studies are carried out in Section 5. In Section 5.1, the permutation tests are compared with the existing asymptotic tests. Permutation test with predictor transformation is considered for non-normal predictors in Section 5.2. For the real data analysis in Section 5.3, the proposed permutation test is applied in a backward elimination procedure to achieve model-free variable selection. We conclude the paper with some discussions in Section 6. For the ease of presentation, all the proofs are delegated to the Appendix.

2. Review of SIR, SAVE and DR

The central space between Y and X, denoted as $\mathcal{S}_{Y|X}$, is a critical concept in sufficient dimension reduction. Let $\beta \in \mathbb{R}^{p\times d}$ be the basis of the central space $\mathcal{S}_{Y|X}$. Then the column space of β satisfies $\mathrm{Span}(\beta) = \mathcal{S}_{Y|X}$ and β satisfies $Y \perp\!\!\!\perp X|\beta^\top X$. For any $\gamma \in \mathbb{R}^{p\times q}$ also satisfying $Y \perp\!\!\!\perp X|\gamma^\top X$, the central space is defined such that $\mathrm{Span}(\beta) \subseteq \mathrm{Span}(\gamma)$ with $d \le q$. The conditional independency $Y \perp\!\!\!\perp X|\beta^\top X$ implies that one can reduce the p-dimensional predictor X to the d-dimensional $\beta^\top X$ for the regression between Y and X. The requirement $\mathrm{Span}(\beta) \subseteq \mathrm{Span}(\gamma)$ guarantees $\beta^\top X$ achieves minimum dimensionality reduction.

Let $\mu = \mathrm{E}(X)$ and $\Sigma = \mathrm{Var}(X)$. Then $Z = \Sigma^{-1/2}(X-\mu)$ is the standardized predictor. The relationship between the Z-scale central space $\mathcal{S}_{Y|Z}$ and the X-scale central space $\mathcal{S}_{Y|X}$ is stated next.

Proposition 2.1. *Suppose $\mathcal{S}_{Y|Z}$ has basis η such that $\mathcal{S}_{Y|Z} = \mathrm{Span}(\eta)$. Then $\Sigma^{-1/2}\eta$ is a basis for $\mathcal{S}_{Y|X}$ and satisfies $\mathcal{S}_{Y|X} = \mathrm{Span}(\Sigma^{-1/2}\eta)$.*

The relationship above is known as the invariance property of the central space. Due to this invariance property, one can first find estimators of the Z-scale central space, and then transform it back to the X-scale.

SIR [6], SAVE [3], and DR [7] are among the most popular central space estimators. Denote $M_{\mathrm{SIR}} = \mathrm{Var}\{\mathrm{E}(Z|Y)\}$, $M_{\mathrm{SAVE}} = \mathrm{E}[\{I_p - \mathrm{Var}(Z|Y)\}^2]$, and $M_{\mathrm{DR}} = \mathrm{E}\{(2I_p - A(Y,\tilde{Y})^2)\}$, where $A(Y,\tilde{Y}) = \mathrm{E}\{(Z-\tilde{Z})(Z-\tilde{Z})^\top|Y,\tilde{Y}\}$ with $(\tilde{Y},\tilde{Z})$ being an independent copy of (Y,Z). Then we have

Proposition 2.2. *Suppose Z is normal. Then $\mathrm{Span}(M) \subseteq \mathcal{S}_{Y|Z}$, where M can be M_{SIR}, M_{SAVE} or M_{DR}.*

We remark that the normality assumption can be relaxed to weaker assumptions. See, for example, the discussions in [7]. Proposition 2.2 suggests that the eigenvectors corresponding to the nonzero eigenvalues of kernel matrices M_{SIR}, M_{SAVE} and M_{DR} can be used to recover the Z-scale central space.

For discrete response Y, assume without loss of generality that Y has support $\Pi = \{1,\dots,H\}$. Then for $h \in \Pi$, $\mathrm{E}(Z|Y=h)$ denotes the within group mean for the hth category of Y. The meaning of $\mathrm{E}(Z|Y)$ for continuous response Y is explained next. Let $\{J_1,\dots,J_H\}$ be a partition of the support of Y. Then $\mathrm{E}(Z|Y)$ is discretized as $\mathrm{E}(Z|Y\in J_h)$, $h=1,\dots,H$. Let f_h be the probability of $Y \in J_h$ and denote $\xi_h = \mathrm{E}(Z|Y\in J_h)$. A discretized version of M_{SIR} thus becomes $M_{\mathrm{SIR}} = \sum_{h=1}^H f_h\xi_h\xi_h^\top$. Similarly, the discretized kernel matrices for SAVE and DR are $M_{\mathrm{SAVE}} = \sum_{h=1}^H A_h^2$ and $M_{\mathrm{DR}} = \sum_{h=1}^H\sum_{k=1}^H B_{hk}^2$ respectively, where $A_h = f_h^{1/2}\{I_p - \mathrm{Var}(Z \mid Y\in J_h)\}$ and $B_{hk} = f_h^{1/2}f_k^{1/2}[2I_p - \mathrm{E}\{(Z-\tilde{Z})(Z-\tilde{Z})^\top|Y\in J_h,\tilde{Y}\in J_k\}]$. Without ambiguity, we denote M_{SIR}, M_{SAVE} and M_{DR} as the discretized kernel matrices for the rest of the paper.

3. Permutation test for predictor contribution with SIR

3.1. Test statistic construction

Recall that the null hypothesis $H_0 : Y \perp\!\!\!\perp X|X_{-i}$ in (1) implies that X_i has no additional contribution to Y given the other predictors. For $i=1,\dots,p$, define $e_i \in \mathbb{R}^p$, where the ith element of e_i is 1 and all the other elements are zero. Alternatively, we can define e_i as the ith column of the identity matrix I_p. The next observation is key to develop the test statistics for (1).

Proposition 3.1. *Suppose $\mathcal{S}_{Y|Z} = \mathrm{Span}(\eta)$ for $\eta \in \mathbb{R}^{p\times d}$. Then $e_i^\top\Sigma^{-1/2}\eta = 0$ if and only if $Y \perp\!\!\!\perp X|X_{-i}$.*

Note that due to Proposition 2.1, $\Sigma^{-1/2}\eta$ is the basis of the X-scale central space. Proposition 3.1 thus implies η, or the basis of the Z-scale central space, can be used to test the conditional independence $Y \perp\!\!\!\perp X|X_{-i}$ at the X-scale.