Communications for Statistical Applications and Methods
2016, Vol. 23, No. 3, 259–268

http://dx.doi.org/10.5351/CSAM.2016.23.3.259
Print ISSN 2287-7843 / Online ISSN 2383-4757

Dimension reduction for right-censored survival regression: transformation approach

Jae Keun Yoo[1,a], Sung-Jin Kim[a], Bi-Seul Seo[a], Hyejung Shin[a], Su-Ah Sim[a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

High-dimensional survival data with large numbers of predictors has become more common. The analysis of such data can be facilitated if the dimensions of predictors are adequately reduced. Recent studies show that a method called sliced inverse regression (SIR) is an effective dimension reduction tool in high-dimensional survival regression. However, it faces incapability in implementation due to a double categorization procedure. This problem can be overcome in the right-censoring type by transforming the observed survival time and censoring status into a single variable. This provides more flexibility in the categorization, so the applicability of SIR can be enhanced. Numerical studies show that the proposed transforming approach is equally good to (or even better) than the usual SIR application in both balanced and highly-unbalanced censoring status. The real data example also confirms its practical usefulness, so the proposed approach should be an effective and valuable addition to usual statistical practitioners.

Keywords: bivariate slicing, right-censored data, sliced inverse regression, sufficient dimension reduction, survival regression, transformation method, unbalanced censoring status

1. Introduction

Survival regression is a study of the conditional distribution of a true survival time T given a set of predictors, saying $\mathbf{X} \in \mathbb{R}^p = (X_1, \ldots, X_p)^{\mathrm{T}}$. A direct regression analysis of $T|\mathbf{X}$ is not possible since the survival time T cannot be fully observed due to the censoring of T. A most popular one among many types of censoring should be right-censoring which takes $\min(T, C)$ as observations of T, where a variable C indicates a censoring variable. In the right censoring scheme, the observed data of $(Y_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, are assumed as n iid realizations of $(T, C, \mathbf{X})$, where $Y = T\delta + C(1 - \delta)$, $\delta = 0, 1$ is an indicator variable with $\delta_{(C \geq T)} = 1$. That is, $\delta_i = 0$ means that a censoring occurs to the i^{th} subject, so the observed survival time is a possible value of C, not T. Hereafter, Y and δ are called observed survival time and censoring status, respectively. These data are typically used for survival regression.

In survival regression of $T|\mathbf{X}$, two popular statistical approaches among many should be the Cox proportional Hazards (CPH) model and the accelerated failure time (AFT) model. The CPH model has a long history and is semi-parametric due to no requirement of specific distributions of T. In the CPH model, the unknown regression coefficients are estimated based on the likelihood partially constructed by the ordered non-censored observed survival time. After fitting the model, the survival time is indirectly interpreted through hazards ratios. The AFT model assumes specific distributions for

[1] Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: peter.yoo@ewha.ac.kr

Published 31 May 2016 / journal homepage: http://csam.or.kr

T, so it construct the full-likelihood from the data to estimate unknown regression coefficients. One advantageous aspect of the AFT model should be the direct interpretation of survival time. However, the optimization of the likelihood in the AFT model is often computationally intensive. For more about the CPH and AFT models, read (Kleinbaum and Klein, 2005).

High-dimensional survival data with large numbers of predictors has become more common. In such cases, the two popular approaches may suffer from the curse of dimensionality. Often, high-dimensional regression analysis can be facilitated, if the dimensions of predictors are adequately reduced. Here, we consider a sufficient dimension reduction methodology called sliced inverse regression (SIR; Li, 1991) to reduce the dimension of predictors. SIR replaces the original p-dimensional predictor with d-dimensional linearly transformed predictors $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ without loss of information on $T|\mathbf{X}$, where $\boldsymbol{\eta}$ is a $p \times d$ matrix. Equivalently, SIR pursues to find $\boldsymbol{\eta}$ such that

$$F_{T|\mathbf{X}}(\cdot) = F_{T|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}}(\cdot), \tag{1.1}$$

where $F(\cdot)$ stands for a distribution function.

There are several advantages of SIR as a dimension reduction tool in survival regression. First, SIR does not require certain parametric distribution for $T|\mathbf{X}$, so it can be applicable to both CPH and AFT models. Second, unlike many local nonparametric methodologies, SIR can often avoid the curse of dimensionality because its estimate converges at the usual $\sqrt{n}$ rate. Third, SIR is easily implemented in practice with dr-package in R. In the later section, SIR and its applicability to survival regression will be discussed in further detail. For more about sufficient dimension reduction and their methodologies (including SIR), read Yoo (2016a, 2016b).

A categorization of a response variable (slicing) should be done to implement SIR in practice. Using SIR in survival regression, slicing Y and δ is the key step, so that the observed survival time Y is categorized within each level of the censoring status δ. Bivariate slicing should be problematic in practice when δ is heavily unbalanced. Few (or no observations) in at least one slice result in no implementation of SIR. This problem can be relaxed by transforming Y and δ into one-dimensional variable when the survival time is right-censored. The transformed variable can provide more flexibility in slicing, so SIR can be implemented despite a heavy imbalance in censoring. We consider two approaches for the transformation as suggested in Datta *et al.* (2007). Hereafter, the categorization of the transformed observed survival time will be called *transformed slicing*.

This paper conducts a comparison study of SIR applications with original bivariate slicing and transformed slicing. The applicability of SIR to survival regression can be enhanced if the latter will give equally good or better results than the former. This enables a usual statistical practitioner to conduct a proper dimension reduction in a high-dimensional survival regression with heavily imbalanced censoring status; therefore, the analysis can be facilitated with dimension reduced predictors.

The organization of the paper is as follows. In Section 2, SIR in survival regression and two transformation methods are discussed. Section 3 is devoted to presenting numerical studies for the performances of SIR for CPH and AFT models with two different censoring rates and real data analysis. In Section 4, we summarize our work.

2. Sliced inverse regression and transformation methods

2.1. Sliced inverse regression in survival regression

Consider a survival regression of $T|\mathbf{X}$ with right-censoring. For notational convenience, we define $\mathcal{S}(\mathbf{M})$ and $\perp\!\!\!\perp$ as a subspace spanned by the columns of $\mathbf{M} \in \mathbb{R}^{p\times q}$ and statistical independence, respectively.

SIR (Li, 1991) constructs a subspace $\mathcal{S}\{E(\mathbf{X}|T)\}$, which is a subspace spanned by the inverse mean $E(\mathbf{X}|T)$ with varying T. SIR estimates $\boldsymbol{\eta}$ to satisfy statement (1.1) through $E(\mathbf{X}|T)$. The estimation of $\boldsymbol{\eta}$ through SIR is:

(a) Divide the observed range of T into h slices J_j, if T is many-valued or continuous. If T is categorical, each category is a slice.

(b) Compute the sample means within each slice, $\bar{\mathbf{X}}_j = (1/n_j)\sum_{T_i \in J_j} \mathbf{X}_i$, $j = 1, \ldots, h$, where n_j is the number of observations within J_j.

(c) $\hat{\mathbf{M}} = \sum_{j=1}^{h} f_j(\bar{\mathbf{X}}_j - \bar{\mathbf{X}})(\bar{\mathbf{X}}_j - \bar{\mathbf{X}})^{\mathrm{T}}$, where $f_j = n_j/n$.

(d) Then $\hat{\mathbf{M}}$ is spectral-decomposed such that $\hat{\mathbf{M}} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\mathbf{m}}_i \hat{\mathbf{m}}_i^{\mathrm{T}}$ with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \cdots \hat{\lambda}_p \geq 0$.

(e) The matrix $\boldsymbol{\eta}$ is estimated by $\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\mathbf{m}}_1, \ldots, \hat{\mathbf{m}}_d)$, where $\hat{\boldsymbol{\Sigma}}$ is usual moment estimator of $\mathrm{cov}(\mathbf{X})$.

Li *et al.* (1999) and Cook (2003) investigate the applicability of SIR to survival regression with $(Y, C, \mathbf{X})$. Li *et al.* (1999) discuss that the direct application of SIR based on slices of Y potentially introduces bias in dimension reduction due to censoring. According Li and his associates, if a condition of $C \perp\!\!\!\perp (\mathbf{X}, T)$ holds, the bias is eliminated, but the condition is too restricted in practice. In order for SIR to work properly under a more general censoring condition of $C \perp\!\!\!\perp T|\mathbf{X}$, a bivariate slicing of Y and δ to categorize Y within each level of δ is suggested.

According to Cook (2003), the bivariate slicing can work in survival regression under a more general condition $C \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, T)$. This condition enables the use of a plausible regression model to estimate $\boldsymbol{\eta}$. Let $\boldsymbol{\gamma}$ be matrices to satisfy $F_{(T,C)|\mathbf{X}}(\cdot) = F_{(T,C)|\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}(\cdot)$. Then the condition forces that $F_{(T,C)|\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}(\cdot) = F_{(T,C)|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}}(\cdot)$. Since (Y, δ) is a function of (T, C), it follows immediately that

$$F_{Y,\delta|\mathbf{X}}(\cdot) = F_{Y,\delta|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}}(\cdot). \tag{2.1}$$

SIR-application for survival regression can be done with bivariate responses of $(Y, \delta)|\mathbf{X}$. This approach requires bivariate slicing of (Y, δ) discussed above.

2.2. Transformation methods

In this subsection, we discuss two transformation methods for right-censored survival data as suggested in Datta *et al.* (2007).

Method 1 (reweighting): The method is based on inverse probability of censoring weighted estimation (Robins and Finkelstein, 2000; Robins and Rotnitzky, 1992; Satten and Datta, 2001; Satten *et al.*, 2001).

Let $S^c(t)$ denote the survival function of the censoring variable C. As long as a condition of $C \perp\!\!\!\perp (T, \mathbf{X})$ holds, it can be estimated by the Kaplan-Meier estimator

$$\hat{S}^c(t) = \prod_{\tau_{(i)} \leq t} \left\{ 1 - \frac{\Delta N^c(\tau_{(i)})}{R(\tau_{(i)})} \right\}, \tag{2.2}$$

where $\tau_{(1)} < \cdots < \tau_{(m)}$ are the distinct ordered censored lifetimes, $\Delta N^c(\tau_{(i)})$ is the number of censored observations at time $\tau_{(i)}$, T_i^c stands for the observed censored time and $R(\tau_{(i)}) = \#\{j : T_j^c \geq \tau_{(i)}\}$ counts the number of individuals at risk of failing just before time $\tau_{(i)}$.

Under this scenario, the unobserved Y_i is replaced by 0. To compensate for this, the observed Y_i is reweighed by the reciprocal of the probability that it corresponds to an uncensored observation. Mathematically, we have $\tilde{Y}_i = \delta_i \log(T_i^c)/\hat{S}^c(T_i^c-)$, where '–' denotes a left limit. It should be noted that $\tilde{Y}$ is constructed from observed data, because According to Koul *et al.* (1981), $E(\tilde{Y}_i|\mathbf{X})$ is approximately equal to $E(Y_i|\mathbf{X})$. Method 1 replaces pairs of (Y_i, δ_i) by $\tilde{Y}_i$.

Method 2 (mean imputation): According to Datta (2005), the usual moment-based sample mean is an inconsistent and asymptotically biased estimator due to right censoring. To relax this problem, Datta (2005) suggests a computation of the sample mean in the usual way by imputing the censored values. In method 2, the censored survival times are imputed by following the guidance in Datta (2005), which is discussed below in detail. So, the observed Y_i is kept intact in method 2, while unobserved Y_i values are replaced by its expected value given that the observed failure time T_i was larger than T_i^c. The Kaplan-Meier curve of the survival function of T can provide its estimate as:

$$Y_i^* = \hat{S}^c(T_i^c)^{-1} \sum_{\tau_{(j)} > T_i^c} \log \tau_{(j)} \Delta \hat{S}_{\tau_{(j)}},$$

where $\hat{S}$ stands for the Kaplan-Meier estimator with the roles of δ and $1-\delta$ exchanged in (2.2) and $\Delta\hat{S}_{\tau_{(j)}}$ is the jump size of $\hat{S}$ at time $\tau_{(j)}$.

In this computation, the largest event time τ_m is chosen as a true failure, although $\delta_m = 0$. This allows $\tau_{(m)}$ to be the largest mass point of the estimated survival curve. In this scheme, we replace each censored observation by its estimated conditional expectation given that the true failure was a value that exceeded the censored observation. Thus, in method 2, pairs of (Y_i, δ_i) are replaced by $\tilde{Y}_i$ such that $\tilde{Y}_i = Y_i$, if $\delta_i = 1$, and $\tilde{Y}_i = Y_i^*$, if $\delta_i = 0$.

3. Numerical studies and real data application

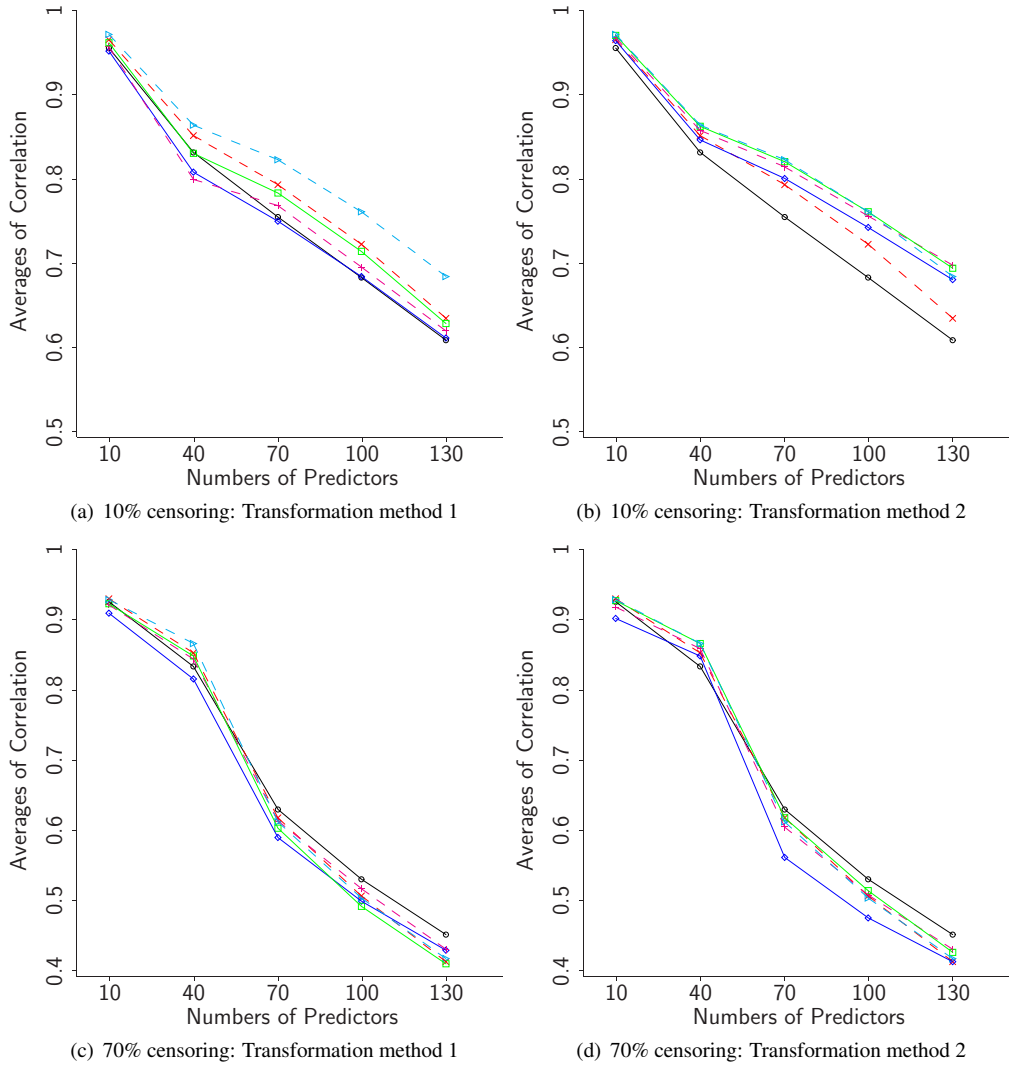
3.1. Numerical studies

We consider two survival regressions of CPH and AFT models. For both models, the predictors of $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ were independently generated from $N(0,1)$, $p = 10, 40, 70, 100, 130$, and one linear combination of $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ was considered, where the first $p/2$ coefficients of $\boldsymbol{\eta}$ are $(p/10)$ replicates of $\boldsymbol{\eta}_0 = (1, 2, 3, 4, 5)$ and its last $(p/2)$ coefficients are all zeros. For example, for $p = 40$, we have $\boldsymbol{\eta} = (\boldsymbol{\eta}_0, \boldsymbol{\eta}_0, \boldsymbol{\eta}_0, \boldsymbol{\eta}_0, 0, 0, \ldots, 0)^{\mathrm{T}}$.

Under this predictor configurations, an AFT model was generated following the model in Section 5 of Datta *et al.* (2007) such that $T|\mathbf{X} \overset{iid}{\sim} \exp(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X} + \varepsilon)$, where a random error ε was sampled from $N(0,1)$ independently of $\mathbf{X}$. That is, under this setup, $T|\mathbf{X}$ follows log-normal distribution. A censoring variable C was sampled from log-normal distribution $\exp\{N(c_0\sqrt{2}, 2)\}$.

A CPH model was simulated by mimicking one in Section 4.2 of Yoo and Lee (2011). The model was generated with a hazard rate $\lambda = \exp(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})$ and a baseline hazard rate λ_0 equal to 1. A censoring time C was sampled from Uniform$(0, c_0)$ independently of $\mathbf{X}$.

For the AFT and CPH models, c_0 was chosen to have 10% and 70% average censoring percentages. In the AFT model, $c_0 = 1.81$ and $c_0 = -0.71$ were selected for 10% and 70% censoring, respectively, while $c_0 = 14.8$ and $c_0 = 0.61$ were used for 10% and 70% censoring, respectively in the CPH model. The total number of iterations were 500 and the sample size n was fixed at 200. The sample size $n = 200$ was used to provide adequate observation for SIR implementation with bivariate slicing with 10% censoring.



(a) 10% censoring: Transformation method 1

(b) 10% censoring: Transformation method 2

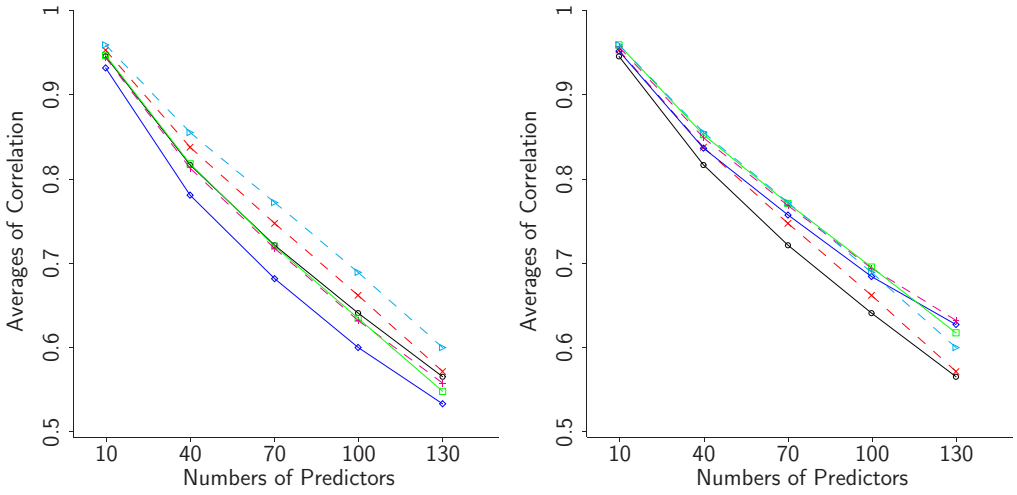
(c) 70% censoring: Transformation method 1

(d) 70% censoring: Transformation method 2

Figure 1: *Direction estimation for accelerated failure time (AFT) model in Section 3.1 (black: $\mathcal{B}_4$; red: $\mathcal{B}_6$; blue: $\mathcal{T}_3^{(\bullet)}$; magenta: $\mathcal{T}_4^{(\bullet)}$; green: $\mathcal{T}_5^{(\bullet)}$; cyan: $\mathcal{T}_6^{(\bullet)}$).*

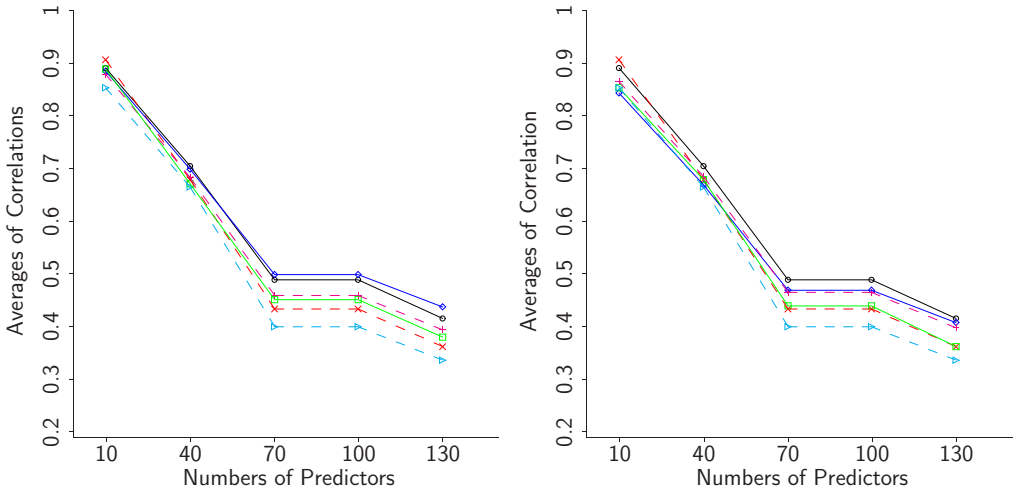
Two popular types of the survival regression of the CPH and AFT models were considered with $p = 10, 40, 70, 100, 130$ under two different censoring percentages of 10% and 70%. It is expected that the transformation methods would produce better estimate results in 10% (which is highly unbalanced), than the usual bivariate slicing; however, the latter should be better (or equally good) to the transformation methods.

To summarize the numerical studies, the averages of absolute correlation coefficients between $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ and $\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{X}$ were computed, which are reported in Figures 1 and 2. The estimate $\hat{\boldsymbol{\eta}}$ were obtained from bivariate slicing with 4 and 6 slices, two transforming slicing schemes with 3, 4, 5 and 6 slices. If the averages are closed to one, it estimates $\boldsymbol{\eta}$ well. The true dimension of $\mathcal{S}(\boldsymbol{\eta})$ is equal to one; therefore,



(a) 10% censoring: Transformation method 1

(b) 10% censoring: Transformation method 2



(c) 70% censoring: Transformation method 1

(d) 70% censoring: Transformation method 2

Figure 2: *Direction estimation for Cox proportional hazards (CPH) model in Section 3.1 (black: $\mathcal{B}_4$; red: $\mathcal{B}_6$; blue: $\mathcal{T}_3^{(\bullet)}$; magenta: $\mathcal{T}_4^{(\bullet)}$; green: $\mathcal{T}_5^{(\bullet)}$; cyan: $\mathcal{T}_6^{(\bullet)}$).*

the percentages of dimension determination of $\hat{d} = 1$ with level 5% were computed and reported in Tables 1–4. The percentages close to 95% indicates good estimation of the dimension.

In Figures 1, 2 and Tables 1–4, the notation of $\mathcal{B}_K$, $K = 4, 6$, "$\mathcal{B}$" and K represent SIR application of bivariate slicing of the censoring status and observed survival time and the number of slices, respectively. A notation of $\mathcal{T}_J^{(I)}$, $I = 1, 2$ and $J = 3, 4, 5, 6$ also stands for SIR application with transforming slicing, where "I" and "J" represent a type of the methods and the numbers of slices, respectively. For example, $\mathcal{B}_4$ means the bivariate slicing SIR application with 4 slices, and $\mathcal{T}_3^{(2)}$ does the transforming SIR application via the method 2 with 3 slices.

According to Figures 1 and 2, with both 10% and 70% censoring, the performances to estimate

Table 1: Percentages of $\hat{d} = 1$ for AFT model with 10% censoring in Section 3.1

p	$\mathcal{B}_4$	$\mathcal{B}_6$	$\mathcal{T}_3^{(1)}$	$\mathcal{T}_4^{(1)}$	$\mathcal{T}_5^{(1)}$	$\mathcal{T}_6^{(1)}$	$\mathcal{T}_3^{(2)}$	$\mathcal{T}_4^{(2)}$	$\mathcal{T}_5^{(2)}$	$\mathcal{T}_6^{(2)}$
10	95.8	93.6	93.0	93.4	93.2	91.2	93.2	92.0	93.0	93.2
40	70.4	54.4	69.0	65.2	55.2	47.2	65.6	61.8	61.4	55.2
70	36.4	29.0	35.8	32.8	31.6	25.4	36.0	32.4	32.8	28.6
100	12.0	8.4	13.0	10.2	7.2	6.0	9.6	12.4	11.6	7.8
130	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2

Table 2: Percentages of $\hat{d} = 1$ for AFT model with 70% censoring in Section 3.1

p	$\mathcal{B}_4$	$\mathcal{B}_6$	$\mathcal{T}_3^{(1)}$	$\mathcal{T}_4^{(1)}$	$\mathcal{T}_5^{(1)}$	$\mathcal{T}_6^{(1)}$	$\mathcal{T}_3^{(2)}$	$\mathcal{T}_4^{(2)}$	$\mathcal{T}_5^{(2)}$	$\mathcal{T}_6^{(2)}$
10	95.4	96.0	96.6	95.2	94.4	95.6	92.6	96.0	95.2	96.4
40	93.0	93.0	95.4	95.0	92.2	87.4	93.2	95.8	96.8	94.0
70	68.8	61.8	81.2	74.6	67.6	62.2	93.4	87.2	78.8	73.4
100	34.2	31.4	43.2	37.4	33.2	26.6	64.4	54.2	40.2	38.2
130	10.2	8.2	9.6	8.6	7.6	6.8	19.2	14.2	11.0	9.2

Table 3: Percentages of $\hat{d} = 1$ for CPH model with 10% censoring in Section 3.1

p	$\mathcal{B}_4$	$\mathcal{B}_6$	$\mathcal{T}_3^{(1)}$	$\mathcal{T}_4^{(1)}$	$\mathcal{T}_5^{(1)}$	$\mathcal{T}_6^{(1)}$	$\mathcal{T}_3^{(2)}$	$\mathcal{T}_4^{(2)}$	$\mathcal{T}_5^{(2)}$	$\mathcal{T}_6^{(2)}$
10	96.0	94.4	96.4	95.6	95.0	95.6	96.6	94.8	94.4	96.4
40	96.4	97.2	94.6	95.4	95.2	96.2	93.0	95.8	97.8	97.0
70	98.4	98.0	96.2	98.0	97.8	97.6	91.2	98.8	98.0	97.2
100	98.8	89.2	94.2	94.6	96.8	97.0	88.4	98.6	95.6	94.4
130	70.8	44.2	55.8	62.6	58.2	58.4	69.2	79.2	62.4	55.6

Table 4: Percentages of $\hat{d} = 1$ for CPH model with 70% censoring in Section 3.1

p	$\mathcal{B}_4$	$\mathcal{B}_6$	$\mathcal{T}_3^{(1)}$	$\mathcal{T}_4^{(1)}$	$\mathcal{T}_5^{(1)}$	$\mathcal{T}_6^{(1)}$	$\mathcal{T}_3^{(2)}$	$\mathcal{T}_4^{(2)}$	$\mathcal{T}_5^{(2)}$	$\mathcal{T}_6^{(2)}$
10	95.6	96.2	95.2	95.6	96.2	93.8	93.4	94.0	97.2	95.8
40	97.4	97.2	96.2	97.0	96.4	97.0	86.0	97.4	96.8	97.8
70	96.6	97.0	96.6	97.2	97.4	97.4	81.4	97.6	96.4	96.4
100	98.6	88.2	99.0	97.4	94.6	93.6	78.0	97.8	97.2	93.2
130	68.4	39.2	91.8	72.2	59.8	52.0	53.4	76.2	58.2	52.8

$\boldsymbol{\eta}$ via the two transforming slicing schemes are similar in both the AFT and CPH models; however, the method 2 produced slightly better results than method 1 in most cases. For the CPH and AFT models with 10% censoring, $\mathcal{T}_6^{(\bullet)}$ yields the best results among the others, and $\mathcal{T}_4^{(\bullet)}$ shows good performance. With 70% censoring, the two transformation slicing schemes, especially $\mathcal{T}_3^{(\bullet)}$ and $\mathcal{T}_4^{(\bullet)}$, are equally good to bivariate slicing; however, the latter is slightly better than the former. This meets our expectation, so the transformation slicing is shown to be good alternatives to bivariate slicing in dimension reduction in survival regression.

For the dimension estimation, Tables 1–4 show that the transformation slicing produces at least equally good or even better results than bivariate slicing. In the AFT model, $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_3^{(2)}$ result in reliable dimension estimation. For the CPH model, under 10% censoring, $\mathcal{T}_4^{(1)}$ and $\mathcal{T}_4^{(2)}$ are better than the others, while $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_4^{(2)}$ are good.

3.2. Real data example: primary biliary cirrhosis data

For illustration purposes, the data commonly used in Tibshirani (1997) and Yoo and Lee (2011) was considered. The data was on primary biliary cirrhosis (PBC), which were collected at the Mayo Clinic

Table 5: p-values for the dimension estimation of primary biliary cirrhosis data in Section 3.2

H_0	$\mathcal{B}_4$	$\mathcal{B}_6$	$\mathcal{T}_3^{(1)}$	$\mathcal{T}_4^{(1)}$	$\mathcal{T}_5^{(1)}$	$\mathcal{T}_6^{(1)}$	$\mathcal{T}^{(}2)_3$	$\mathcal{T}^{(}2)_4$	$\mathcal{T}^{(}2)_5$	$\mathcal{T}^{(}2)_6$
$d = 0$	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
$d = 1$	0.00	0.00	0.73	0.82	0.91	0.91	0.07	0.09	0.01	0.00
$d = 2$	0.01	0.00	N/A	0.85	0.99	0.99	N/A	0.45	0.27	0.08
$d = 3$	N/A	0.26	N/A	N/A	0.97	0.99	N/A	N/A	0.97	0.89
Decision	$\hat{d} > 2$	$\hat{d} = 3$	$\hat{d} = 1$	$\hat{d} = 1$	$\hat{d} = 1$	$\hat{d} = 1$	$\hat{d} = 1$	$\hat{d} = 1$	$\hat{d} = 2$	$\hat{d} = 2$

between 1974 and 1986. The data consists of 19 variables with 276 observations after removing all missing values. The variables used in the analysis are: Y = the number of days between registration and the earlier of death or censoring; $\delta = 1$, if Y is time to death; 0 otherwise; X_1 Treatment code: 1 = D-penicillamine, 2 = placebo; X_2 Age in years; X_3 Gender: 0 = male, 1 = female; X_4 Presence of ascites: absent = 0 or present = 1; X_5 Presence of hepatomegaly: absent = 0 or present = 1; X_6 Presence of spiders: 0 = no or 1 = yes; X_7 Presence of edema: absent and no diuretic therapy = 0, present but no diuretic therapy or edema resolved by diuretics = 0.5 or present despite diuretic therapy = 1; X_8 Serum bilirubin, in mg/dL; X_9 Serum cholesterol, in mg/dL; X_{10} Albumin, in g/dL; X_{11} Urine copper, in μg/day; X_{12} Alkaline phosphatase, in U/L; X_{13} SGOT, in U/mL; X_{14} Triglycerides, in mg/dL; X_{15} Platelet count; coded value is number of platelets per cubic mL of blood divided by 1,000; X_{16} Prothrombin time, in seconds; X_{17} Histologic state of disease, graded 1, 2, 3, or 4.

Yoo and Lee (2011) did the application of SIR and ordinary least squares to the PBC data for model-free predictor test. In Tibshirani (1997), the data was successfully fitted with the CPH model with 17 predictors. We considered SIR application with the bivariate slicing with 4 and 6 slices and the two transformation slicing schemes with 3, 4, 5, 6 slices.

First, Table 5 shows the p-values for the dimension estimation obtained from SIR implementation with dr-package in R. According to the table, $\mathcal{B}_4$ and $\mathcal{B}_6$ estimate the dimension as three or possibly larger. However, all $\mathcal{T}_\bullet^{(1)}$, $\mathcal{T}_3^{(2)}$ and $\mathcal{T}_4^{(2)}$ determine $\hat{d} = 1$, while $\mathcal{T}_5^{(2)}$ and $\mathcal{T}_6^{(2)}$ do $\hat{d} = 2$. According to numerical studies, $\mathcal{T}_3^{(1)}$, $\mathcal{T}_4^{(1)}$, $\mathcal{T}_3^{(2)}$, and $\mathcal{T}_4^{(2)}$ showed better dimension estimations. Following this guidance, it is concluded that $\hat{d} = 1$. This indicates that the usual bivariate slicing overestimates the true dimension. Therefore, the consideration of the transformation slicing relaxes the potential complexity in dimension reduction. This proves the usefulness of the proposed transforming slicing in practice.

Next, the scatter plot matrix of all the estimates $\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{X}$ from SIR applications are presented in Figure 3. According to the figure, $\mathcal{B}_\bullet$s and $\mathcal{T}_\bullet^{(2)}$s yield the highly-correlated estimates, while $\mathcal{T}_\bullet^{(1)}$s provide a different one. Combining this with the dimension estimation results, either $\mathcal{T}_3^{(2)}$ or $\mathcal{T}_4^{(2)}$ should be preferred over others. Their correlations with the CPH fit are 0.96 and 0.98, respectively, and confirms that the dimension reduction using the transformation methods would be successful in practice.

4. Conclusion

High-dimensional survival data with large numbers of predictors is more common and the direct application of the two popular statistical approaches of AFT model and CPH model may face and suffer from the curse of dimensionality. The analysis of such data can be facilitated if the dimensions of predictors are adequately reduced. Recent studies show SIR (Li, 1991). SIR requires the categorization (slicing) of the observed survival time within each level of the censoring status. For example, some categories may have inadequate observations to implement SIR in a case of highly-unbalanced censoring. This problem can be overcome in the right-censoring type by transforming the observed

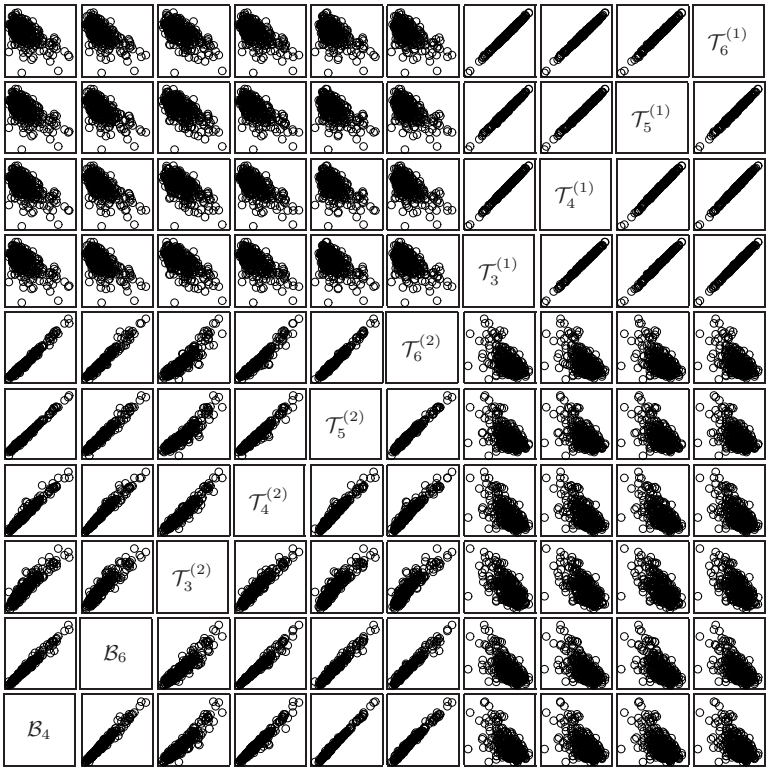


Figure 3: *Scatterplot matrix of $\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{X}$ from various sliced inverse regression application for the primary biliary cirrhosis data in Section 3.2.*

survival time and censoring status into a single variable. The applicability of SIR can be enhanced because it provides more flexibility in the categorization. For the transformation method, we adopted two approaches suggested in Datta *et al.* (2007).

Numerical studies indicate that the two transformation slicing schemes are equally good to (or even better) than usual bivariate slicing in dimension reduction in both balanced and highly-unbalanced censoring status. The real data example also confirms its practical usefulness; therefore, the proposed approach should be an effective and valuable addition to usual statistical practitioners.

The proposed approach is restricted in right-censoring type survival data; however, SIR application of the bivariate slicing does not have this restriction and remains a possible shortcoming of the proposed approach. The research for a transformation approach with interval and left censoring types is in progress. SIR can be implemented with dr-package in R, and the two transformation methods will be available on the webpage of the author, http://home.ewha.ac.kr/~yjkstat/transformat ion.txt.

Acknowledgements

For the corresponding author Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2014R1A2A1A11049389 and 2009-0093827). For Sung-Jin Kim, Bi-Seul Seo, and Su-Ah Sim, this work was supported by the BK21 Plus Project through the National Research

Foundation of Korea (NRF) funded by the Korean Ministry of Education (22A20130011003).

References

Cook RD (2003). Dimension reduction and graphical exploration in regression including survival analysis, *Statistics in Medicine*, **22**, 1399–1413.

Datta S (2005). Estimating the mean life time using right censored data, *Statistical Methodology*, **2**, 65–69.

Datta S, Le-Rademacher J, and Datta S (2007). Predicting patient survival from microarray Data by accelerated failure time modeling using partial least squares and LASSO, *Biometrics*, **63**, 259–271.

Kleinbaum DG and Klein M (2005). *Survival Analysis: A Self-Learning Text* (2nd ed), Springer, New York.

Koul H, Susarla V, and Van Ryzin J (1981). Regression analysis with randomly right-censored data, *Annals of Statistics*, **9**, 1276–1288.

Li KC (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–327.

Li KC, Wang JL, and Chen CH (1999). Dimension reduction for censored regression data, *Annals of Statistics*, **27**, 1–23.

Robins JM and Finkelstein DM (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests, *Biometrics*, **56**, 779–788.

Robins JM and Rotnitzky A (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In NP Jewell, K Dietz, and VT Farewell (Eds), *AIDS Epidemiology: Methodological Issues* (pp. 297–331), Birkhäuser, Boston.

Satten GA and Datta S (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average, *American Statistician*, **55**, 207–210.

Satten GA, Datta S, and Robins J (2001). Estimating the marginal survival function in the presence of time dependent covariates, *Statistics & Probability Letters*, **54**, 397–403.

Tibshirani R (1997). The lasso method for variable selection in the Cox model, *Statistics in Medicine*, **16**, 385–395.

Yoo JK (2016a). Tutorial: Dimension reduction in regression with a notion of sufficiency, *Communications for Statistical Applications and Methods*, **23**, 93–103.

Yoo JK (2016b). Tutorial: Methodologies for sufficient dimension reduction in regression, *Communications for Statistical Applications and Methods*, **23**, 105–117.

Yoo JK and Lee K (2011). Model-free predictor tests in survival regression through sufficient dimension reduction, *Lifetime Data Analysis*, **17**, 433–444.

Received April 30, 2016; Revised May 19, 2016; Accepted May 21, 2016