Computational Statistics and Data Analysis 55 (2011) 1656–1664

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

# Estimation of inverse mean: An orthogonal series approach

Qin Wang [a,*], Xiangrong Yin [b]

[a] Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284, United States

[b] Department of Statistics, 204 Statistics Building, The University of Georgia, Athens, GA 30602, United States

ARTICLE INFO

*Article history:*
Received 13 November 2009
Received in revised form 3 September 2010
Accepted 23 October 2010
Available online 30 October 2010

*Keywords:*
Sufficient dimension reduction
Central subspace
Sliced inverse regression
Orthogonal series

ABSTRACT

In this article, we propose the use of orthogonal series to estimate the inverse mean space. Compared to the original slicing scheme, it significantly improves the estimation accuracy without losing computation efficiency, especially for the heteroscedastic models. Compared to the local smoothing approach, it is more computationally efficient. The new approach also has the advantage of robustness in selecting the tuning parameter. Permutation test is used to determine the structural dimension. Moreover, a variable selection procedure is incorporated into this new approach, which is particularly useful when the model is sparse. The efficacy of the proposed method is demonstrated through simulations and a real data analysis.



## 1. Introduction

Sufficient dimension reduction (Li, 1991; Cook, 1998) has recently received much attention as an efficient tool to tackle the challenging problem of high dimensional data analysis. In full generality, the goal of regression is to elicit information on the conditional distribution of a univariate response $Y$ given a $p$-dimensional predictor vector $\mathbf{X}$. Sufficient dimension reduction is to find a $k$-dimensional projection subspace $\mathcal{S} = \text{Span}\{\mathbf{B} = (\beta_1, \beta_2, \ldots, \beta_k)\}$ with $k \le p$ such that

$$Y \perp\!\!\!\perp \mathbf{X} | P_{\mathcal{S}}\mathbf{X}, \tag{1}$$

where $\beta$'s are unknown $p \times 1$ vectors, $\perp\!\!\!\perp$ indicates independence and $P$ stands for a projection operator in the standard inner product. The subspace $\mathcal{S}$ is then called a *dimension reduction subspace* for $Y|\mathbf{X}$. When the intersection of all subspaces satisfying (1) also satisfies (1), it is called the *central subspace* (*CS*) and is denoted by $\mathcal{S}_{Y|\mathbf{X}}$. Its dimension $d_{Y|\mathbf{X}} = \dim(\mathcal{S}_{Y|\mathbf{X}})$ is defined as the *structural dimension* of the regression. Under some mild conditions, the *CS* exists (Cook, 1998; Yin et al., 2008). The *CS*, which represents the minimal subspace preserving the original information of $Y|\mathbf{X}$, is unique and the main focus of dimension reduction. Let $\mathbf{Z} = \Sigma_{\mathbf{X}}^{-\frac{1}{2}}(\mathbf{X} - E(\mathbf{X}))$, where $\Sigma_{\mathbf{X}}$ is the covariance matrix of $\mathbf{X}$, assumed to be positive definite. Then $\Sigma_{\mathbf{X}}^{-\frac{1}{2}}\mathcal{S}_{Y|\mathbf{Z}} = \mathcal{S}_{Y|\mathbf{X}}$. Hence, without loss of generality, we may work at either $\mathbf{Z}$- or $\mathbf{X}$-scale.

Sliced inverse regression (Li, 1991, SIR) is the first and most well-known method for sufficient dimension reduction. It investigates the trajectory of the inverse mean curve $E(\mathbf{Z}|Y)$. Under the so-called linearity condition that $E(\mathbf{Z}|\mathbf{B}^T\mathbf{Z})$ is linear in $\mathbf{B}^T\mathbf{Z}$, $\mathcal{S}_{E(\mathbf{Z}|Y)} \subseteq \mathcal{S}_{Y|\mathbf{Z}}$. Since then, many related studies have been carried out in both theory and applications. Hsing and Carroll (1992) established the asymptotic properties of SIR estimates when each slice only contains 2 observations. Zhu and Ng (1995) extended this idea to allow for a fixed number of observations per slice. Zhu and Fang (1996) bypassed the slicing step and used kernel smoothing to estimate $\text{cov}[E(\mathbf{Z}|Y)]$. Bura and Cook (2001a) suggested a parametric approach called parametric inverse regression. Fung et al. (2002) developed a variant version of SIR, CANCOR, where $B$-spline basis function replaced simple slicing. Xia et al. (2002) proposed an alternative derivation of SIR through the combination of local

* Corresponding author.
*E-mail addresses:* qwang3@vcu.edu (Q. Wang), xryin@stat.uga.edu (X. Yin).

0167-9473/$ – see front matter 
doi:10.1016/j.csda.2010.10.022

linear expansion and projection pursuit, known as inverse minimum average variance estimation (IMAVE). Bura (2003) also used local linear smoother to estimate the inverse mean function. On the other hand, Schott (1994), Velilla (1998), Bura and Cook (2001b) and Zhu et al. (2006) developed different methods to estimate the structural dimension $d_{Y|\mathbf{X}}$, under different scenarios. SIR is a powerful method due to its simplicity. However, it still has limitations. One of the issues is estimation efficiency. The finite-sample performance of SIR is not very satisfactory when the dimension is more than 2 and can be poor for heteroscedastic models.

A new direction for sufficient dimension reduction that deserves serious consideration is functional data analysis. See Ferraty and Vieu (2006) for an extensive review on functional data. Due to infinite dimensional in functional data, one technical difficulty is in inverting the ill-conditioned covariance matrix. To overcome this issue, Ferré and Yao (2003, 2005) replaced the matrix by a sequence of finite rank operators, with bounded inverse and converging to the covariance matrix, and used an equivalent eigen-space combining with a generalized inverse to avoid the inversion of the functional covariance matrix, respectively. Ait-Saidi et al. (2008) investigated dimension reduction methods assuming a single index functional model. Amato et al. (2006) extended SIR and others to functional data through appropriate wavelet decompositions. Recently, a platform from finite to infinite dimensional settings for inverse regression dimension reduction problem is provided by Hsing and Ren (2009) using an RKHS formulation.

In this article, we propose the use of orthogonal series to estimate the inverse mean function. As a useful nonparametric method, orthogonal series estimation is computationally efficient and can improve the estimation accuracy of SIR significantly, especially for the heteroscedastic models. Adopting the covariance matrix estimation techniques proposed by Ferré and Yao (2003, 2005), our method could be applied to functional data as well. The rest of the article is organized as follows. Section 2 gives a brief review of the estimation of SIR. The new approach based on orthogonal series estimation is detailed in Section 3. Section 4 introduces a Lasso type procedure to select informative variables. Section 5 discusses the permutation procedure used to choose the structural dimension. Simulation studies and a real data example are in Section 6. Section 7 concludes our discussion.

## 2. A brief review

Let $\{(\mathbf{X}_i^T, Y_i), i = 1, \dots, n\}$ be a random sample from $(\mathbf{X}^T, Y)$, where $\mathbf{X} = (X_1, \dots, X_p)^T \in \mathcal{R}^p$ and $Y \in \mathcal{R}$, and assume that $d_{Y|\mathbf{X}}$ is known. The SIR algorithm proposed by Li (1991) can be summarized as follows:

1. Standardize $\mathbf{X}_i$:$\mathbf{Z}_i = \hat{\mathbf{\Sigma}}_{\mathbf{X}}^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}})$, where $\bar{\mathbf{X}}$ and $\hat{\Sigma}_{\mathbf{X}}$ are the sample mean and sample covariance matrix respectively. Then divide $Y_i$ for $i = 1, \dots, n$ into $H$ slices and let $\hat{p}_h$ be the proportion of $Y_i$ that falls in slice $h \in \{1, 2, \dots, H\}$;
2. Within each slice $h$, compute the sample mean of $\mathbf{Z}$ and denote by $\bar{\mathbf{Z}}_h$. Form a sample SIR matrix $\hat{V} = \sum_{h=1}^{H} \hat{p}_h \bar{\mathbf{Z}}_h \bar{\mathbf{Z}}_h^T$, and find the eigen-structure of $\hat{V}$;
3. The $d_{Y|\mathbf{X}}$ eigenvectors $(\hat{\eta}_i, i = 1, \dots, d_{Y|\mathbf{X}})$ corresponding to the $d_{Y|\mathbf{X}}$ largest eigenvalues are the estimated directions of $\mathcal{S}_{E(\mathbf{Z}|Y)}$. Back to the $\mathbf{X}$ scale, $\hat{\beta}_i = \Sigma_{\mathbf{X}}^{-\frac{1}{2}} \hat{\eta}_i, i = 1, \dots, d_{Y|\mathbf{X}}$.

In this article, we use orthogonal series, a more flexible nonparametric tool, to estimate the inverse mean function. Our approach can be regarded as an alternative to the *Principal Fitted Component* model, proposed by Cook (2007) and Cook and Forzani (2008). In particular, one model they proposed is

$$\mathbf{X}_y = E(\mathbf{X}) + \Gamma \alpha f_y + \sigma \epsilon,$$

where $\mathbf{X}_y$ denotes a random vector distributed as $\mathbf{X}|Y = y$, $\Gamma \in \mathcal{R}^{p\times d}$, $d < p$, $\Gamma^T\Gamma = I_d$, $\alpha \in \mathcal{R}^{d\times r}$ and $d \le r$. $f_y \in \mathcal{R}^r$ is a known vector-valued function of response with $\sum_y f_y = 0$, $\sigma > 0$ and the error vector $\epsilon \in \mathcal{R}^p$. Inverse regression plots of $\mathbf{X}_y$ versus $y$ can be used to find suitable choices of $f_y$, then the sufficient dimension reduction subspace $\mathcal{S}(\Gamma)$ is estimated from the maximum likelihood. The authors also mentioned other possibilities for basis functions to be used for $f_y$. Rather than just estimating $f_y$, we use orthogonal series to estimate the inverse mean function without any particular model assumption.

## 3. Alternative estimation for SIR

### 3.1. Orthogonal series estimation

Suppose that a regression function of $y$ given $t$ can be represented as $y = \mu(t) + \epsilon$, where $\mu(t)$ is the mean function and $\epsilon$ is the random error. If it is reasonable to assume that $\mu(t)$ is a smooth function, many classes of functions can then be used to approximate $\mu(t)$. In general,

$$y = \sum_{j=0}^{\infty} \theta_j \varphi_j(t) + \epsilon, \tag{2}$$

where $\{\varphi_j\}$ is a basis function and $\theta_j$'s are the unknown Fourier coefficients. Once a basis function is chosen, the estimation of $\mu(t)$ is equivalent to the estimation of those Fourier coefficients. In practice, not all of them are estimable since only a finite number of observations are available. The approximation $\hat{\mu}(t) = \sum_{j=0}^{J} \hat{\theta}_j \varphi_j(t)$ is often used and known as a series estimator. More details on the properties of series estimator and the choice of smoothing parameter $J$ can be found in Härdle (1990) and Eubank (1999).