Hiroshima Math. J.
41 (2011), 41–53

Estimation on inverse regression using principal components of covariance matrix of sliced data

Tomoyuki Akita

(Received June 3, 2010)
(Revised August 20, 2010)

Abstract. Li (1991) introduced the so called sliced inverse regression (SIR) which reduces the dimension of the input variable in regression analysis, without any model-fitting process. Akita et al. (2009) proposed an improvement of SIR, called SIR, which uses conditional covariate matrices. In this paper, developing the approach of PCA-SIR, we propose yet another improvement of SIR, which we call PCA-SIR2. Simulation results produced by SIR, PCA-SIR and PCA-SIR2 are compared.

1. Introduction

We consider regression problem with the response variate y and p-dimensional covariate $\boldsymbol{x} = (x_1, \dots, x_p)'$. Li (1991) introduced the so called *sliced inverse regression* (SIR) for the linear or non-linear model

$$y = f(\boldsymbol{\beta}_1'\boldsymbol{x}, \boldsymbol{\beta}_2'\boldsymbol{x}, \dots, \boldsymbol{\beta}_K'\boldsymbol{x}, \varepsilon), \tag{1}$$

where $f(\cdot)$ is an unknown link function and $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K$ are unknown parameter vectors in $\mathbf{R}^p$, and ε is the error, an unobservable random variable which is independent of $\boldsymbol{x}$. For $p \times p$ matrix A, we write $\mathscr{R}[A]$ for the subspace of $\mathbf{R}^p$ spanned by the column vectors of A. SIR enables us to estimate the basis of

$$\mathscr{B} = \mathscr{R}[\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K].$$

This space $\mathscr{B}$, which is the subspace of $\mathbf{R}^p$ spanned by $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K$, is called the *effective dimension reduction space* (EDR).

It should be noted that, within the framework of (1), we have various popular regression models, such as the linear regression model

$$y = a + \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon, \tag{2}$$

the single index model

$$y = h(\boldsymbol{\beta}'\boldsymbol{x}, \varepsilon), \tag{3}$$

2000 *Mathematics Subject Classification*. Primary 62J02; Secondary 62H10.
Key words and phrases. non-linear model, sliced inverse regression, dimension reduction.

and the projection pursuit regression model

$$y = h_1(\boldsymbol{\beta}_1'\boldsymbol{x}) + h_2(\boldsymbol{\beta}_2'\boldsymbol{x}) + \cdots + h_K(\boldsymbol{\beta}_K'\boldsymbol{x}) + \varepsilon. \tag{4}$$

Let $\boldsymbol{\eta}(y) = E(\boldsymbol{x}|y) - E(\boldsymbol{x})$ be the inverse regression curve. If the covariate $\boldsymbol{x}$ is elliptically distributed, then SIR can be used to estimate $\mathscr{B}$. For, in this case, we have

$$S_{\boldsymbol{\eta}} \subset \mathscr{R}[\Sigma\boldsymbol{\beta}_1, \dots, \Sigma\boldsymbol{\beta}_K], \tag{5}$$

where $\Sigma = \mathrm{Cov}(\boldsymbol{x})$ and $S_{\boldsymbol{\eta}} = \{\boldsymbol{\eta}(y) : y \in \mathrm{Domain}(\boldsymbol{\eta})\}$ is the range of $\boldsymbol{\eta}$.

We put $\Sigma_{\boldsymbol{\eta}} = \mathrm{Cov}(\boldsymbol{\eta}(y))$. Let λ_i, $i = 1, \dots, p$, be the eigenvalues of the generalized eigenvalue problem

$$\Sigma_{\boldsymbol{\eta}}\boldsymbol{\gamma}_i = \lambda_i \Sigma \boldsymbol{\gamma}_i. \tag{6}$$

and let $\boldsymbol{\gamma}_i$, $i = 1, \dots, p$, be the corresponding eigenvectors. If $\lambda_i > 0$, then

$$\boldsymbol{\gamma}_i \in \mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \dots, \Sigma^{1/2}\boldsymbol{\beta}_K], \tag{7}$$

(see Li (1991)). In the estimation of the EDR space $\mathscr{B}$ via SIR, we calculate an "estimator of $\boldsymbol{\gamma}_i$" from samples. However, if $f(z, \varepsilon)$ is an even function of z, then $E(\boldsymbol{x}|y) = \mathbf{0}$, so that we cannot estimate $\mathscr{B}$ by SIR.

There are several methods for estimating $\mathscr{B}$. Among them are SIR, sliced average variance estimation (SAVE) (Cook and Weisberg, 1991), graphical regression (Cook, 1994, 1995), parametric inverse regression (Bura and Cook, 2001), partial SIR (Chiaromonte et al., 2002), ESAVE$_a$ (Zhu et al., 2007) and PCA-SIR (Akita et al., 2009). In this paper, developing the approach of PCA-SIR based on conditional covariance matrices, we propose yet another improvement of SIR which we call PCA-SIR2.

This paper is organized as follows. In section 2, we consider the sample covariance matrices formed from sliced samples and explain the theoretical background of PCA-SIR2. In section 3, we present the algorithm of PCA-SIR2. Simulation results to compare SIR, PCA-SIR and PCA-SIR2 are given in section 4. In section 5, we consider the convergence of estimators.

2. Theoretical background of estimation

Let $\{(y_i, \boldsymbol{x}_i) : i = 1, \dots, n\}$ be a set of data from model (1), $\mathscr{I}_h$ an interval of $\mathbf{R}$ depending on $h \in \{1, 2, \dots, \eta\}$, $I_h = \{i : y_i \in \mathscr{I}_h\}$, and $n_h = \#(I_h)$. We define the covariance matrix $\hat{\Sigma}_h$ of the sliced data $\{y_i\}_{i \in I_h}$ by

$$\hat{\Sigma}_h = \frac{1}{n_h}\sum_{i=1}^{n} 1_{\mathscr{I}_h}(y_i)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)' \qquad \text{with } \bar{\boldsymbol{x}}_h = \frac{1}{n_h}\sum_{i=1}^{n} 1_{\mathscr{I}_h}(y_i)\boldsymbol{x}_i, \tag{8}$$

where 1_A is the indicator function of A. We also define the sample covariance matrix $\hat{\Sigma}$ by

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)' \qquad \text{with } \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i. \tag{9}$$

The next theorem concerns the eigenvectors of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$.

Theorem 1. *Let H be a $p \times p$ orthogonal matrix whose column vectors consist of the normalized orthogonal basis of $\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \ldots, \Sigma^{1/2}\boldsymbol{\beta}_K]$ and that of its orthogonal complement. Then, as $n \to \infty$*

$$\hat{\Sigma}_h \to \Sigma^{1/2}H'\begin{pmatrix} \Lambda & O' \\ O & I_{p-K} \end{pmatrix}H\Sigma^{1/2} \qquad \text{in probability}, \tag{10}$$

where $\Lambda = \mathrm{diag}(\sigma_{h1}, \ldots, \sigma_{hK})$ with σ_{hi} being the (i,i) entry of

$$E[H\Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1/2}H' \mid y \in \mathscr{I}_h].$$

We can prove Theorem 1 easily by using law of large numbers. The implication of Theorem 1 is as follows: From Theorem 1, we see that the subspace spanned by K eigenvectors of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ converges to $\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \ldots, \Sigma^{1/2}\boldsymbol{\beta}_K]$, and the subspace spanned by the remaining $(p-K)$ eigenvectors to its orthogonal complement. Let $\lambda_{h1}, \ldots, \lambda_{hp}$ be the eigenvalues of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ and $\hat{\boldsymbol{h}}_{h1}, \ldots, \hat{\boldsymbol{h}}_{hp}$ the corresponding eigenvectors. Then K eigenvectors, say, $\hat{\boldsymbol{h}}_{h1}, \ldots, \hat{\boldsymbol{h}}_{hK}$ by changing the numbers if necessary, converge to the basis of $\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \ldots, \Sigma^{1/2}\boldsymbol{\beta}_K]$, and $\hat{\boldsymbol{h}}_{hK+1}, \ldots, \hat{\boldsymbol{h}}_{hp}$ to the basis of the orthogonal complement of $\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \ldots, \Sigma^{1/2}\boldsymbol{\beta}_K]$. Since this holds for any $h \in \{1, 2, \ldots, \eta\}$ and $\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \ldots, \Sigma^{1/2}\boldsymbol{\beta}_K]$ does not depend on h, we have, for $h, h' \in \{1, 2, \ldots, \eta\}$, $i = 1, 2, \ldots, K$ and $j = K+1, K+2, \ldots, p$,

$$(\hat{\boldsymbol{h}}_{hi}, \hat{\boldsymbol{h}}_{h'j})^2 \to 0 \qquad \text{as } n \to \infty,$$

so

$$\sum_{h' \neq h}^{\eta}\sum_{j=K+1}^{p}(\hat{\boldsymbol{h}}_{hi}, \hat{\boldsymbol{h}}_{h'j})^2 \to 0 \qquad \text{as } n \to \infty.$$

From the observation above, we are led to the algorithm PCA-SIR2 described in section 3. It enables us to find the basis of the EDR space $\mathscr{B}$.

3. PCA-SIR2

Given a set of data $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, our new method PCA-SIR2 proceeds in the following five steps:

Step 1. Divide the range of y into η slices

$$I_h = \{i : \ell_{h-1} \le y_i < \ell_h\}, \qquad h = 1, 2, \dots, \eta$$

where $-\infty = \ell_0 < \ell_1 < \cdots < \ell_{\eta-1} < \ell_\eta = \infty$.

Step 2. Calculate covariance matrices $\hat{\Sigma}_h$ in (8) and $\hat{\Sigma}$ in (9).

Step 3. For $h = 1, 2, \dots, \eta$, let $\lambda_{h1} \ge \lambda_{h2} \ge \cdots \ge \lambda_{hp}$ be the eigenvalues of matrix $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ and $\hat{\boldsymbol{h}}_{h1}, \dots, \hat{\boldsymbol{h}}_{hp}$ the corresponding normalised eigenvectors. Calculate the following squared inner products between eigenvectors of $\hat{\Sigma}_h$ and those of $\hat{\Sigma}_1, \dots, \hat{\Sigma}_{h-1}, \hat{\Sigma}_{h+1}, \dots, \hat{\Sigma}_\eta$:

$$(\hat{\boldsymbol{h}}_{hi}, \hat{\boldsymbol{h}}_{h'j})^2, \qquad h' \in \{1, 2, \dots, \eta\} \backslash \{h\}, \qquad i, j = 1, 2, \dots, p.$$

Step 4. For each $i = 1, 2, \dots, p$, withih each slice $h = 1, 2, \dots, \eta$,

$$z_1 = (\hat{\boldsymbol{h}}'_{hi}\hat{\boldsymbol{h}}_{11})^2,\ z_2 = (\hat{\boldsymbol{h}}'_{hi}\hat{\boldsymbol{h}}_{12})^2, \dots, z_{(\eta-1)p} = (\hat{\boldsymbol{h}}'_{hi}\hat{\boldsymbol{h}}_{\eta p})^2,$$

into $z_1^* \le z_2^* \le \cdots \le z_{(\eta-1)p}^*$ and calculate $c_i = \sum_{j=1}^{(\eta-1)(p-K)} z_j^*$. Let $c_{i_1}, \dots, c_{i_K}$ be the K smallest values of c_j's. We use subscript i_k in Step 5.

Step 5. We consider $\mathscr{R}[\hat{\boldsymbol{h}}_{hi_1}, \dots, \hat{\boldsymbol{h}}_{hi_K}]$ as the estimator of

$$\mathscr{R}[\Sigma^{1/2}\boldsymbol{\beta}_1, \dots, \Sigma^{1/2}\boldsymbol{\beta}_K].$$

Let A be $p \times (K\eta)$ matrix given by

$$A = (\hat{\boldsymbol{h}}_{11_1}, \dots \hat{\boldsymbol{h}}_{11_K}, \dots, \hat{\boldsymbol{h}}_{\eta\eta_1}, \dots, \hat{\boldsymbol{h}}_{\eta\eta_K}).$$

For the eigenvectors $\hat{\boldsymbol{\xi}}_1, \dots, \hat{\boldsymbol{\xi}}_K$ corresponding to the K largest eigenvalues of AA', output

$$\hat{\Sigma}^{1/2}\hat{\boldsymbol{\xi}}_1, \dots, \hat{\Sigma}^{1/2}\hat{\boldsymbol{\xi}}_K$$

as the basis of the EDR space $\mathscr{B}$.

4. Simulation results

In this section, we show some results of simulation studies which we conducted to compare SIR, PCA-SIR, and PCA-SIR2. We adapt the criterion proposed by Li (1991) to measure the distance between $\mathscr{R}[\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K]$ and $\mathscr{R}[\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_K]$. Let $B = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K)$ and $\hat{B} = (\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_K)$. Then the squared multiple correlation coefficient $R^2(\hat{B})$, which is the product of eigenvalues of $\hat{B}'B(B'B)^{-1}B'\hat{B}(\hat{B}'\hat{B})^{-1}$, is employed to measure the distance between the hyperplanes. If $K = 1$, then

$$R^2(\hat{\boldsymbol{\beta}}_1) = \frac{(\hat{\boldsymbol{\beta}}_1'\boldsymbol{\beta}_1)^2}{\|\hat{\boldsymbol{\beta}}_1\|^2\|\boldsymbol{\beta}_1\|^2}.$$

From 1000 runs of the Monte Carlo simulation, we report means of R^2.

Table 1. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^3 + \varepsilon$, $n = 100$, normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.970641	0.957759	0.217619	0.934050	0.051357
3	0.987819	0.985986	0.913249	0.975297	0.954613
4	0.991398	0.988077	0.952120	0.983929	0.985268
8	0.996129	0.958998	0.950359	0.992669	0.993622
12	0.997324	0.558445	0.939480	0.993599	0.994312
16	0.997810	0.071469	0.925421	0.992154	0.992863

Table 2. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^3 + \varepsilon$, $n = 200$, normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.985722	0.983108	0.206642	0.970314	0.035771
3	0.994223	0.994332	0.978660	0.988859	0.985630
4	0.996121	0.995853	0.985268	0.992773	0.994567
8	0.998345	0.997904	0.982537	0.997012	0.997480
12	0.998873	0.997872	0.978614	0.998004	0.998229
16	0.999089	0.996258	0.975286	0.998344	0.998457

Table 3. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + \varepsilon$, $n = 100$, normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.192511	0.981387	0.973642	0.901018	0.760304
3	0.185635	0.972527	0.969161	0.906328	0.909477
4	0.183586	0.968558	0.967148	0.840081	0.923895
8	0.217607	0.942713	0.949333	0.733058	0.915318
12	0.221793	0.903547	0.933142	0.609051	0.889294
16	0.222166	0.872368	0.917211	0.477671	0.849218

First, we consider the following models with $K = 1$:

(1) $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^3 + \varepsilon$,

(2) $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + \varepsilon$.

In these models, the covariate $\boldsymbol{x}$ and the error ε follow the normal distributions $N(\mathbf{0}, I_5)$ and $N(0, 0.1)$, respectively, where I_p is the $p \times p$ identity matrix. In this simulation, we put $\boldsymbol{\beta}_1 = (1, -1, 0, 0, 0)'$, and the results are shown in Tables 1–4. In the tables, PCA and PC2 mean PCA-SIR and PCA-SIR2, respectively. We see that the values of R^2 for PCA-SIR2 are very small when $H = 2$, so the method should not be used in this case. On the other hand, if $H \geq 3$, then the values of R^2 for PCA-SIR2 are nearly equal to 1 as those for other methods.

Table 4. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + \varepsilon,\ n = 200,$ normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.202815	0.992029	0.990366	0.971826	0.826442
3	0.199905	0.988165	0.987506	0.971914	0.968792
4	0.197157	0.985882	0.985542	0.933977	0.960459
8	0.206676	0.979370	0.980238	0.928101	0.974716
12	0.238394	0.972510	0.975074	0.904689	0.975755
16	0.245747	0.967654	0.971526	0.858137	0.974453

Table 5. $y = (\boldsymbol{\beta}_1'\boldsymbol{x}) + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon,\ n = 100,$ normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.548823	0.788497	0.629719	0.654998	0.409186
3	0.662408	0.818942	0.694473	0.663842	0.550517
4	0.665339	0.813094	0.758186	0.683578	0.556134
8	0.670205	0.685610	0.793902	0.683995	0.526532
12	0.668120	0.599823	0.781846	0.640539	0.509771
16	0.658361	0.553725	0.779070	0.601200	0.502736

Table 6. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon,\ n = 100,$ normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.367651	0.933492	0.905053	0.509906	0.557671
3	0.387267	0.919098	0.894192	0.544662	0.678187
4	0.374500	0.902262	0.886403	0.560528	0.667932
8	0.394949	0.812020	0.791434	0.570816	0.598108
12	0.395353	0.764159	0.737009	0.543973	0.544599
16	0.410187	0.711335	0.699161	0.522666	0.481134

Next, we considered the following models with $K = 2$:

(3) $y = (\boldsymbol{\beta}_1'\boldsymbol{x}) + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon,$

(4) $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon.$

The covariate $\boldsymbol{x}$ and the error ε follow the normal distribution $N(\mathbf{0}, I_{10})$ and $N(0, 0.1)$, respectively. In this simulation, $\boldsymbol{\beta}_1 = (1, -1, 0, \dots, 0)'$ and $\boldsymbol{\beta}_2 = (0, -1, 1, 0, \dots, 0)$, and the results are shown in Tables 5–8. We see that SAVE and $\mathrm{ESAVE}_{1/2}$ (ESAVE with parameter $\alpha = 1/2$) have good behaviors when $K = 2$ and the covariate is normally distributed, with the behavior of PCA-SIR2 is not so bad.

Finally, we consider some cases where the distribution of covariate is not normal. We take the same models as (1)–(4) above but the distribution of covariate $\boldsymbol{x}$ is replaced by a multivariate t-distribution. The error ε is normally distributed. The results are shown in Tables 9–12. From the simulation

Table 7. $y = (\boldsymbol{\beta}_1' \boldsymbol{x})^3 + \varepsilon$, $n = 100$, non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.938578	0.555209	0.196973	0.754648	0.067951
3	0.956962	0.576702	0.645265	0.891007	0.604149
4	0.964734	0.451926	0.768416	0.933611	0.890786
8	0.973892	0.111958	0.755786	0.956693	0.974071
12	0.977530	0.047813	0.692849	0.957635	0.972764
16	0.979310	0.034598	0.640405	0.939672	0.959533

Table 8. $y = (\boldsymbol{\beta}_1' \boldsymbol{x})^3 + \varepsilon$, $n = 200$, non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.965554	0.748075	0.192696	0.852529	0.055208
3	0.975228	0.785978	0.812536	0.939305	0.750028
4	0.979388	0.709462	0.886198	0.960360	0.950969
8	0.984761	0.333782	0.874620	0.976572	0.992109
12	0.987209	0.135501	0.828582	0.980614	0.994120
16	0.988560	0.064414	0.787917	0.982642	0.994590

Table 9. $y = (\boldsymbol{\beta}_1' \boldsymbol{x})^2 + \varepsilon$, $n = 100$, non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.194276	0.973348	0.959292	0.826993	0.596467
3	0.203009	0.940863	0.940968	0.869197	0.831137
4	0.197529	0.910557	0.915456	0.834491	0.866139
8	0.222312	0.785917	0.810268	0.693930	0.801604
12	0.239064	0.685352	0.721137	0.546928	0.706295
16	0.257178	0.615964	0.658510	0.494763	0.588851

Table 10. $y = (\boldsymbol{\beta}_1' \boldsymbol{x})^2 + \varepsilon$, $n = 200$, non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.204408	0.989556	0.987534	0.903803	0.659781
3	0.199815	0.978519	0.978429	0.949175	0.934868
4	0.204919	0.965646	0.966047	0.935784	0.943013
8	0.220897	0.905043	0.913564	0.874847	0.947013
12	0.243639	0.843446	0.862153	0.832493	0.943414
16	0.249150	0.783236	0.812710	0.774091	0.921320

Table 11. $y = (\boldsymbol{\beta}_1'\boldsymbol{x}) + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon,\ n = 100,$ non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.540453	0.662824	0.622784	0.684377	0.430781
3	0.627499	0.673844	0.684804	0.676313	0.521725
4	0.623323	0.659906	0.701966	0.674850	0.517324
8	0.622478	0.583122	0.675519	0.649888	0.497773
12	0.622994	0.541763	0.645940	0.615523	0.464434
16	0.613880	0.514178	0.622011	0.563132	0.444411

Table 12. $y = (\boldsymbol{\beta}_1'\boldsymbol{x})^2 + (\boldsymbol{\beta}_2'\boldsymbol{x})^2 + \varepsilon,\ n = 100,$ non-normal

	SIR	SAVE	ESAVE	PCA	PC2
2	0.360737	0.880715	0.867530	0.514416	0.453818
3	0.377416	0.829433	0.829160	0.520488	0.557645
4	0.370769	0.794728	0.798521	0.524498	0.561269
8	0.379395	0.698324	0.707542	0.533072	0.488556
12	0.395061	0.638532	0.656351	0.523567	0.449569
16	0.399511	0.597496	0.616404	0.506230	0.409344

results, we see that when $K = 1$ and the distribution of $\boldsymbol{x}$ is not normal, the behaviors of SIR and SAVE are not so good, while those of PCA-SIR and PCA-SIR2 are good.

5. Consistency

In this section, we consider an inner product of eigenvectors of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ and $\hat{\Sigma}^{-1/2}\hat{\Sigma}_{h'}\hat{\Sigma}^{-1/2}$ to prove the convergence of $\hat{\boldsymbol{\beta}}_1$ when $K = 1$. We consider the model

$$y = f(\boldsymbol{\beta}_1'\boldsymbol{x}, \varepsilon), \qquad \boldsymbol{\beta}_1 = (1, 0, \dots, 0)', \qquad \boldsymbol{x} \sim N_p(\boldsymbol{0}, I_p), \qquad \varepsilon \amalg \boldsymbol{x},$$

where $f(z, \varepsilon)$ is an even function of z.

Remark. The model

$$y = f(\boldsymbol{\beta}_1'\boldsymbol{x}, \varepsilon), \qquad \boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma), \qquad \varepsilon \amalg \boldsymbol{x},$$

is reduced to the above one by a suitable affine transformation.

Suppose that the data $(y_i, \boldsymbol{x}_i)$ $i = 1, \dots, n$, are given. As in section 2, we put

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i, \qquad n_h = \sum_{i=1}^{n} 1_h(y_i), \qquad \bar{\boldsymbol{x}}_h = \frac{1}{n_h}\sum_{i=1}^{n} 1_h(y_i)\boldsymbol{x}_i,$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})', \qquad \hat{\Sigma}_h = \frac{1}{n_h}\sum_{i=1}^{n} 1_h(y_i)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_h)',$$

where $1_h(\cdot) = 1_{[\ell_{h-1}, \ell_h)}(\cdot)$. We consider the asymptotic variance of the eigenvectors of a multiple eigenvalue in the same way as Anderson (1963).

We introduce some notation. Let $\sigma_h = E[x_1^2 \mid y \in \mathscr{I}_h]$, where x_1 is the first element of $\boldsymbol{x}$. Let γ_1 be an eigenvector corresponding to σ_h, and let $\gamma_2, \dots, \gamma_p$ be a basis of the eigenspace corresponding to the eigenvalue 1. Then $\Sigma_h = \operatorname{diag}(\sigma_h, \mathbf{1}'_{p-1})$. We put $\Gamma = (\gamma_1, \dots, \gamma_p)$ and $\Delta = \operatorname{diag}(\sigma_h, \mathbf{1}'_{p-1})$. Let $d_1 > \cdots > d_p$ be the eigenvalues of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ and $\boldsymbol{c}_i$ an eigenvector corresponding to d_i for $i = 1, 2, \dots, p$.

We consider the matrix $\Gamma'\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}\Gamma$. Since both $\hat{\Sigma}$ and $\hat{\Sigma}_h$ have asymptotic normal distributions,

$$\sqrt{n}(\Gamma'\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}\Gamma - \Gamma'\Sigma_h\Gamma)$$

converges to a multivariate normal distribution with mean $\mathbf{0}$. Eigenvalues and eigenvectors of $\Gamma'\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}\Gamma$ are $d_1, \dots, d_p$ and $\boldsymbol{e}_1 = \Gamma'\boldsymbol{c}_1, \dots, \boldsymbol{e}_p = \Gamma'\boldsymbol{c}_p$, respectively. Let $D = \operatorname{diag}(d_1, \dots, d_p)$ and $E = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_p)$. For uniqueness, we take the eigenvector $\boldsymbol{e}_i$ such that $e_{ii} > 0$. We put $T = \Gamma'\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}\Gamma$ and $U = \sqrt{n}(T - \Delta)$. Then $T = EDE'$ holds.

We write

$$\Delta = \begin{pmatrix} \sigma_h & \mathbf{0}' \\ \mathbf{0} & I_{p-1} \end{pmatrix}, \qquad U = \begin{pmatrix} u_{11} & \boldsymbol{u}'_{12} \\ \boldsymbol{u}_{21} & U_{22} \end{pmatrix}, \qquad E = \begin{pmatrix} e_{11} & \boldsymbol{e}'_{12} \\ \boldsymbol{e}_{21} & E_{22} \end{pmatrix},$$

$$D = \operatorname{diag}(d_1, d_2, \dots, d_p) = \begin{pmatrix} d_1 & \\ & D_2 \end{pmatrix},$$

$$p_1 = \sqrt{n}(d_1 - \sigma_h), \qquad P_2 = \sqrt{n}(D_2 - I_{p-1}), \qquad P = \begin{pmatrix} p_1 & \mathbf{0}' \\ \mathbf{0} & P_2 \end{pmatrix} = \sqrt{n}(D - \Delta),$$

$$\boldsymbol{f}'_{12} = \sqrt{n}\boldsymbol{e}'_{12}, \qquad \boldsymbol{f}_{21} = \sqrt{n}\boldsymbol{e}_{21}.$$

We are ready to state the next lemma.

Lemma 1. *We have, in probability,* $e_{11} \to 1$, $\boldsymbol{e}'_{12} \to \mathbf{0}'$ *and* $\boldsymbol{e}_{21} \to \mathbf{0}$, *as* $n \to \infty$. *The asymptotic distribution of* E_{22} *is the unique normalized Haar measure on the group* $O(p-1)$ *of* $(p-1) \times (p-1)$ *orthogonal matrices.*

We refer to Anderson (1963) for the details of the proof of Lemma 1. Here, we give an outline of it.

Since $EE' = I_p$, we have

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & I_{p-1} \end{pmatrix} = \begin{pmatrix} e_{11}^2 & \mathbf{0}' \\ \mathbf{0} & E_{22}E'_{22} \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} 0 & e_{11}\boldsymbol{f}'_{21} + \boldsymbol{f}'_{12}E'_{22} \\ e_{11}\boldsymbol{f}_{21} + E_{22}\boldsymbol{f}_{12} & O \end{pmatrix} + \frac{1}{n}W.$$

Comparing the blocks, we get

$$e_{11}^2 = 1 - \frac{1}{n} w_{11}, \qquad E_{22}E_{22}' = I_{p-1} - \frac{1}{n} W_{22},$$

$$\mathbf{0}' = e_{11}\boldsymbol{f}_{21}' + \boldsymbol{f}_{12}'E_{22}' + \frac{1}{\sqrt{n}}\boldsymbol{w}_{12}, \qquad \mathbf{0} = e_{11}\boldsymbol{f}_{21} + E_{22}\boldsymbol{f}_{12} + \frac{1}{\sqrt{n}}\boldsymbol{w}_{21}.$$

On the other hand, T can be expanded as

$$\begin{aligned} T &= \varDelta + \frac{1}{\sqrt{n}} U \\ &= E\left(\varDelta + \frac{1}{\sqrt{n}} P\right)E' \\ &= \begin{pmatrix} \sigma_h e_{11}^2 & \mathbf{0}' \\ \mathbf{0} & E_{22}E_{22}' \end{pmatrix} \\ &\quad + \frac{1}{\sqrt{n}} \begin{pmatrix} p_1 e_{11}^2 & \sigma_h e_{11}\boldsymbol{f}_{21}' + \boldsymbol{f}_{12}'E_{22}' \\ \sigma_h e_{11}\boldsymbol{f}_{21} + E_{22}\boldsymbol{f}_{12} & E_{22}P_2E_{22}' \end{pmatrix} + \frac{1}{n} M. \end{aligned}$$

Hence,

$$\begin{aligned} T &= \varDelta - \frac{1}{n} \begin{pmatrix} \sigma_h w_{11} & \mathbf{0}' \\ \mathbf{0} & W_{22} \end{pmatrix} \\ &\quad + \frac{1}{\sqrt{n}} \begin{pmatrix} p_1 e_{11}^2 & \sigma_h e_{11}\boldsymbol{f}_{21}' + \boldsymbol{f}_{12}'E_{22}' \\ \sigma_h e_{11}\boldsymbol{f}_{21} + E_{22}\boldsymbol{f}_{12} & E_{22}P_2E_{22}' \end{pmatrix} + \frac{1}{n} M, \end{aligned}$$

then

$$u_{11} = p_1 e_{11}^2 + \frac{1}{\sqrt{n}}(m_{11} - \sigma_h), \qquad U_{22} = E_{22}P_2E_{22}' + \frac{1}{\sqrt{n}}(M_{22} - W_{22}),$$

$$\boldsymbol{u}_{12}' = \sigma_h e_{11}\boldsymbol{f}_{21}' + \boldsymbol{f}_{21}'E_{22}' + \frac{1}{\sqrt{n}}\boldsymbol{m}_{12}', \qquad \boldsymbol{u}_{21} = \sigma_h e_{11}\boldsymbol{f}_{21} + E_{22}\boldsymbol{f}_{12} + \frac{1}{\sqrt{n}}\boldsymbol{m}_{12}.$$

We see that E_{22} is asymptotically orthogonal and the limiting distributions of $p_1 e_{11}^2$ and $E_{22}P_2E_{22}'$ are equal to those of u_{11} and U_{22}, respectively. We also find that $E_{22}\boldsymbol{f}_{21}$ and $-E_{22}\boldsymbol{f}_{12}$ are asymptotically equivalent, and the limiting distribution of both coincides with that of $(1/(\sigma_h - 1))\boldsymbol{u}_{12}$. So, we may consider the asymptotic variance of U_{22} to derive the asymptotic distribution of E_{22}.

Using the asymptotically normal random variables

$$V_h = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} 1_h(y_i)\boldsymbol{x}_i\boldsymbol{x}_i' - p_h\Sigma_h\right\}, \qquad V_h = (v_{ij}^{(h)}),$$

$$\boldsymbol{z}_h = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} 1_h(y_i)\boldsymbol{x}_i\right\}, \qquad q_h = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} 1_h(y_i) - p_h\right\},$$

$\hat{\Sigma}_h$ and $\hat{\Sigma}^{-1/2}$ can be expanded as

$$\hat{\Sigma}_h = \Sigma_h + \frac{1}{\sqrt{n}}\frac{1}{p_h}(V_h - q_h\Sigma_h) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$\hat{\Sigma}^{-1/2} = I - \frac{1}{2\sqrt{n}}\sum_{h=1}^{H} V_h + o_p\left(\frac{1}{\sqrt{n}}\right),$$

whence $\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2}$ as

$$\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2} = \Sigma_h + \frac{1}{\sqrt{n}}\left\{\frac{1}{p_h}(V_h - q_h\Sigma_h) - \frac{1}{2}\sum_{k=1}^{H}(V_k\Sigma_h + \Sigma_h V_k)\right\} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

In the next Lemma, we list some of the expectations which we need later.

LEMMA 2. *Let* $1 \le h \le k \le \eta$, *and* $i, j, l, m \ge 2$, $i \ne j$. *Then*

$$E[v_{ii}^{(h)}v_{ii}^{(k)}] = \begin{cases} p_h(3 - p_h) & (h = k) \\ -p_h p_k & (h \ne k) \end{cases}, \qquad E[v_{ii}^{(h)}v_{jj}^{(k)}] = \begin{cases} p_h(1 - p_h) & (h = k) \\ -p_h p_h & (h \ne k) \end{cases},$$

$$E[v_{ij}^{(h)}v_{ij}^{(k)}] = \begin{cases} p_h & (h = k) \\ 0 & (h \ne k) \end{cases}, \qquad E[v_{ii}^{(h)}v_{jl}^{(k)}] = 0, \qquad E[v_{ij}^{(h)}v_{lm}^{(h)}] = 0,$$

$$E[q_h v_{ii}^{(k)}] = \begin{cases} p_h(1 - p_h) & (h = k) \\ -p_h p_k & (h \ne k) \end{cases}, \qquad E[q_k q_h] = \begin{cases} p_h(1 - p_h) & (h = k) \\ -p_h p_k & (h \ne k) \end{cases},$$

$$E[q_h v_{ij}^{(k)}] = 0.$$

We define $W^{(h)} = (w_{ij}^{(h)})$ by $W^{(h)} = \sqrt{n}(\hat{\Sigma}^{-1/2}\hat{\Sigma}_h\hat{\Sigma}^{-1/2} - \Sigma_h)$. Then $W^{(h)}$ converges to normal random matrix with mean $\mathbf{0}$. From Lemma 2, we have, for $i, j \ge 2$,

$$E[w_{ii}^{(h)}w_{ii}^{(k)}] = \begin{cases} 2(p_h - 1)/p_h & (h = k) \\ -2 & (h \ne k) \end{cases},$$

$$E[w_{ij}^{(h)}w_{ij}^{(k)}] = \begin{cases} (p_h - 1)/p_h & (h = k) \\ -2 & (h \ne k) \end{cases},$$

$$E[w_{ii}^{(h)}w_{jl}^{(k)}] = E[w_{il}^{(h)}w_{jm}^{(k)}] = 0.$$

We see that the limiting marginal distribution of $U_{22}^{(h)}$ is normal with expectation $\mathbf{0}$, and the density of U_{22} is proportional to

$$\exp\left(-\frac{1}{2}\left(\sum_{i=2}^{p} u_{ii}^2 + \sum_{2\le i<j} u_{ij}^2\right)\right) = \exp\left(-\frac{1}{2}(\mathrm{tr}(U_{22}))^2\right).$$

Thus, the limiting joint distribution of (E_{22}, P_2) is given by

$$\exp\left(-\frac{1}{2}(\mathrm{tr}(E_{22}P_2E_{22}'))^2\right)|J(U_{22} \to (E_{22}, P_2))|$$
$$= K(P_2)\exp\left(-\frac{1}{2}(\mathrm{tr}(P_2))^2\right)\prod_{i<j}(p_i - p_j),$$

where $J(U_{22} \to (E_{22}, P_2))$ is the Jacobian which depends on only P_2 (See, Muirhead (1982)). The distribution of T_2 is the same as that of $Q'T_2Q$ for any orthogonal matrices Q. Hence the distribution of EQ is the same as that of E except for the effect of $e_{ii} > 0$. This implies that the distribution of E is the unique normalized Haar measure on $O(p-1)$.

If

$$p_h + p_k \neq 1, \qquad p_h + p_k + 3p_hp_k \neq 1, \tag{11}$$

then the asymptotic covariance matrix of joint distribution of $W^{(h)}$ and $W^{(k)}$ becomes nonsingular. On the other hand, if (11) does not hold, then it becomes singular. In particular, if $\eta = 2$, then the asymptotic covariance matrix becomes singular. Hence the estimation via PCA-SIR2 may fail if $\eta = 2$. In fact, the simulation results on R^2 for $\eta = 2$ in section 4 suggest that this is the case. When $\eta = 2$, which is the singular case, the eigenvectors of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_2\hat{\Sigma}^{-1/2}$ are determined by the eigenvectors of $\hat{\Sigma}^{-1/2}\hat{\Sigma}_1\hat{\Sigma}^{-1/2}$. In the nonsingular case, the two vectors are not bisected at right angle with probability one. So, let

$$E^{(h)} = (\boldsymbol{e}_1^{(h)}, \dots, \boldsymbol{e}_p^{(h)}), \qquad E^{(k)} = (\boldsymbol{e}_1^{(k)}, \dots, \boldsymbol{e}_p^{(k)}),$$

be the matrices of the eigenvectors of the h-th and k-th slices. Then, for $i, j \ge 2$,

$$(\boldsymbol{e}_1^{(h)})'\boldsymbol{e}_1^{(k)} = 1 + o_p(n^{-1}), \qquad (\boldsymbol{e}_1^{(h)})'\boldsymbol{e}_j^{(k)} = o_p(n^{-1/2}),$$

and $((\boldsymbol{e}_i^{(h)})'\boldsymbol{e}_j^{(k)})^2$ has a non-degenerate distribution on $[0, 1]$. Thus if n is enough large, then the matrix A in step 5 of section 3 is made of $\boldsymbol{e}_1^{(h)}$'s. Thus, the estimator $\hat{\Sigma}^{1/2}\hat{\boldsymbol{\xi}}_1$ in PCA-SIR2 converges to $\boldsymbol{\beta}_1$ in probability. In general, the case $K \ge 2$, similar argument may be hold.

If $f(z, \varepsilon)$ is an odd function of z, then the asymptotic behavior of the covariance matrix becomes very complex. We leave the problem of consistency in this case open here.

Acknowledgement

The author is deeply grateful to his supervisor, Professor Hirofumi Wakaki, for his continual guidance and valuable ideas and comments.

References

[1] Akita, T., Tanaka, S., Kawanishi, D., and Wakaki, H. (2009). An improvement method of sliced inverse regression using principal component analysis. Journal of Statistics and Application. Vol. 4 **2-3** 489–498.

[2] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. Ann. Math. Statist. **34** 122–148.

[3] Bura, E. and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. J. R. Stat. Soc. Ser. B Stat. Methodol. **63** 393–410.

[4] Chiaromonte, F., Cook, R. D. and Li, B. (2002). Sufficient dimension reduction in regressions with categorical predictors. Ann. Statist. **30** 475–497.

[5] Cook, R. D. (1994). On the interpretation of regression plots. J. Amer. Statist. **89** 177–189.

[6] Cook, R. D. (1998). Regression Graphics. Wiley, New York.

[7] Cook, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. Ann. Statist. **32** 1062–1092.

[8] Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction," by K. C. Li. J. Amer. Statist. **86** 328–339.

[9] Duan, N. and Li, K. C. (1991). Slicing regression: a link free regression method. Ann. Statist. **19** 505–530.

[10] Li, K. C. (1991). Sliced inverse regression for dimension reduction. (with discussion). J. Amer. Statist. Assoc. **86** 316–342.

[11] Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. John Wiley & Sons, New York.

[12] Zhu, L. X., Ohtaki, M. and Li, Y. (2007). On hybrid methods of inverse regression-based algorithms. Comput. Statist. **51** 2621–2635.

Tomoyuki Akita
Department of Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526, *Japan*
E-mail: d082905@hiroshima-u.ac.jp