

# Sliced Inverse Regression for Survival Data

Maya Shevlyakova · Stephan  
Morgenthaler

Received: date / Accepted: date

**Abstract** We apply the univariate sliced inverse regression (SIR) to survival data. Our approach is different from the other papers on this subject. The right-censored observations are taken into account during the slicing of the survival times by assigning each of them with equal weight to all of the slices with longer survival.

We test this method with different distributions for the two main survival data models, the accelerated lifetime model and Cox's proportional hazards model. In both cases and under different conditions of sparsity, sample size and dimension of parameters, this non-parametric approach finds the data structure and can be viewed as a variable selector.

**Keywords** survival data · sliced inverse regression

## 1 Introduction

Dimension reduction aims to select a few new variables, which often are linear combinations of the original ones, and which describe the important aspects of the observed data well enough. Sufficient dimension reduction requires no model assumption, yet it provides a tool for subsequent model formulation and prediction. The main idea is based on the regression of a vector  $Y$  on a  $p$ -dimensional predictor  $x$  by estimating the central subspace  $S$  (of dimension  $k < p$ ) which contains most of the information about our response  $Y$ . This

---

F. Author Maya Shevlyakova  
Ecole Polytechnique Fédérale de Lausanne, Switzerland  
Tel.: +41-21-6932784  
Fax: +41-21-6934250  
E-mail: maya.shevlyakova@epfl.ch

S. Author Stephan Morgenthaler  
Ecole Polytechnique Fédérale de Lausanne, Switzerland  
E-mail: stephan.morgenthaler@epfl.ch

can be written as  $Y \perp x P_s x$ , where  $\perp$  designates independence and  $P_s$  denotes an orthogonal projection onto  $S$ . Under minor conditions ([2],[5]), such a subspace, denoted as  $S_{Y|x}$ , exists and is unique.

Sliced inverse regression (SIR) is one of the methods of dimension reduction, aiming to estimate  $S_{Y|x}$  without any parametric assumptions on the conditional distribution of  $Y|x$ . It was suggested in 1991 by Li [5] and is based on the spectral decomposition of  $\text{Cov}(E(X|Y))$ . Such an estimate of  $\text{Cov}(E(X|Y))$  is formed by dividing the response  $Y$  into non-overlapping slices and averaging the covariates  $x$  within each of the slice. A version for survival data appeared in [6], where the censoring was handled by a double slicing technique, treating the observed and censored responses separately. Later on, an idea of inverse regression was generalized by Cook [1], introducing a minimum discrepancy approach in order to find the central subspace  $S_{Y|x}$ . Latest results on bivariate estimation can be found in [13].

An alternative to bivariate dimension reduction method is reweighting the censored observations. In literature it is mainly done using the method of inverse probability weighting estimation. Details on this approach can be found, for example, in [12]. SIR in survival analysis has been addressed in [9] and [11]. Some papers consider regularized estimation, for instance, [10].

SIR has been adopted for high-dimensional data, usually as the second step of a two-way procedure. For the microarray data, [8] suggests the principal component analysis for dimension reduction and further application of the found components as the input data for the SIR. More specific problems of linear dimension reduction methods under high-dimensionality are discussed in [7], where the author presents several possible ways to handle the  $n < p$  problem.

We suggest another way of treating the censored observations, in a somewhat naive and straightforward way. Nevertheless, the method shows quite good results for estimating the coefficients of the covariates and in most cases can compare with the method based on the minimum discrepancy approach, all the while being simpler and not as computationally challenging.

### 1.1 Cox's proportional hazards model (PH)

Cox's proportional hazards model [3] assumes the following relation for the hazard function for an individual with characteristics  $x$ :

$$\lambda_x(t) = \lambda_0(t)e^{\beta^T x}, \quad (1)$$

where  $\lambda_0(t)$  is the baseline hazard function which is unknown and  $\beta$  contains the regression coefficients of interest. The covariate  $x$  may contain the clinical and patient history, as well as genomic and treatment information.

Cox's model is a semi-parametric model. While no assumptions about the form of  $\lambda_0(t)$  are made, we assume a parametric form for the effect of the predictors on the hazard.

Parameter estimates are obtained by maximizing the partial likelihood function, which in the absence of tied survival times is

$$L(\beta) = \prod_{i=1}^n \frac{e^{\beta^T x_i}}{\sum_{j \in R_i} e^{\beta^T x_j}}.$$

This product is computed over all the events (failures) and for each event we define the set  $R_i$  as the list of individuals at risk of failure time  $i$ . Censored data are naturally taken into account.

## 1.2 Accelerated lifetime model (ALT)

Another model for survival data which seems reasonable in many studies is the accelerated lifetime (failure time) model. Cox's model suggests proportionality between hazards, while accelerated lifetime model considers different form for the hazards of different individuals. The survival time is

$$T = \exp(\beta^T x) T_0, \quad (2)$$

where  $T_0$  denotes the base survival time and  $\exp(\beta^T x)$  describes the effect of the parameters of the individual on his survival time. If  $\beta^T x > 0$ , the failure occurs later than for the base situation. Alternatively, when  $\beta^T x < 0$ , one speaks of accelerated lifetime since the failure happens earlier.

## 2 Data and Assumptions

To formalize the setup and the assumptions, we adopt the notations from [6]. Let  $Y^o$  be the true (unobservable) lifetime and  $C$  the censoring time.  $\delta = I(Y^o \leq C)$  is the censoring indicator, and we observe  $Y = \min\{Y^o, C\}$ . We assume that  $Y^o$  follows the model

$$Y^o = f(\beta_1^T x, \beta_2^T x, \dots, \beta_k^T x, \epsilon), \quad (3)$$

and that

$$C \perp Y^o \mid x. \quad (4)$$

Without loss of generality, we take unit directions, that is,  $\beta_i^T \beta_i = 1$  for  $i = 1, \dots, k$ . Moreover, we assume that the  $x$ -values are realizations of a random variable  $X$  which satisfies for any vector  $b$ ,

$$E(b^T X \mid \beta_1^T X, \beta_2^T X, \dots, \beta_k^T X) \text{ is linear in } \beta_1^T X, \dots, \beta_k^T X. \quad (5)$$

In (3),  $\beta$  are unknown  $p$ -vectors,  $\epsilon$  is a random variable independent of  $x$ , and the  $f$  is an arbitrary unknown function on  $\mathbb{R}^{k+1}$ . Condition (4) ensures identifiability under the random censoring scheme and condition (5) is called the linearity condition and is satisfied for elliptically symmetric distributions. Some studies suggest that SIR is not very sensitive to the violation of (5) (see [2], for example).

The data sample consists of  $n$  i.i.d. observations  $(Y_i, x_i, \delta_i, i = 1, \dots, n)$ . The continuous random variables  $Y^o$  and  $C$  are not simultaneously observed.

### 3 Sliced Inverse Regression

We avoid dealing directly with a possible high-dimensional covariate vector  $x$  by switching to the inverse problem. Instead of estimating  $Y$  as a function of  $x$ , we regress  $x$  against  $Y$ , which transforms a high-dimensional regression problem into a set of one-dimensional regression problems.

The inverse regression curve  $E(X|Y)$  under the conditions (3) and (5) will fall into a  $k$ -dimensional affine subspace determined by the linear combinations  $\beta_1^T X, \dots, \beta_k^T X$ . For estimation, it is important to standardize the covariates  $x$  to have 0 mean and the identity covariance,  $Z = \Sigma^{-1/2}(x - \mu)$ ,  $\Sigma = \text{Cov}(X)$  and  $\mu = E(X)$ . We can then rewrite the formula (3) as

$$Y^o = f(\eta_1^T z, \eta_2^T z, \dots, \eta_k^T z, \epsilon), \quad (6)$$

where  $\eta_k = \beta_k \Sigma^{1/2}$ .

To estimate the regression curve  $E(Z|Y)$ , we estimate  $\mu$  and  $\Sigma$  and compute the corresponding  $z$ . Then, we slice the sorted  $Y$  into several intervals and we compute the corresponding slice means of  $z$ . The principal component analysis on the slice means of  $z$  defines the most important  $k$ -dimensional subspace for tracking the inverse regression curve  $E(Z|Y)$ . The original directions of  $\beta$  are found after re-transformation back to the original scale.

Throughout this paper, our case of interest is  $k = 1$ , i.e. the single important direction in the subspace  $S_{Y|x}$ . Such an approach is called univariate SIR and corresponds to a regression-type problem. It is often preferred because the results are easily visualized. SIR has been studied in multivariate cases, and other papers in survival treat this case.

#### 3.1 Censored data

As mentioned earlier, the idea of SIR is to divide the response into a fixed number of nonoverlapping intervals and to average the covariates within each of the slice to build the estimate of  $E(X|Y)$ . The question is, where do the censored observations go? Ignoring them is not an option, especially when we do not have a lot of data. We use the following approach: given that the individual was right-censored at time  $t$  which falls in the slice  $s$ , the event for this individual could have taken place anytime after  $t$ . As a first approximation, we attribute this event with equal weights to all subsequent slices. The total sum of the weights naturally equals one. This allows us to use the covariate information of the censored observations. Let us illustrate it with a quick example:

Suppose we have 7 observations, listed below:

$$10, 11^*, 13, 15^*, 17^*, 18, 20. \quad (7)$$

For this data we choose four slices: 10-12, 13-15, 16-18 and 19-20. We then create a matrix of weights which shows in which slice each observation falls.

**Table 1** An example of the weight matrix  $W$  for the dataset (7)

obs	slice	slice 2	slice 3	slice 4
10	1	0	0	0
11*	0.14	0.29	0.29	0.29
13	0	1	0	0
15*	0	0	0.5	0.5
17*	0	0	0.33	0.66
18	0	0	1	0
20	0	0	0	1

The first censored observation 11\* is in the middle of the first slice, hence it is assigned to the second half of this slice and to the next 3 slices, giving a weight of  $1/7$  to the first slice and  $2/7$  to the slices 2-4. The last two censored observations 15\* and 17\* will be taken into account in the slices 3 and 4 but with the different weights. The observation 15\* is considered with the equal weights of  $1/2$ , while the 17\* will have the weight of  $1/3$  in the slice 3 and  $2/3$  in the slice 4. This matrix, listed in Table 1 is used at the third step of the SIR algorithm, when computing the slice mean for covariates. The slice ranges are computed by the R function from the package **dr**, which aims to put approximately the same number of observations in every slice.

Originally slicing was performed on the response  $Y$  being a function of covariates of some form. In survival case, it is the unobservable survival time  $Y^o$  which satisfies (3), whereas in the two regression models, the log-hazard or the log-survival serve as responses.

Our redistribution of weights is surely simplistic since it assumes the risk of an event to be equally distributed among all the remaining slices. In reality, an individual is most likely to experience an event with higher probability in some slices rather than others. But on the other hand, it is easy to implement and gives an idea of the possible contributing covariates. We are currently exploring other options for distributing the weights to the slices.

### 3.2 Algorithm

Let us present the steps of the sliced inverse regression algorithm. We operate on the data  $(Y_i, x_i, \delta_i, i = 1, \dots, n)$ , and  $x_i \in \mathbb{R}^p$  is a vector containing the covariates for the  $i$ -th observation.

1. Standardize  $x$  to get:

$$z_i = \hat{\Sigma}^{-1/2}(x_i - \bar{x}),$$

where  $\hat{\Sigma}^{-1/2}$  and  $\bar{x}$  are the sample covariance matrix and sample mean of  $x_1, \dots, x_n$ , respectively.

2. Divide the range of  $Y_1, \dots, Y_n$  into  $S$  nonoverlapping slices  $H_s, s = 1, \dots, S$ .  $n_s$  denotes the number of observations within slice  $H_s$ , and  $I_{H_s}$  is the

indicator function for this slice:

$$n_s = \sum_{i=1}^n I_{H_s}(Y_i).$$

3. Create the matrix of weights  $W$  for all the observations, as described above. Let  $W_i$  denote the row in the matrix based on the values of  $\delta_i$ . Compute the mean of  $z_i$  over all slices, denoted by  $\bar{z}_s$ .

$$\bar{z}_s = \frac{1}{n_s} \sum_{i=1}^n W_i z_i I_{H_s}(Y_i).$$

4. Calculate the estimate of the weighted covariance matrix

$$\hat{V} = S^{-1} \sum_{s=1}^S n_s \bar{z}_s \bar{z}_s^T.$$

5. Identify the eigenvalues  $\hat{\lambda}_i$  and eigenvectors  $\hat{\eta}_i$  of  $\hat{V}$ , where we choose the indices such that the eigenvalues are sorted in decreasing order.
6. Transform the standardized directions  $\hat{\eta}_i$  back to the original scale.

$$\hat{\beta}_i = \hat{\Sigma}^{-1/2} \hat{\eta}_i.$$

When assuming the dimension of the subspace  $S$  to be  $k = 1$ , we look for the first eigenvector  $\hat{\beta}_1$  which contains the dominant directions. Note that there is a connection between  $\beta_1$  and the regression coefficient  $\beta$  in either (1) or (2). We have  $\beta_1 = \pm\beta/|\beta|$ .

#### 4 Simulation results and comparison with other methods

Simulation studies were performed to evaluate this approach under different conditions. We start with presenting the results under different models and later on compare the estimation with the other methods from [11] and [9].

##### 4.1 Cox's proportional hazards model

The main idea was to generate the survival times following different distributions under the given models and assess the estimation of the  $\beta$  coefficient via our SIR approach. While the covariates were generated from the normal distribution, we studied Weibull, exponential and Gompertz survival times in the PH case. We chose the following setup:  $p = 5$ ,  $X = (X[1], \dots, X[5]) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 2)$ , and the regression coefficient  $\beta = (0.5, -0.5, \frac{1}{\sqrt{2}}, 0, 0)^T$  which we chose to be of length 1. While different sample sizes and censoring patterns were considered, in the tables below we list the averaged results for  $n = 50$  and  $n = 500$ , with censoring percentages of 25% and 50%, all estimated in 1000 runs. Censoring time was computed as a random uniform variable.

**Table 2** SIR estimates and standard deviations of the coefficients of a PH model. 25% of the observations are right-censored. The sign of  $\hat{\beta}_1$  was chosen such that  $\hat{\beta}_1 > 0$ .

	$\hat{\beta}_1[1]$	$\hat{\beta}_1[2]$	$\hat{\beta}_1[3]$	$\hat{\beta}_1[4]$	$\hat{\beta}_1[5]$
Exponential, n=50	0.56 (0.04)	-0.41 (0.06)	0.67 (0.06)	-0.11 (0.05)	0.18 (0.04)
Weibull, n=50	0.43 (0.05)	-0.44 (0.05)	0.77 (0.05)	-0.09 (0.04)	0.13 (0.04)
Gompertz, n=50	0.45 (0.03)	-0.49 (0.04)	0.73 (0.02)	-0.07 (0.04)	0.15 (0.03)
Exponential, n=500	0.54 (0.01)	-0.52 (0.01)	0.66 (0.01)	-0.01 (0.01)	0.04 (0.01)
Weibull, n=500	0.54 (0.01)	-0.51 (0.01)	0.67 (0.01)	0.00 (0.01)	0.03 (0.01)
Gompertz, n=500	0.53 (0.01)	-0.48 (0.01)	0.71 (0.01)	0.00 (0.01)	0.02 (0.01)

From Table 2 we can see that the coefficients are close to the true ones. Even on relatively small samples ( $n = 50$ ), the method performs rather well. Having larger samples brings more accuracy, shrinking the fourth and the fifth coefficients more towards zero. The standard deviations, as expected, get smaller with larger samples.

**Table 3** SIR estimates and standard deviations of the coefficients of a PH model. 50% of the observations are right-censored. The sign of  $\hat{\beta}_1$  was chosen such that  $\hat{\beta}_1 > 0$ .

	$\hat{\beta}_1[1]$	$\hat{\beta}_1[2]$	$\hat{\beta}_1[3]$	$\hat{\beta}_1[4]$	$\hat{\beta}_1[5]$
Exponential, n=50	0.56 (0.06)	-0.48 (0.07)	0.61 (0.07)	0.23 (0.08)	0.21 (0.08)
Weibull, n=50	0.47 (0.06)	-0.51 (0.08)	0.71 (0.09)	0.12 (0.07)	0.07 (0.07)
Gompertz, n=50	0.73 (0.03)	-0.43 (0.05)	0.52 (0.04)	0.04 (0.04)	0.12 (0.04)
Exponential, n=500	0.51 (0.01)	-0.47 (0.01)	0.72 (0.01)	-0.01 (0.01)	0.00 (0.01)
Weibull, n=500	0.52 (0.02)	-0.48 (0.02)	0.71 (0.01)	-0.01 (0.02)	0.01 (0.01)
Gompertz, n=500	0.48 (0.01)	-0.49 (0.01)	0.73 (0.01)	0.04 (0.01)	0.01 (0.01)

Table 3 contains the same results as Table 2, except that a larger proportion of the data was censored (50% instead of 25%). There is much more noise for the smaller sample size which makes the correct estimation of the non-zero variables quite challenging. While the estimates are not so close to their true values, such a procedure can be viewed as a variable selector, to distinguish the most important variables. The larger samples ( $n = 500$ ) do not seem to be influenced much by the severe censoring.

In general, the underlying distribution does not seem to play a major role in the successful recovery of the coefficients. One observes a slight underestimation of the second coefficient, which may be due to a slight bias caused by the equal reweighting. On larger sample this effect is less present.

## 4.2 Accelerated Lifetime model

As a next step, we assess how our procedure performs under the ALT model. Here we also used the exponential and the Weibull distributions, replacing the

Gompertz with the log-normal one. The regression coefficients  $\beta$  remained the same as above, and the similar setups of sample sizes and censoring patterns were considered.

**Table 4** SIR estimates and standard deviations of the coefficients of an ALT model. 25% of the observations are right-censored. The sign of  $\hat{\beta}_1$  was chosen such that  $\hat{\beta}_1 > 0$ .

	$\hat{\beta}_1[1]$	$\hat{\beta}_1[2]$	$\hat{\beta}_1[3]$	$\hat{\beta}_1[4]$	$\hat{\beta}_1[5]$
Exponential, n=50	0.53 (0.07)	-0.40 (0.10)	0.74 (0.11)	-0.04 (0.08)	-0.09 (0.08)
Weibull, n=50	0.50 (0.11)	-0.54 (0.13)	0.66 (0.15)	-0.04 (0.12)	-0.16 (0.11)
Log-Normal, n=50	0.50 (0.06)	-0.37 (0.06)	0.77 (0.05)	-0.09 (0.08)	0.11 (0.07)
Exponential, n=500	0.54 (0.02)	-0.48 (0.02)	0.70 (0.02)	0.03 (0.03)	0.02 (0.03)
Weibull, n=500	0.50 (0.04)	-0.50 (0.04)	0.71 (0.03)	0.02 (0.03)	0.01 (0.03)
Log-Normal, n=500	0.50 (0.02)	-0.52 (0.02)	0.69 (0.02)	0.02 (0.02)	0.03 (0.02)

**Table 5** SIR estimates and standard deviations of the coefficients of an ALT model. 50% of the observations are right-censored. The sign of  $\hat{\beta}_1$  was chosen such that  $\hat{\beta}_1 > 0$ .

	$\hat{\beta}_1[1]$	$\hat{\beta}_1[2]$	$\hat{\beta}_1[3]$	$\hat{\beta}_1[4]$	$\hat{\beta}_1[5]$
Exponential, n=50	0.71 (0.07)	-0.46 (0.09)	0.52 (0.09)	0.01 (0.09)	0.10 (0.10)
Weibull, n=50	0.74 (0.11)	-0.20 (0.13)	0.63 (0.15)	0.09 (0.13)	0.11 (0.14)
Log-normal, n=50	0.71 (0.07)	-0.60 (0.08)	0.34 (0.08)	0.11 (0.09)	0.10 (0.09)
Exponential, n=500	0.52 (0.02)	-0.52 (0.02)	0.67 (0.02)	0.08 (0.02)	0.02 (0.02)
Weibull, n=500	0.54 (0.04)	-0.47 (0.04)	0.70 (0.03)	0.01 (0.04)	0.03 (0.04)
Log-normal, n=500	0.55 (0.02)	-0.51 (0.02)	0.67 (0.02)	0.01 (0.03)	-0.02 (0.03)

In general, the results listed in Tables 4 and 5 present similar properties to the proportional hazards model. One could notice again an underestimation of the second parameter in case of the mild censoring (25%) on lower samples. The estimation under the severe censoring (50%) is of a slightly worse quality, as compared to the PH case.

Naturally, both the degree of censoring and the sample size crucially influence the results. The larger the sample size, the better (and more accurate) estimates we get. The same pattern applies to the degree of censoring. But the sufficiently large sample size can compensate for the severely censored data. If we have a lot of data, we can get good results disregarding the fact that a major part of it has been censored.

### 4.3 Comparison with other methods

Next, we compare our results to those in [11], where the authors also used the accelerated lifetime and Cox's model on the basis of inverse regression with

quadratic discrepancy function. We apply our method to several models from their article. Since the authors of [11] were comparing their method to the double slicing method listed in [6] by computing the mean angle between the basis vector and the eigenvector estimate, we did the same for our method. Let us briefly present these models.

Model 1 has the censoring percentage of 45% under the following setup:  $p=6$ ,  $x = (x_1, \dots, x_6) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . The true survival time  $Y^o$  is generated from

$$Y^o = \exp(x_1 + x_3)\epsilon_1,$$

where  $\epsilon_1$  follows the exponential distribution with parameter 1. The censoring time  $C$  is generated from

$$C = \exp(x_1 + x_2 + x_3)^4.$$

Model 2 has the same setup as model 1, but the true survival time  $Y^o$  is generated from

$$Y^o = -\log(\epsilon_2) / \exp(x_1 + x_3),$$

where  $\epsilon_2$  follows the uniform distribution on  $[0, 1]$ . The censoring time  $C$  is generated from

$$C = \exp(x_1 + x_2 + |x_3|)^2.$$

Figures 1 and 2 show the mean angles between the true basis and the estimates from 100 simulations runs for Model 1 and Model 2 respectively. The true basis equals  $\beta = (1, 0, 1, 0, 0, 0)^T$  for  $p = 6$ . The considered procedures are listed as SIR (our method) and IR (the method by [11]). Each figure has two plots, the part (a) presents the mean angles for a fixed number of predictors ( $p = 6$ ) as the sample size grows, while the part (b) is a function of  $p$  on the sample size of  $n = 200$ . We can see that our method definitely performs worse as the number of parameters grows. When it comes to a fixed  $p$ , it really depends on the model. While the computational cost of the inverse regression via minimum discrepancy approach is unknown to us, it is surely higher than the method we execute.

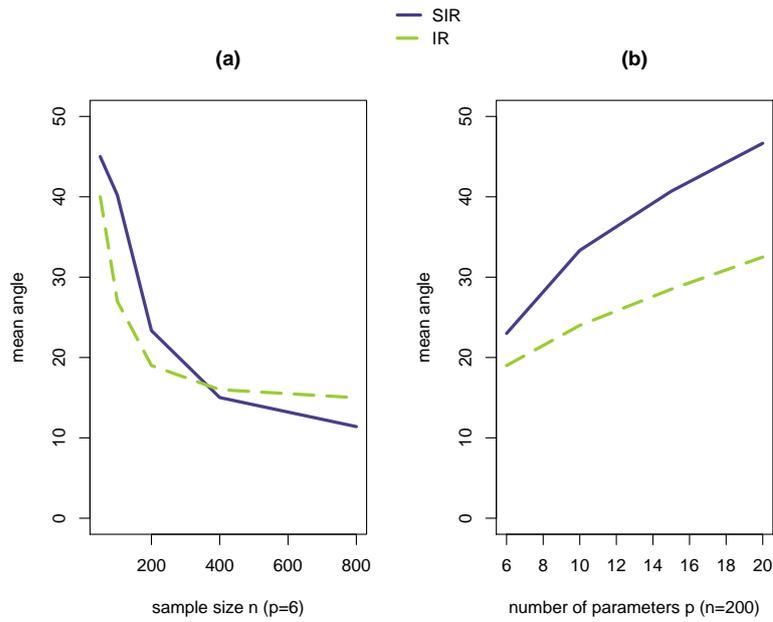
We considered another comparison with a method described in [9], which is the inverse censoring probability weighted estimation. The idea is to use the inverse of the survival function of the censored time to weight the uncensored observations. Authors also view this as a variable selection method. We adopt one of the models they consider and test it in our case.

Model 3 has the the censoring percentage of 40% and  $p=10$  covariates  $x = (x_1, \dots, x_{10}) \sim \mathcal{N}(0, 1)$ . The true survival time is generated from

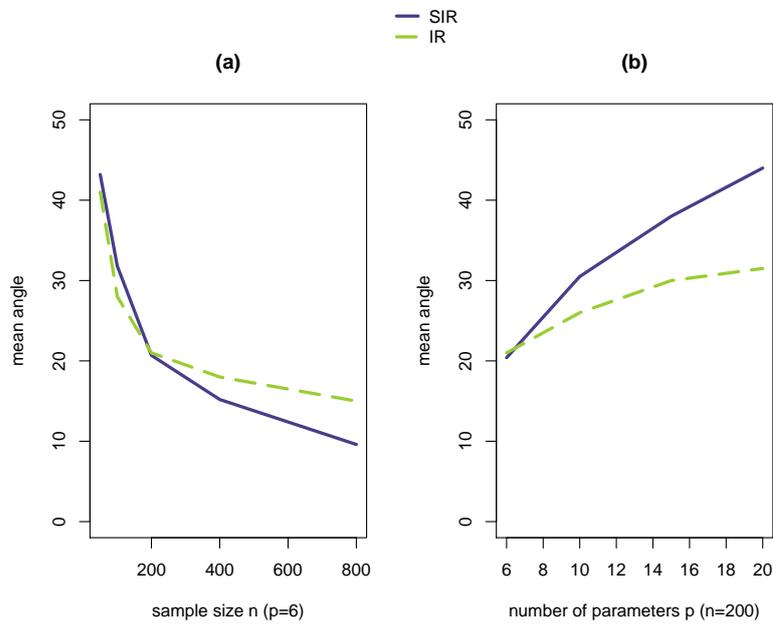
$$Y^o = \exp(-2.5 + \sin(0.1\pi\beta^T x) + 0.1(\beta^T x + 2)^2 + 0.25\epsilon_3),$$

where  $\epsilon_3$  follows the extreme value distribution, and  $\beta = (1, 1, 1, 0, 0, 0, 0, 0, 0, 0)^T$ . The censoring time is generated as

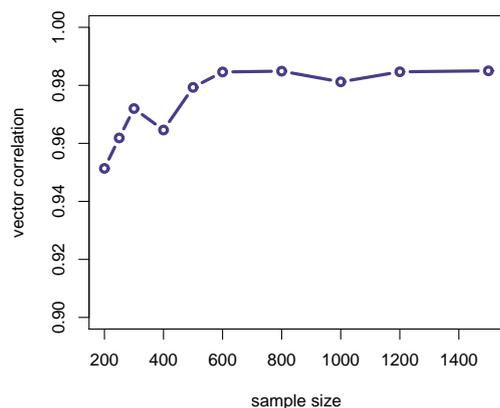
$$C = \exp(c + \beta_c^T X + \epsilon),$$



**Fig. 1** Mean angles between the basis and SIR (our) and IR (alternative) estimates in Model 1.



**Fig. 2** Mean angles between the basis and SIR (our) and IR (alternative) estimates in Model 2.



**Fig. 3** Correlation coefficient between the true and estimated basis for Model 3.

with  $\beta_c = (-1, 0, 0, 1, 0, 0, 0, 0, 0)^T$ , and  $c$  being a constant that controls the censoring percentage.

Figure 3 shows the median correlation coefficient versus the sample size for Model 3, as suggested in [9]. The authors were comparing the different methods of estimating the inverse of the survival function of the censored time, and their resulting curves vary in between 0.8 and 0.98 range for the correlation coefficient. Our results are comparable with their best techniques.

#### 4.4 Discussion

We argue that the SIR procedure with slicing the survival time can be viewed as a variable selector disregarding the underlying model. Why the variable selector? Since our regression coefficients are the components of the first eigenvector, we can never really get their precise values (in simulations above we normalized the results to match the true  $\beta$ ), the best we hope for is the data structure (the signs of the components, their ratio to each other, etc.) While SIR was originally created for dimension reduction, we mostly view it as a tool for pre-selection of the variables and understanding the data structure of different types of survival regressions, if the sample size allows for it. While we concentrate on the one-component SIR, taking the basis of the central subspace to be 1, its general method will surely work as well. Testing for the dimension of the subspace and interpretation of results in the multivariate case is beyond the scope of this paper.

Another open question is how the right-censored observations should be treated. Is it correct to simply put them into the higher hazard slices? This most likely creates a bias in estimation. Censoring itself is known to introduce

a bias for the central subspace of interest and there is no easy way to account for it. We admit the suggested imperfection of our implementation but we find that its simplicity compensates for it. We are currently investigating other ways to disperse such observations over the remaining slices.

While we do not formally introduce the test for predictors here, we rely on bootstrap methods for this task. This is a standard procedure in survival case since no asymptotic variance for censored SIR has been developed. When computing the confidence intervals for predictors, we use the percentile approach from [4].

## 5 Conclusion

Our procedure adapts the idea of sliced inverse regression and applies it to survival data. Censored observations are redistributed in posterior slices. The method allows to pre-select the variables of interest which later can be fitted if necessary by any appropriate technique. The approach in non-parametric and uncovers the data structure even under the severe censoring.

**Acknowledgements** The authors gratefully acknowledge the support by the Swiss National Science Foundation. The authors would also like to thank the referees for their helpful and constructive criticisms.

## References

1. Cook, D.R., Ni, L.: Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *Journal of the American Statistical Association* **100**, 410–428 (2005)
2. Cook, D.R., Weisberg, S.: Discussion of "sliced inverse regression" by k.c. li. *Journal of the American Statistical Association* **86**, 328–332 (1991)
3. Cox, D.R.: Regression models and life tables (with discussion). *Journal of Royal Statistical Society* **34**, 187–200 (1972)
4. Efron, B.: Censored data and bootstrap. *Journal of the American Statistical Association* **76**(374), 312–319 (1981)
5. Li, K.C.: Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86**(414), 316–327 (1991)
6. Li, K.C., Wang, J.L., Chen, C.H.: Dimension reduction for censored regression data. *The Annals of Statistics* **27**(1), 1–23 (1999)
7. Li, L.: Dimension reduction for high-dimensional data. In: *Statistical Methods in Molecular Biology*, vol. 620, pp. 417–434. Springer Protocols (2010)
8. Li, L., Li, H.: Dimension reduction methods for microarrays with application to censored survival data. *Bioinformatics* **20**(18), 3406–3412 (2004)
9. Li, L., Lu, W.: Sufficient dimension reduction with missing predictors. *Journal of the American Statistical Association* **103**(482), 822–831 (2008)
10. Lu, W., Li, L.: Sufficient dimension reduction for censored regressions. *Biometrics* **67**, 513–523 (2011)
11. Nadkarni, N.V., Zhao, Y., Kosorok, M.: Inverse regression estimation for censored data. *Journal of the American Statistical Association* **106**(493), 178–190 (2011)
12. Rotnitzky, A., Robins, J.: Inverse probability weighted estimation in survival analysis (2005)
13. Wen, X., Cook, D.R.: New approaches to model-free dimension reduction for bivariate regression. *Journal of Statistical Planning and Inference* **139**(3), 734–748 (2009)