Computational Statistics and Data Analysis 53 (2008) 334–343

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Response dimension reduction for the conditional mean in multivariate regression

Jae Keun Yoo [a,*], R. Dennis Cook [b]

[a] Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY 40202, USA
[b] School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 22 August 2007
Received in revised form 21 July 2008
Accepted 21 July 2008
Available online 26 July 2008

ABSTRACT

Sufficient dimension reduction methodologies in regression have been developed in the past decade, focusing mostly on predictors. Here, we propose a methodology to reduce the dimension of the response vector in multivariate regression, without loss of information about the conditional mean. The asymptotic distributions of dimension test statistics are chi-squared distributions, and an estimate of the dimension reduction subspace is asymptotically efficient. Moreover, the proposed methodology enables us to test response effects for the conditional mean. Properties of the proposed method are studied via simulation.

1. Introduction

One goal of *Sufficient dimension reduction* (SDR) in regression is to replace the predictors $\mathbf{X} \in \mathbb{R}^p$ with a lower dimensional projection $\mathbf{P}_{\mathcal{S}}\mathbf{X}$ onto a subspace $\mathcal{S}$ of $\mathbb{R}^p$, without loss of information on selected aspects of the conditional distribution of $\mathbf{Y}|\mathbf{X}$, where $\mathbf{Y} = (Y_1, \ldots, Y_r)^{\mathrm{T}}$ is the response vector with $r \geq 2$ and $\mathbf{P}_{\mathcal{S}}$ stands for the orthogonal projection operator onto $\mathcal{S}$ with the usual inner product space. We assume throughout, that the data $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \ldots, n$, are a random sample of $(\mathbf{X}, \mathbf{Y})$ with finite fourth moments. We use $\mathcal{S}(\mathbf{B})$ to denote the subspace of $\mathbb{R}^p$ spanned by the columns of $\mathbf{B} \in M_{p\times r}(\mathbb{R})$.

Methodologies for SDR in multivariate regression have been recently developed by Li et al. (2003), Cook and Setodji (2003), Setodji and Cook (2004) and Yoo and Cook (2007). Li et al. (2003) provided a method to reduce the dimensions of both $\mathbf{X}$ and $\mathbf{Y}$, by applying sliced inverse regression (SIR; (Li, 1991)) twice in additive-error multivariate regression models. Focusing on estimation of the multivariate central mean subspace, Cook and Setodji (2003) developed a method for SDR in multivariate regression to reduce the dimension of $\mathbf{X}$ alone. They used a set of the OLS coefficient vectors $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\,\mathrm{cov}(\mathbf{X}, \mathbf{Y}^{\mathrm{T}})$, where $\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{cov}(\mathbf{X})$, to estimate the multivariate central mean subspace. Setodji and Cook (2004) proposed K-means inverse regression to reduce $\mathbf{X}$ without loss of information on the conditional distribution of $\mathbf{Y}|\mathbf{X}$. K-means inverse regression involves application of a K-means algorithm to cluster $\mathbf{Y}$. These clusters can then be used as slices, and SIR applied in usual way. Yoo and Cook (2007) developed an optimal version of Cook and Setodji (2003).

Analysis of repeated measures, longitudinal data, or curve or time series data is often difficult, due to high dimensionality of $\mathbf{Y}$, although the dimension of $\mathbf{X}$ is relatively low. The study of such data would be facilitated if we could find a low dimensional linear transform of $\mathbf{Y}$ that adequately describes the regression relationship. Unfortunately, most SDR methods in multivariate regression have focused on reducing the dimension of $\mathbf{X}$, not $\mathbf{Y}$. The only method to provide response dimension reduction was proposed by Li et al. (2003), but it has notable deficiencies since, as discussed in Section 6, it does not perform well with correlated responses, it requires user-selected tuning parameters that can affect performance, and its asymptotic

* Corresponding author. Tel.: +1 5028523987; fax: +1 5028523294.
E-mail address: peter.yoo@louisville.edu (J.K. Yoo).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.07.029

properties are unknown. In this article, we propose SDR methodology to address these deficiencies, with the specific goal of reducing the dimension of $\mathbf{Y}$ without loss of information on $E(\mathbf{Y}|\mathbf{X})$. For illustration, the method is applied to epilepsy data by Thall and Vail (1990).

The organization of this article is as follows. We define working targets and investigate their properties in Section 2. Section 3 is devoted to developing our methodology to reduce $\mathbf{Y}$, including related inference procedures. In Section 4 we define hypotheses for response effects and address the test statistics and their asymptotic distributions. Section 5 contains simulation studies and data analysis. In Section 6 we summarize our work and compare it with the method proposed by Li et al. (2003). Proofs for most results are given in the Appendix.

2. Dimension reduction of responses

2.1. Linear response reduction

Consider a multivariate regression $\mathbf{Y}|\mathbf{X}$, and define $\mathbf{L}$ to be a $r \times q$ matrix with smallest possible rank $q \leq r$ so that $E(\mathbf{Y}|\mathbf{X}) = \mathbf{A}E(\mathbf{L}^{\mathrm{T}}\mathbf{Y}|\mathbf{X})$,, where $\mathbf{A}$ is a $r \times q$ matrix. This says that $\mathbf{X}$ can be thought of as influencing $\mathbf{L}^{\mathrm{T}}\mathbf{Y}$ and all other conditional mean components are determined from $E(\mathbf{L}^{\mathrm{T}}\mathbf{Y}|\mathbf{X})$ via $\mathbf{A}$. Let Ω denote the support of $\mathbf{X}$. Assuming that $\{E(\mathbf{L}^{\mathrm{T}}\mathbf{Y}|\mathbf{X} = x)|x \in \Omega\}$ spans a q dimensional subspace of $\mathbb{R}^p$, it can be shown that $\mathbf{A}$ is a generalized inverse of $\mathbf{L}^{\mathrm{T}}$: $\mathbf{L}^{\mathrm{T}}\mathbf{A}\mathbf{L}^{\mathrm{T}} = \mathbf{L}^{\mathrm{T}}$. Without loss of generality, we take $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{L}(\mathbf{L}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{L})^{-1}$ with $\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathrm{cov}(\mathbf{Y}) > 0$. Then $\mathbf{L}\mathbf{A}^{\mathrm{T}}$ forms the orthogonal projection operator $\mathbf{P}_{\mathbf{L}(\boldsymbol{\Sigma}_{\mathbf{Y}})}$ for $\mathcal{S}(\mathbf{L})$ relative to the inner product $\langle v_1, v_2\rangle_{\boldsymbol{\Sigma}_{\mathbf{Y}}} = v_1^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}v_2$, so we have

$$E(\mathbf{Y}|\mathbf{X}) = E[\{\mathbf{P}^{\mathrm{T}}_{\mathbf{L}(\boldsymbol{\Sigma}_{\mathbf{Y}})}\mathbf{Y}|\mathbf{X}\}]. \quad (1)$$

This says that $E(\mathbf{Y}|\mathbf{X})$ varies in the subspace spanned by $\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{L}$. Statement (1) implies that $E(\mathbf{Y}|\mathbf{X} = x) \in \mathcal{S}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{L})$ for all $x \in \Omega$.

If $q < r$, we accomplish the dimension reduction for $\mathbf{Y}$ without loss of information about $E(\mathbf{Y}|\mathbf{X})$. This dimension reduction will be called *linear response reduction*. A matrix $\mathbf{L}$ satisfying (1) always exists for any multivariate regressions, because we can set $\mathbf{L} = \mathbf{I}_r$. The following example is provided as the illustration:

Example 1. Suppose $\{\mathbf{X} = (X_1, X_2, X_3)^{\mathrm{T}}, \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^{\mathrm{T}}\}^{\mathrm{T}} \sim N(0, \mathbf{I}_6)$. Define $\mathbf{Y} = (Y_1, Y_2, Y_3)^{\mathrm{T}}$ as follows: $Y_1 = X_1 + \varepsilon_1$, $Y_2 = X_1 + \varepsilon_2$ and $Y_3 = \varepsilon_3$. Then $\mathbf{L} = (1, 1, 0)^{\mathrm{T}}$, and $\mathbf{P}^{\mathrm{T}}_{\mathbf{L}(\boldsymbol{\Sigma}_{\mathbf{Y}})} = \{(0.5, 0.5, 0)^{\mathrm{T}}, (0.5, 0.5, 0)^{\mathrm{T}}, (0, 0, 0)^{\mathrm{T}}\}$.

The following lemma shows that linear response reductions are equivariant under non-singular transformations of $\mathbf{Y}$.

Lemma 1. *Suppose that* $\mathbf{L}$ *satisfies* (1) *for* $\mathbf{Y}|\mathbf{X}$*. Let* $\mathbf{G}$ *be an* $r \times r$ *non-singular matrix. Then,* $\mathbf{G}^{-\mathrm{T}}\mathbf{L}$ *satisfies* (1) *for* $\mathbf{G}\mathbf{Y}|\mathbf{X}$*.*

Let $\mathbf{W}$ be any matrix that satisfies (1) and is orthogonal to $\mathbf{L}$ in the inner product $\langle\cdot, \cdot\rangle_{\boldsymbol{\Sigma}_{\mathbf{Y}}}$. For $\mathbf{L}^* = (\mathbf{L}, \mathbf{W})$, (1) holds, because $\mathbf{P}_{\mathbf{L}^*(\boldsymbol{\Sigma}_{\mathbf{Y}})} = \mathbf{P}_{\mathbf{L}(\boldsymbol{\Sigma}_{\mathbf{Y}})} + \mathbf{P}_{\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{Y}})}$ and $\mathbf{P}^{\mathrm{T}}_{\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{Y}})}E(\mathbf{Y}|\mathbf{X} = x) = 0$ for all $x \in \Omega$ by the fact that $E(\mathbf{Y}|\mathbf{X}) \in \mathcal{S}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{L})$. This implies the existence of multiple $\mathbf{L}$s satisfying (1). The $\mathbf{L}$ with the smallest rank can be defined uniquely as suggested by the following proposition.

Proposition 1. *Let* $\mathbf{L}$ *and* $\mathbf{L}^*$ *both satisfy* (1)*. Let the columns of the matrix* $\boldsymbol{\phi}$ *be a basis for* $\mathcal{S}(\mathbf{L}) \cap \mathcal{S}(\mathbf{L}^*)$*. Then* $\boldsymbol{\phi}$ *also satisfies* (1)*.*

2.2. Conditional response reduction

Reconsider Example 1 with Y_2 defined as $Y_2 = Y_1^2 + \varepsilon_2$. Then the matrix $\mathbf{L} = \{(1, 0, 0)^{\mathrm{T}}, (0, 1, 0^{\mathrm{T}})\}$ satisfies (1) and has smallest rank. However, all the information on $E(\mathbf{Y}|\mathbf{X})$ with respect to the responses comes from Y_1 alone. To cover this situation, we suppose that the following conditional statements holds for some $\mathbf{M} \in M_{r\times q}(\mathbb{R})$:

$$\begin{aligned} E(\mathbf{Y}|\mathbf{X}) &= E\{E(\mathbf{Y}|\mathbf{X}, \mathbf{M}^{\mathrm{T}}\mathbf{Y})|\mathbf{X}\} \\ &= E\{E(\mathbf{Y}|\mathbf{M}^{\mathrm{T}}\mathbf{Y})|\mathbf{X}\}. \end{aligned} \quad (2)$$

The equivalence (2) always holds, because the identity matrix is one of the possible choices for $\mathbf{M}$. The conditional mean $E(\mathbf{Y}|\mathbf{M}^{\mathrm{T}}\mathbf{Y})$ is a function, say g, of $\mathbf{M}^{\mathrm{T}}\mathbf{Y}$, so we can rewrite the right hand side of (2) as $E\{g(\mathbf{M}^{\mathrm{T}}\mathbf{Y})|\mathbf{X}\}$. Consequently, $E(\mathbf{Y}|\mathbf{X})$ can be fully characterized by the regression $g(\mathbf{M}^{\mathrm{T}}\mathbf{Y})|\mathbf{X}$, and we achieve dimension reduction of $\mathbf{Y}$ if $q < r$. We will call this *conditional response reduction*. In Example 1, (2) holds for $\mathbf{M} = (1, 1, 0)^{\mathrm{T}}$, while, in the revised version of Example 1, (2) holds for $\mathbf{M} = (1, 0, 0)^{\mathrm{T}}$.

For conditional response reduction, the relationship between $\mathbf{G}\mathbf{Y}|\mathbf{X}$ and $\mathbf{Y}|\mathbf{X}$ is the same as linear response reduction:

Lemma 2. *Suppose that* $\mathbf{M}$ *satisfies* (2) *for* $\mathbf{Y}|\mathbf{X}$*. Let* $\mathbf{G}$ *be a* $r \times r$ *non-singular matrix. Then* $\mathbf{G}^{-\mathrm{T}}\mathbf{M}$ *accomplishes conditional response reduction for* $\mathbf{G}\mathbf{Y}|\mathbf{X}$*.*