Communications of the Korean Statistical Society
2012, Vol. 19, No. 3, 501–506
DOI: http://dx.doi.org/10.5351/CKSS.2012.19.3.501

Investigating SIR, DOC and SAVE for the Polychotomous Response

Hakbae Lee[1,a], Heemin Lee[a]

[a]Department of Applied Statistics, Yonsei University

Abstract

This paper investigates the central subspace related with SIR, DOC and SAVE when the response has more than two values. The subspaces constructed by SIR, DOC and SAVE are investigated and compared. The SAVE paradigm is the most comprehensive. In addition, the SAVE coincides with the central subspace when the conditional distribution of predictors given the response is normally distributed.

Keywords: Sufficient dimension reduction, central subspace, SIR, DOC, SAVE.

1. Introduction

The sufficient dimension reduction without loss of the original regression information is summarized by the central subspaces (Cook,1994) containing all information on the regression. Sliced Inverse Regression (SIR; Li, 1991), Principal Hessian Directions (pHd; Li, 1992) and Sliced Average Variance Estimation (SAVE; Cook and Weisberg, 1991) are some well known methods to estimate the central subspace in regression. Cook and Lee (1999) suggested Difference of Covariances(DOC) when the response has only two values. This paper investigates the central subspace related with SIR, DOC and SAVE when the response has more than two values.

Consider a regression problem consisting of a univariate response variable Y and a $p \times 1$ random vector of predictors $X = (X_1, \ldots, X_p)^T \in R^p$. Let $\boldsymbol{\eta}$ denote a fixed $p \times q$, $q \leqslant p$ matrix so that

$$Y \perp\!\!\!\perp X \,|\, \boldsymbol{\eta}^T X$$

This statement says that the distribution of $Y \,|\, X$ is the same as that of $Y \,|\, \boldsymbol{\eta}^T X$ for all values of X in its marginal sample space. It implies that the $p \times 1$ predictor X can be replaced by the $q \times 1$ predictor vector $\boldsymbol{\eta}^T X$ without loss of the original regression information, and represents a useful reduction in the dimension of the predictor vector.

Let $P_{\boldsymbol{\eta}}$ denote the orthogonal projection onto the subspace constructed by $\boldsymbol{\eta}$ and $Q_{\boldsymbol{\eta}} = I - P_{\boldsymbol{\eta}}$. Let Z denote the standardized predictor of X: $Z = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}(X - \mu)$ where $\boldsymbol{\mu} = E(X)$ and $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathrm{Cov}(X)$. Cook (1994) suggested the foundation of dimension reduction and the central subspace as follows:

Let $\mathcal{S}$ denote a subspace, and $\mathcal{S}(\boldsymbol{\eta})$ denote the subspace constructed by $\boldsymbol{\eta}$.

If $Y \perp\!\!\!\perp X \,|\, \boldsymbol{\eta}^T X$, then $\mathcal{S}(\boldsymbol{\eta})$ is defined to be a dimension reduction subspace(DRS) for the regression of Y on X.

This research was supported by Fund for Supporting Basic Science Research in the College of Business and Economics, Yonsei University of 2010(2010-1-0161).

[1] Corresponding author: Associate Professor, Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea. E-mail: hblee@yonsei.ac.kr

If $\mathcal{S}_{y|x}$ is a DRS and $\mathcal{S}_{y|x} \subset \mathcal{S}_{drs}$ for all DRSs $\mathcal{S}_{drs}$, then a subspace $\mathcal{S}_{y|x}$ is defined to be the central subspace for the regression of Y on X.

Let $\mathcal{S}_{y|x}(\boldsymbol{\eta})$ with the basis $\boldsymbol{\eta}$ be the central subspace for the regression of Y on X, and let $\mathcal{S}_{y|z}$ be the central subspace for the regression of Y on $Z = A^T X$ where A is a full rank, $p \times p$ matrix. Then $\mathcal{S}_{y|z} = A^{-1}\mathcal{S}_{y|x}$.

If $E(Z \mid \eta^T Z) = P_\eta Z$(linearity condition), then $E(Z \mid Y) \in \mathcal{S}_{y|z}$ and $\mathrm{Var}(Z \mid Y) = Q_\eta + P_\eta \mathrm{Var}(Z \mid Y) P_\eta$. If $E(Z \mid \eta^T Z) = P_\eta Z$ and $\mathrm{Var}(Z \mid \eta^T Z) = Q_\eta$(constant covariance condition), then $E(Z \mid Y) \in \mathcal{S}_{y|z}$ and $S(I_p - \mathrm{Var}(Z \mid Y)) \subset \mathcal{S}_{y|z}$.

$E(Z \mid \eta^T Z) = P_\eta Z$ and $\mathrm{Var}(Z \mid \eta^T Z) = Q_\eta$ hold when X is normally distributed (Cook, 1998).

Let $\boldsymbol{\mu}_j = E(Z \mid Y = j)$, $\boldsymbol{\Sigma}_j = \mathrm{Cov}(Z \mid Y = j)$, $j = 1, \ldots, g$, and $f_j = \Pr(Y = j)$. We assume $0 < f_j < 1$ and $\sum_{j=1}^{g} f_j = 1$. Finally, let $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{g-1})$ and $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_{g-1})$ where $\boldsymbol{\nu}_j = \boldsymbol{\mu}_{j+1} - \boldsymbol{\mu}_j$ and $\boldsymbol{\Delta}_j = \boldsymbol{\Sigma}_{j+1} - \boldsymbol{\Sigma}_j$ for $j = 1, \ldots, g-1$.

2. SIR, DOC, SAVE and Central Subpace

2.1. Sliced Inverse Regression(SIR)

To review the ideas behind SIR (Li, 1991), we assume that the response Y is continuous. SIR is based on a discrete version of Y: the range of Y is partitioned into g fixed, non-overlapping slices, $J_1, \ldots, J_g$, and Y is replaced with a discrete response $\tilde{Y} = s$ when $Y \in J_s$ for $s = 1, \ldots, g$. Clearly, because $\tilde{Y}$ is a function of Y, $\mathcal{S}_{\tilde{y}|x} \subset \mathcal{S}_{y|x}$ where $\mathcal{S}_{\tilde{y}|x}$ is the central subspace for the regression of $\tilde{Y}$ on X. In practice, SIR is based on computing the intraslice averages of the standardized predictors Z. In this paper, since the response Y is polychotomous with g values, the kernel matrix of SIR is given by

$$\sum_{j=1}^{g} \Pr(Y = j) E(Z \mid Y = j) E(Z \mid Y = j)^T = \sum_{j=1}^{g} f_j \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T.$$

When the response has more than two values, the relation between the central subspace, $\mathcal{S}_{y|x}$ and the subspace constructed by SIR, $\mathcal{S}_{SIR}$ is summarized by the following result.

Proposition 1. *Let $\mathcal{S}_{SIR}$ denote the subspace constructed by SIR, and the linearity condition $E(Z \mid \eta^T Z) = P_\eta Z$ hold. Then*

$$\mathcal{S}_{SIR} = \mathcal{S}\left(\sum_{j=1}^{g} f_j \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T\right) = \mathcal{S}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}) = \mathcal{S}(\boldsymbol{\nu}) \subset \mathcal{S}_{y|x}. \tag{2.1}$$

Proof: Because $\mathcal{S}(A) = \mathcal{S}(AA^T)$ for any matrix A and $\mathcal{S}(A) = \mathcal{S}(AB)$ for a nonsingular matrix B,

$$\begin{aligned}
\mathcal{S}_{SIR} &= \mathcal{S}\left(\sqrt{f_1}\boldsymbol{\mu}_1, \sqrt{f_2}\boldsymbol{\mu}_2, \ldots, \sqrt{f_g}\boldsymbol{\mu}_g\right) = \mathcal{S}\left(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_g\right) \\
&= \mathcal{S}\left(\left(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_g\right)\begin{pmatrix} f_1 & -1 & 0 & \ldots & 0 \\ f_2 & 1 & -1 & \ldots & 0 \\ f_3 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{g-1} & 0 & 0 & \ldots & -1 \\ f_g & 0 & 0 & \ldots & 1 \end{pmatrix}\right) \\
&= \mathcal{S}(\mathbf{0}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}) = \mathcal{S}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}) = \mathcal{S}(\boldsymbol{\nu}).
\end{aligned}$$

The third equality holds since the post-multiplied matrix is full rank and $\sum_{j=1}^{g} f_j \boldsymbol{\mu}_j = 0$. The direct application of linearity condition $E(Z \mid \eta^T Z) = P_\eta Z$ shows $\mathcal{S}_{SIR} \subset \mathcal{S}_{y|x}$. Hence the result follows. □

This proposition shows that the subspace $\mathcal{S}_{SIR}$ constructed by SIR coincides with the subspace by $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1})$ when the response has more than two values. Furthermore, SIR provides a part of the central subspace $\mathcal{S}_{y|x}$.

2.2. Differences of Covariances(DOC)

For the binary response, the subspace constructed by DOC (Cook and Lee, 1999) is

$$\mathcal{S}_{DOC} = \mathcal{S}(\mathrm{Cov}(Z \mid Y = 2) - \mathrm{Cov}(Z \mid Y = 1)) = \mathcal{S}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1) = \mathcal{S}(\boldsymbol{\Delta}_1).$$

For the polychotomous response Y with g values, the kernel matrix of DOC is given by

$$\boldsymbol{\Delta} = \left(\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_{g-1}\right).$$

When the response Y is polychotomous, the relation between the central subspace, $\mathcal{S}_{y|x}$ and the subspace constructed by DOC, $\mathcal{S}_{DOC}$ is summarized by the following result.

Proposition 2. *Let $\mathcal{S}_{DOC}$ denote the subspace constructed by DOC, and also the linearity and constant covariance conditions, that is, $E(Z \mid \eta^T Z) = P_\eta Z$ and $\mathrm{Var}(Z \mid \eta^T Z) = Q_\eta$ hold. Then*

$$\mathcal{S}_{DOC} = \mathcal{S}(\boldsymbol{\Delta}) = \mathcal{S}\left(\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_{g-1}\right) \subset \mathcal{S}_{y|x}. \tag{2.2}$$

Proof: Since the linearity and constant covariance conditions hold, it is obvious that $\mathcal{S}_{DOC} \subset \mathcal{S}_{y|x}$ by Cook and Lee (1999). Hence the result follows. □

This proposition implies that the subspace $\mathcal{S}_{DOC}$ constructed by DOC provides a part of the central subspace $\mathcal{S}_{y|x}$.

2.3. Sliced Average Variance Estimation(SAVE)

Cook and Weisberg (1991) proposed SAVE to overcome the inability of SIR to detect certain types of nonlinear regression relationships. Let us consider the population kernel matrix for SAVE to be

$$\boldsymbol{\Omega} = E\left\{\left(\mathbf{I}_p - \boldsymbol{\Sigma}_y\right)^2\right\} = \sum_{j=1}^{g} \Pr(Y = j)\left(\mathbf{I}_p - \boldsymbol{\Sigma}_j\right)^2.$$

For the binary response, the subspace constructed by SAVE (Cook and Lee, 1999) is

$$\mathcal{S}_{SAVE} = \mathcal{S}(\boldsymbol{\Omega}) = \mathcal{S}_{DOC} = \mathcal{S}\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1\right) = \mathcal{S}\left(\boldsymbol{\nu}_1, \boldsymbol{\Delta}_1\right).$$

For the polychotomous response Y with g values, the kernel matrix of SAVE is given by

$$\boldsymbol{\Omega} = \left(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}, \boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{g-1}\right) = (\boldsymbol{\nu}, \boldsymbol{\Delta}).$$

When the response Y is polychotomous, the following proposition shows the relation between the central subspace, $\mathcal{S}_{y|x}$ and the subspace constructed by SAVE, $\mathcal{S}_{SAVE}$.

Proposition 3. *Let $\mathcal{S}_{SAVE}$ denote the subspace constructed by SAVE, and also the linearity and constant covariance conditions, that is, $E(Z\,|\,\eta^T Z) = P_\eta Z$ and $\mathrm{Var}(Z\,|\,\eta^T Z) = Q_\eta$ hold. Then*

$$\mathcal{S}_{SAVE} = \mathcal{S}(\mathbf{\Omega}) = \mathcal{S}(\boldsymbol{\nu}, \mathbf{\Delta}) = \mathcal{S}\left(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}, \mathbf{\Delta}_1, \mathbf{\Delta}_2, \ldots, \mathbf{\Delta}_{g-1}\right) = \mathcal{S}_{SIR} \oplus \mathcal{S}_{DOC}. \tag{2.3}$$

Consequently, $\mathcal{S}_{SIR} \subset \mathcal{S}_{SAVE}$, and $\mathcal{S}_{DOC} \subset \mathcal{S}_{SAVE}$.

Proof: Because $\mathcal{S}(A) = \mathcal{S}(AA^T)$ for any matrix A and $\mathcal{S}(A) = \mathcal{S}(AB)$ for a nonsingular matrix B,

$$\begin{aligned}
\mathcal{S}_{SAVE} &= \mathcal{S}(\mathbf{\Omega}) = \mathcal{S}\left(E\left(\mathbf{I}_p - \mathbf{\Sigma}_y\right)^2\right) = \mathcal{S}\left(\sqrt{f_1}\left(\mathbf{I}_p - \mathbf{\Sigma}_1\right), \sqrt{f_2}\left(\mathbf{I}_p - \mathbf{\Sigma}_2\right), \ldots, \sqrt{f_g}\left(\mathbf{I}_p - \mathbf{\Sigma}_g\right)\right) \\
&= \mathcal{S}\left(\mathbf{I}_p - \mathbf{\Sigma}_1, \mathbf{I}_p - \mathbf{\Sigma}_2, \ldots, \mathbf{I}_p - \mathbf{\Sigma}_g\right) \\
&= \mathcal{S}\left(\left(\mathbf{I}_p - \mathbf{\Sigma}_1, \mathbf{I}_p - \mathbf{\Sigma}_2, \ldots, \mathbf{I}_p - \mathbf{\Sigma}_g\right)\begin{pmatrix} \mathbf{I}_p & \mathbf{I}_p & 0 & \ldots & 0 \\ 0 & -\mathbf{I}_p & \mathbf{I}_p & \ldots & 0 \\ 0 & 0 & -\mathbf{I}_p & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \mathbf{I}_p \\ 0 & 0 & 0 & \ldots & -\mathbf{I}_p \end{pmatrix}\right) \\
&= \mathcal{S}\left(\mathbf{I}_p - \mathbf{\Sigma}_1, \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1, \mathbf{\Sigma}_3 - \mathbf{\Sigma}_2, \ldots, \mathbf{\Sigma}_g - \mathbf{\Sigma}_{g-1}\right) \\
&= \mathcal{S}\left(\mathbf{I}_p - \mathbf{\Sigma}_1, \mathbf{\Delta}_1, \mathbf{\Delta}_2 \ldots, \mathbf{\Delta}_{g-1}\right).
\end{aligned}$$

Since $\mathbf{I}_p = \mathrm{Cov}(Z) = E(\mathrm{Cov}(Z\,|\,Y)) + \mathrm{Cov}(E(Z\,|\,Y)) = E(\mathbf{\Sigma}_y) + \mathrm{Cov}(\boldsymbol{\mu}_y)$,

$$\mathbf{I}_p - \mathbf{\Sigma}_1 = E(\mathbf{\Sigma}_y) + \mathrm{Cov}(\boldsymbol{\mu}_y) - \mathbf{\Sigma}_1 = \sum_{j=2}^{g} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1) + \sum_{j=1}^{g} f_j \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T.$$

Now we are to show that $\sum_{j=2}^{g} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1)$ is a linear combination of $\mathbf{\Delta}_1, \mathbf{\Delta}_2, \ldots, \mathbf{\Delta}_{g-1}$ using the mathematical induction. Consider the case of $g = 3$,

$$\begin{aligned}
\sum_{j=2}^{3} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1) &= f_2(\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1) + f_3(\mathbf{\Sigma}_3 - \mathbf{\Sigma}_1) = f_2(\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1) + f_3(\mathbf{\Sigma}_3 - \mathbf{\Sigma}_2 + (\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1)) \\
&= f_2\mathbf{\Delta}_1 + f_3\mathbf{\Delta}_2 + f_3\mathbf{\Delta}_1 = (f_2 + f_3)\mathbf{\Delta}_1 + f_3\mathbf{\Delta}_2 = \sum_{j=1}^{2} c_j \mathbf{\Delta}_j,
\end{aligned}$$

where c_j is the constant composed of f_i's.

This means that $\sum_{j=2}^{3} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1)$ is a linear combination of $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. Next, suppose that $\sum_{j=2}^{k} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1)$ is a linear combination of $\mathbf{\Delta}_1, \ldots, \mathbf{\Delta}_{k-1}$, that is,

$$\sum_{j=2}^{k} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1) = \sum_{j=1}^{k-1} c_j \mathbf{\Delta}_j.$$

Let's consider the case of $g = k + 1$.

$$\sum_{j=2}^{k+1} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1) = \sum_{j=2}^{k} f_j(\mathbf{\Sigma}_j - \mathbf{\Sigma}_1) + f_{k+1}(\mathbf{\Sigma}_{k+1} - \mathbf{\Sigma}_1)$$

$$
\begin{aligned}
&= \sum_{j=2}^{k} f_j(\boldsymbol{\Sigma}_j - \boldsymbol{\Sigma}_1) + f_{k+1}\left((\boldsymbol{\Sigma}_{k+1} - \boldsymbol{\Sigma}_k) + (\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{k-1}) + \cdots + (\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\right) \\
&= \sum_{j=1}^{k-1} c_j \boldsymbol{\Delta}_j + f_{k+1} \sum_{j=1}^{k} \boldsymbol{\Delta}_j = \sum_{j=1}^{k} c_j^* \boldsymbol{\Delta}_j,
\end{aligned}
$$

where c_j^* is the constant composed of f_i's.

As a result, this shows that $\sum_{j=2}^{g} f_j(\boldsymbol{\Sigma}_j - \boldsymbol{\Sigma}_1)$ is a linear combination of $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{g-1}$. Thus, by Proposition 1, $\mathcal{S}_{SAVE}$ reduces to

$$
\begin{aligned}
\mathcal{S}_{SAVE} = \mathcal{S}(\boldsymbol{\Omega}) &= \mathcal{S}\left(\sum_{j=1}^{g} f_j \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T, \boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{g-1}\right) \\
&= \mathcal{S}(\boldsymbol{\nu}, \boldsymbol{\Delta}) = \mathcal{S}(\boldsymbol{\nu}) \oplus \mathcal{S}(\boldsymbol{\Delta}) = \mathcal{S}_{SIR} \oplus \mathcal{S}_{DOC}
\end{aligned}
$$

because the property of direct sum $\oplus$ implies that $\mathcal{S}(A, B) = \mathcal{S}(A) \oplus \mathcal{S}(B)$. Hence the results follow. □

This proposition shows that SAVE is the most comprehensive procedure without requiring the linearity or constant covariance conditions. If the linearity and constant covariance conditions hold, but $Z \mid Y$ is not normally distributed, we will still have $\mathcal{S}_{SAVE} \subset \mathcal{S}_{y|x}$.

The following fact is that conditional normality of $Z \mid Y$ guarantees equality of the central and SAVE subspaces.

Proposition 4. *Suppose that $Z \mid Y$ follows a non-singular multivariate normal distribution: $Z \mid (Y = j) \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j),\ j = 1, \ldots, g$. Then $\mathcal{S}_{y|z} = \mathcal{S}_{SAVE}$.*

Proof: Let's consider only two values $(j,\ j+1)$ of the response with g values. Suppose that $Z \mid (Y = j)$ has a density $\mathbf{p}_j$,

$$
\log \frac{\Pr(Y = j+1 \mid Z)}{\Pr(Y = j \mid Z)} = \log \frac{\mathbf{p}_{j+1}(Z)}{p_j(Z)} + \text{constant}.
$$

Because $Z \mid (Y = j) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, for $j = 1, \ldots, g$,

$$
2 \log \frac{\mathbf{p}_{j+1}(Z)}{p_j(Z)} = \text{constant} + Z^T \left(\boldsymbol{\Sigma}_j^{-1} - \boldsymbol{\Sigma}_{j+1}^{-1}\right) Z + 2Z^T \left(\boldsymbol{\Sigma}_{j+1}^{-1} \boldsymbol{\mu}_{j+1} - \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j\right)
$$

by Seber (1984, p.283). The result of Cook and Lee (1999) reduces to

$$
\mathcal{S}\left(\boldsymbol{\Sigma}_j^{-1} - \boldsymbol{\Sigma}_{j+1}^{-1}, \boldsymbol{\Sigma}_{j+1}^{-1} \boldsymbol{\mu}_{j+1} - \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j\right) = \mathcal{S}\left(\boldsymbol{\Delta}_j, \boldsymbol{\nu}_j\right), \quad j = 1, \ldots, g-1.
$$

It follows immediately from this characterizing expression that

$$
\begin{aligned}
\mathcal{S}_{y|z} &= \mathcal{S}(\boldsymbol{\Delta}_1, \boldsymbol{\nu}_1) \oplus \mathcal{S}(\boldsymbol{\Delta}_2, \boldsymbol{\nu}_2) \oplus \cdots \oplus \mathcal{S}\left(\boldsymbol{\Delta}_{g-1}, \boldsymbol{\nu}_{g-1}\right) \\
&= \mathcal{S}\left(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{g-1}, \boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{g-1}\right) = \mathcal{S}(\boldsymbol{\nu}, \boldsymbol{\Delta}) = \mathcal{S}_{SAVE}.
\end{aligned}
$$

Hence the results follow. □

3. Discussion

In this paper, we extend and generalize the part in the result by Cook and Lee (1999) to the case where the response has more than two values. In practice, the conditional normal distribution of Z given Y guarantees that the subspace constructed by the method SAVE coincides with the central subspace. Li and Zhu (2007) investigated the asymptotic distribution for SAVE as the general version. For the practical use, the asymptotic distribution of test statistic for SAVE to determine the structural dimensionality is under investigation.

References

Cook, R. D. (1994). On the interpretation of regression plots, *Journal of the American Statistical Association*, **89**, 177–189.

Cook, R. D. (1998). *Regression Graphics*, John Wiley & Sons, New York.

Cook, R. D. and Lee, H. (1999). Dimension reduction in binary response regression, *Journal of the American Statistical Association*, **94**, 1187–1200.

Cook, R. D. and Weisberg, S. (1991). Discussion of sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–342.

Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma, *Journal of the American Statistical Association*, **87**, 1025–1039.

Li, Y. and Zhu, L. (2007). Asymptotics for sliced average variance estimation, *The Annals of Statistics*, **35**, 41–69.

Seber, G. (1984). *Multivariate Observations*, John Wiley & Sons, New York.

Received April 22, 2012; Revised April 30, 2012; Accepted May 8, 2012