Computational Statistics and Data Analysis 119 (2018) 74–85

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A new nonparametric screening method for ultrahigh-dimensional survival data

Yanyan Liu [a], Jing Zhang [a], Xingqiu Zhao [b,*]

[a] School of Mathematics and Statistics, Wuhan University, Wuhan, China
[b] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 24 September 2017
Accepted 8 October 2017
Available online 23 October 2017

Keywords:
Model-free
Sure screening property
Ultrahigh-dimensional survival data
Variable screening

ABSTRACT

For ultrahigh-dimensional data, sure independent screening methods can effectively reduce the dimensionality while ensuring that all the active variables can be retained with high probability. However, most existing screening procedures are developed for ultrahigh-dimensional complete data and cannot be applicable to censored survival data. To address the new challenges from censoring, a novel model-free screening method was proposed through the Kolmogorov–Smirnov test statistic that is specially tailored to the ultrahigh-dimensional survival data. The sure screening property was established under some mild regularity conditions, and its superior performance over existing screening methods is demonstrated by our extensive simulation studies. A real data example of gene expression is used to illustrate the application of the proposed fully nonparametric screening procedure.

1. Introduction

Ultrahigh-dimensional survival data are widely encountered in many fields such as genomics, medicine, and public health. Examples of such data include the breast cancer data (Van Houwelingen et al., 2006; Song et al., 2014) and the mantle cell lymphoma data (Rosenwald et al., 2003) that motivated this research. For high-dimensional variable selection, there are numerous regularization methods available, such as the LASSO (Tibshirani, 1996), the smoothly clipped absolute deviation (Fan and Li, 2001), the adaptive LASSO (Zou, 2006), the Dantzig selector (Candes and Tao, 2007), and the minimax concave penalty (Zhang, 2010). These penalized procedures, however, may not perform well for a very large number of covariates because ultrahigh dimensionality brings simultaneous challenges of computational expediency, statistical accuracy and algorithmic stability (Fan et al., 2009). Thus, to apply the well-developed regularization methods, the first important step is to study the dimension reduction for ultrahigh-dimensional data.

To effectively reduce the dimensionality, several model-based and model-free screening methods have been proposed. For example, Fan and Lv (2008) and Fan and Song (2010) presented sure independence screening (SIS) methods based on linear and generalized linear regression models; Fan et al. (2011) proposed a nonparametric screening procedure based on the ultrahigh-dimensional additive models using the B-spline approximation; Wu and Yin (2015) developed a conditional quantile screening procedure based on a nonparametric regression model with heterogeneous errors. To avoid specifying a particular model structure, He et al. (2013), Li et al. (2012a), Zhu et al. (2011), Li et al. (2012b), and Mai and Zou (2013, 2015) among others developed different model-free screening methods using the quantile-adaptive approach, Kendall's τ correlation, ranking, the distance correlation learning, and the fused Kolmogorov filter, respectively.

* Corresponding author.
E-mail address: xingqiu.zhao@polyu.edu.hk (X. Zhao).

https://doi.org/10.1016/j.csda.2017.10.003
0167-9473/

The methods mentioned above are mainly designed for ultrahigh-dimensional complete data, while some of them may be applicable to ultrahigh-dimensional censored data through some particular treatment. In contrast, there is limited research for ultrahigh-dimensional censored data. For example, Fan et al. (2010), Tibshirani (2009), and Zhao and Li (2012) investigated SIS methods for the Cox proportional hazards model; Gorst-Rasmussen and Scheike (2013) proposed a screening procedure for a general class of single-index hazard rate models; Song et al. (2014) developed a model-free rank independence screening method.

Note that model-free screening procedures are much more robust in the sense that the sure screening property can hold under much weaker conditions, and choice of statistical modeling approaches after variable screening can be much more flexible (Mai and Zou, 2015). In particular, the fused Kolmogorov filter proposed by Mai and Zou (2015) has the superior performance over existing screening methods for ultrahigh-dimensional data. However, the fused Kolmogorov filter in Mai and Zou (2015) cannot handle the censored survival data. So, our goal is to develop a new fully nonparametric model-free variable screening method for ultrahigh-dimensional survival data using an appropriate Kolmogorov–Smirnov measure.

The remainder of the article is organized as follows. In Section 2, we introduce a different Kolmogorov–Smirnov measure by taking censoring into account and propose a model-free screening procedure for different types of covariates. The sure screening property of the proposed procedure is presented in Section 3. We assess the finite-sample performance of the proposed screening procedure in Section 4 via extensive simulation studies. A gene expression data example from the mantle cell lymphoma study is analyzed in Section 5. Some concluding remarks are made in Section 6. All technical proofs are presented in Appendix.

2. Screening procedures via fused Kolmogorov–Smirnov statistic

For ultrahigh-dimensional survival data, suppose that the observations $\{X_i, \Delta_i, \mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{ip})^{\mathrm{T}} : i = 1, \ldots, n\}$ are independent copies of $\{X, \Delta, \mathbf{Z} = (Z_1, \ldots, Z_p)^{\mathrm{T}}\}$, where $\mathbf{Z}$ is a p-dimensional vector of covariates, T and C denote the survival and censoring times, respectively, $X = \min(T, C)$, and $\Delta = I(T \leq C)$. For ease of exposition, we assume that the censoring mechanism is random, i.e., the survival time T and the censoring time C are conditionally independent given $\mathbf{Z}$. Let τ denote the end time of the study.

Consider the conditional survival function

$$S(t|\mathbf{Z}) = Pr(T > t|\mathbf{Z})$$

given $\mathbf{Z}$. In an ultrahigh-dimensional setting, the dimensionality p greatly exceeds sample size n. A popular assumption is the sparsity assumption that only a small subset of covariates actually contribute to the conditional survival function $S(t|\mathbf{Z})$. To identify those active ones from p covariates, we define the set of active covariates as

$$\mathcal{A} = \{j : S(t|\mathbf{Z}) \text{ depends on } Z_j, j = 1, \ldots, p\}.$$

Our goal is to recover the active set $\mathcal{A}$ as precisely as possible based on the censored observations.

To accommodate the continuous or general discrete covariates, we need to use the idea of slicing as used in Mai and Zou (2015) and construct a Kolmogorov–Smirnov statistic to measure the dependence between each covariate and survival time. For each covariate Z_j, we define a partition

$$\mathbf{\Lambda}_j = \{[a_l^j, a_{l+1}^j) : \ a_l^j < a_{l+1}^j, \ l = 0, \ldots, \Lambda_j - 1 \text{ and } \bigcup_{l=0}^{\Lambda_j - 1} [a_l^j, a_{l+1}^j) \setminus \{a_0^j\} = \mathbb{R}\},$$

where $a_0^j = -\infty$, $a_{\Lambda_j}^j = \infty$ and $j = 1, \ldots, p$. Each $[a_l^j, a_{l+1}^j)$ is called a slice. Then define a random variable $I_j \in \{1, \ldots, \Lambda_j\}$ such that $I_j = l + 1$ if and only if $Z_j \in [a_l^j, a_{l+1}^j)$. Let

$$K_j^{\mathbf{\Lambda}_j} = \max_{l_1, l_2} \sup_{0 \leq t \leq \tau} |S_j(t|I_j = l_1) - S_j(t|I_j = l_2)|,$$

where $S_j(t|I_j = l_1) = Pr(T > t|I_j = l_1)$. It is obvious that when Z_j ($j = 1, \ldots, p$) takes finite values such that each possible value forms a slice, Z_j is independent of T if and only if $K_j^{\mathbf{\Lambda}_j} = 0$. If Z_j is continuous or discrete, the following lemma shows that $K_j^{\mathbf{\Lambda}_j}$ also sheds light on the dependence between Z_j and T. Following Lemma 1 of Mai and Zou (2015), we have the following results:

(i) Z_j is independent of T if and only if $K_j^{\mathbf{\Lambda}_j} = 0$ for any partition $\mathbf{\Lambda}_j$;

(ii) If Z_j is not independent of T and for any fixed $z \in \mathbb{R}$, $Pr(Z_j \leq z|T = t)$ is not a constant in t, then $K_j^{\mathbf{\Lambda}_j} \neq 0$ for any partition $\mathbf{\Lambda}_j$.

The above results play the theoretical foundation so that we can use $K_j^{\mathbf{\Lambda}_j}$ to judge if Z_j is independent of T. Given the observed data $\{X_i, \Delta_i, \mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{ip})^{\mathrm{T}} : i = 1, \ldots, n\}$, the estimate of $K_j^{\mathbf{\Lambda}_j}$ is defined as

$$\widehat{K}_j^{\mathbf{\Lambda}_j} = \max_{l_1, l_2} \sup_{0 \leq t \leq \tau} |\widehat{S}_j(t|I_j = l_1) - \widehat{S}_j(t|I_j = l_2)|,$$

where $\widehat{S}_j(t|I_j = l)$ is the Kaplan–Meier estimator of $S_j(t|I_j = l)$ based on the subsample $\{X_i, \Delta_i, \mathbf{Z}_i : i \in D_{lj}\}$ with $D_{lj} = \{i : Z_{ij} \in [a_{l-1}^j, a_l^j), i = 1, \ldots, n\}$. To be specific,

$$\widehat{S}_j(t|I_j = l) = \prod_{i \in D_{lj}} \left\{ 1 - \frac{1}{\sum_{k \in D_{lj}} I(X_k \geq X_i)} \right\}^{\Delta_i I(X_i \leq t)}.$$

To evaluate $\widehat{K}_j^{\mathbf{\Lambda}_j}$, we suggest an intuitive uniform slicing to partition data into Λ_j slices as follows:

(a) If Z_j is categorical with levels $1, \ldots, \Lambda_j$ or Z_j is discrete with finite possible values $1, \ldots, \Lambda_j$, we set $I_j = Z_j$;
(b) If Z_j is discrete and can take infinite values such as $1, 2, \ldots$, we set $I_j = Z_j$ if $Z_j < \Lambda_j$ and $I_j = \Lambda_j$ if $Z_j \geq \Lambda_j$;
(c) If Z_j is continuous, we let the partition $\mathbf{\Lambda}_j$ contain the intervals bounded by the $\frac{l}{\Lambda_j}$th sample quantiles of Z_j for $l = 0, 1, \ldots, \Lambda_j$.

Furthermore, motivated by Cook and Zhang (2014) and Mai and Zou (2015), we use the idea of fusion to improve the efficiency of the Kolmogorov–Smirnov measure. Suppose that for Z_j in cases (b) and (c), we have N_j different partitions $\mathbf{\Lambda}_{kj}$ $(k = 1, \ldots, N_j)$, where each partition $\mathbf{\Lambda}_{kj}$ contains Λ_{kj} intervals. Then we set

$$\widehat{K}_j = \sum_{k=1}^{N_j} \widehat{K}_j^{\mathbf{\Lambda}_{kj}}$$

as an estimate of $K_j = \sum_{k=1}^{N_j} K_j^{\mathbf{\Lambda}_{kj}}$. As Mai and Zou (2015) suggested, we choose $\Lambda_{kj} \leq \lceil \log n \rceil$ for all $1 \leq k \leq N_j$ so that there is a decent sample size within each slice for all slicing schemes, where $\lceil x \rceil$ denotes the integer part of x. In practice, we take $\Lambda_{kj} = 3, \ldots, \lceil \log n \rceil$ for each partition $\mathbf{\Lambda}_{kj}$. Naturally, for Z_j in case (a), we set $N_j = 1$. Then based on $\widehat{K}_j$ $(j = 1, \ldots, p)$, we define the estimated active set as

$$\widehat{\mathcal{A}}(d_n) = \{1 \leq j \leq p : \widehat{K}_j \text{ is amongst the first } d_n \text{ largest of all } \widehat{K}_j\},$$

where d_n is a prespecified positive integer. In the following, the proposed procedure is called the fused Kolmogorov–Smirnov statistic-based sure independence screening (KS-SIS).

3. Sure screening property

In this section, we establish the sure screening property of the proposed screening method. If we know the distribution of Z_j, then we can consider an oracle uniform slicing to form partitions $\mathbf{\Lambda}_{kj}$ $(k = 1, \ldots, N_j)$ by replacing the $\frac{l}{\Lambda_{kj}}$th sample quantiles of Z_j with the $\frac{l}{\Lambda_{kj}}$th theoretical quantiles of Z_j for each continuous covariate Z_j in an intuitive uniform slicing scheme. For this slicing, we set $K_j^{(o)}(\mathbf{\Lambda}_{kj}) = K_j^{\mathbf{\Lambda}_{kj}}$, $K_j^{(o)} = \sum_{k=1}^{N_j} K_j^{(o)}(\mathbf{\Lambda}_{kj})$, $\widehat{K}_j^{(o)}(\mathbf{\Lambda}_{kj}) = \widehat{K}_j^{\mathbf{\Lambda}_{kj}}$ and $\widehat{K}_j^{(o)} = \sum_{k=1}^{N_j} \widehat{K}_j^{(o)}(\mathbf{\Lambda}_{kj})$. Here, if Z_j is discrete, these oracle quantities are the same as those under an intuitive uniform slicing scheme. Then we can obtain the oracle active set as

$$\widehat{\mathcal{A}}(oracle) = \{1 \leq j \leq p : \widehat{K}_j^{(o)} \text{ is amongst the first } d_n \text{ largest of all } \widehat{K}_j^{(o)}\}.$$

Let $G(\cdot)$ denote the survival function of censoring time. Throughout this paper, c denotes a generic positive constant which may take different values in different places. To establish the sure screening property, we need the following mild regularity conditions.

C1. There exists a set $\mathcal{B}$ such that $\mathcal{A} \subset \mathcal{B}$ and

$$\Delta_{\mathcal{B}} = \min_{j \in \mathcal{B}} \min_{1 \leq k \leq N_j} K_j^{(o)}(\mathbf{\Lambda}_{kj}) - \max_{j \notin \mathcal{B}} \max_{1 \leq k \leq N_j} K_j^{(o)}(\mathbf{\Lambda}_{kj}) > 0.$$

C2. If Z_j is continuous, then for any d_1, d_2 with $\Pr(Z_j \in [d_1, d_2)) \leq \frac{2}{\Lambda_{0j}}$,

$$|S_j(t|z_1) - S_j(t|z_2)| \leq \frac{\Delta_{\mathcal{B}}}{8}$$

for all t, j and $z_1, z_2 \in [d_1, d_2)$, where $\Lambda_{0j} = \min_k\{\Lambda_{kj}\}$.

Condition C1 is the key condition which has been widely used for establishing the sure screening property of marginal screening methods in the literature. Condition C2 guarantees that the sample quantiles of Z_j $(j = 1, 2, \ldots, p)$ are close enough to the population quantiles of Z_j. Obviously, this result is expected for many distributions of Z_j. Here, we can see that the conditions for our method are much weaker than those required for many existing nonparametric screening methods. We make no assumption on the distribution of the covariates. Moreover, we do not impose any moment conditions for predictors Z_j's. Compared with the exponential moment condition C1 in Li et al. (2012b) and C3 in Zhu et al. (2011), ours are weaker. Conditions C1 and C2 are similar to those given in Mai and Zou (2015) with more detailed discussions.

Theorem 1. *Suppose conditions C1 and C2 hold. Define*

$$\eta_1 = cNp(\log^2 n)\exp\left(-c\frac{n\Delta_{\mathcal{B}}^2}{\log n} + c\Delta_{\mathcal{B}}\sqrt{n}\right) + cNp(\log^2 n)\exp\left(-c\frac{n}{\log^2 n} + c\Delta_{\mathcal{B}}\sqrt{n}\right),$$

and

$$\eta_2 = cNp(\log^2 n)\exp\left(-c\frac{n\Delta_{\mathcal{B}}^2}{\log n} + c\Delta_{\mathcal{B}}\sqrt{n}\right) + cNp\exp\left(-c\frac{n}{\log^2 n}\right)$$

with $N = \max(N_j, j = 1, \ldots, p)$. *If* $\Lambda_{kj} \le [\log n]$ *for* $k = 1, 2, \ldots, N_j$ *and* $d_n \ge |\mathcal{B}|$, *we have the following conclusions:*

(1) $Pr(\mathcal{A} \subset \widehat{\mathcal{A}}(oracle)) \ge 1 - \eta_1$;
(2) $Pr(\mathcal{A} \subset \widehat{\mathcal{A}}(d_n)) \ge 1 - \eta_2$.

Therefore, both the oracle fused KS-SIS and the fused KS-SIS enjoy the sure screening property with probability going to 1 *if* $\Delta_{\mathcal{B}} \gg n^{-1/2}\{\log n \log(pN \log n)\}^{1/2}$.

From the two results in Theorem 1, we can see that the fused method by an intuitive uniform slicing is as efficient as one by an oracle uniform slicing. On the other hand, we can see that our method enjoys the sure screening property with a probability tending to one for a reasonably large d_n. This can be easily achieved in practice.

4. Simulation studies

We examined the finite sample performance of the proposed method and made comparisons with existing methods via simulation studies. For convenience, we denoted the proposed Kolmogorov–Smirnov statistic-based sure independence screening procedure as fused KS-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013) as FAST-SIS, the principled sure independent screening procedure of Zhao and Li (2012) as P-SIS, the censored rank independence screening of Song et al. (2014) as CRIS, and the quantile adaptive screening procedure of He et al. (2013) at quantile level 0.5 as QASIS. To demonstrate the robustness of the proposed method, we considered survival data generated from different models including the Cox proportional hazards model, nonlinear model, and linear transformation model.

Example 1. Suppose that survival time T followed the Cox proportional hazards model with the conditional hazard function given by

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t)\exp(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_0),$$

where the baseline hazard function was set to be $\lambda_0(t) = (t-0.5)^2$ and the ultrahigh-dimensional covariate $\mathbf{Z} = (Z_1, \ldots, Z_p)$ followed a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{\Sigma} = (0.8^{|i-j|})$ for $i, j = 1, \ldots, p$. We set the true parameter $\boldsymbol{\beta}_0 = (0.35, 0.35, 0.35, 0.35, 0.35, 0, \ldots, 0)^{\mathrm{T}}$, i.e., only the first five predictors are active. We took the censoring time $C = \widetilde{C} \wedge \tau$, where $\widetilde{C}$ was generated from Unif$(0, \tau + 2)$ and the study duration τ was chosen to yield the desirable censoring rates (CR) of 20% and 40%, respectively.

Example 2. We generated the survival time T from the following nonlinear model with interaction:

$$\log T = (2 + \sin Z_1)^2 + (1 + Z_5)^3 + 3Z_{10}^2 + Z_1 Z_{10} + \epsilon,$$

where the error term ϵ followed the standard normal distribution. The setup for censoring time C was the same as those in Example 1. Here, the predictor Z_1 was generated from the discrete uniform distribution over $\{-2, -1.5, \ldots, 2, 2.5\}$, which is a categorical variable taking each value with an equal probability of 0.1, and the remaining $(p-1)$ covariates $(Z_2, \ldots, Z_p)$ followed a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{\Sigma} = (0.8^{|i-j|})$ for $i, j = 1, \ldots, (p-1)$.

Example 3. Suppose that the survival time T took the following general transformation model adapted from Song et al. (2014)

$$H(T) = -\boldsymbol{\beta}'\mathbf{Z} + \sigma(\mathbf{Z})\epsilon,$$

where $H(t) = \log\{0.5(e^{2t} - 1)\}$, and ϵ followed the standard normal distribution. We set the true parameter $\boldsymbol{\beta} = (1, 0.7, 0_6', 0.8, 1.0, 0_{p-10}')^{\mathrm{T}}$, i.e., only four predictors are active. The setups for covariate $\mathbf{Z}$ and censoring time C were the same as those in Example 1. For $\sigma(\mathbf{Z})$, we considered two cases: $\sigma(\mathbf{Z}) = \sigma_1(\mathbf{Z}) = 1$ or $\sigma(\mathbf{Z}) = \sigma_2(\mathbf{Z}) = \exp(Z_{20} + Z_{21} + Z_{22})$. For KS-SIS, CRIS, FAST-SIS, P-SIS and CSIRS, the sizes of the true active predictor set for these two cases are $p_1 = 4$ and $p_2 = 7$, respectively. As QASIS with quantile level 0.5 can only detect the predictors affecting the median, the corresponding sizes are $p_1 = p_2 = 4$.

We took the sample sizes $n = 50, 100, 200$, the number of covariates $p = 2000$ and the censoring rates CR = 20%, 40%. For each configuration, we repeated 500 times.

Table 1
The median and IQR of $\mathcal{S}$, the selection proportions $\mathcal{P}_e$ and $\mathcal{P}_a$ among 500 replications for Example 1 with the censoring rate CR $= 20\%, 40\%$.

CR	n	Method	Median	IQR	$\mathcal{P}_e$					$\mathcal{P}_a$
					X_1	X_2	X_3	X_4	X_5	
20%	50	Fused KS-SIS	44	131	0.538	0.770	0.790	0.728	0.554	0.204
		CRIS	80	227	0.436	0.646	0.678	0.600	0.402	0.174
		FAST-SIS	6	5	0.880	0.976	0.990	0.986	0.916	0.800
		P-SIS	6	3	0.904	0.990	1.000	0.998	0.946	0.856
		QASIS	128	195	0.220	0.370	0.436	0.336	0.240	0.014
	100	Fused KS-SIS	5	1	0.976	1.000	1.000	0.998	0.986	0.964
		CRIS	7	9	0.890	0.970	0.988	0.946	0.884	0.808
		FAST-SIS	5	0	0.998	1.000	1.000	1.000	0.998	0.996
		P-SIS	5	0	1.000	1.000	1.000	1.000	0.998	0.998
		QASIS	11	14	0.890	0.966	0.984	0.970	0.880	0.758
	200	Fused KS-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		CRIS	5	0	0.994	1.000	1.000	1.000	0.998	0.992
		FAST-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		P-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		QASIS	5	1	1.000	1.000	1.000	1.000	1.000	1.000
40%	50	Fused KS-SIS	90	234	0.388	0.586	0.650	0.592	0.402	0.102
		CRIS	103	297	0.350	0.514	0.612	0.530	0.392	0.104
		FAST-SIS	8	13	0.802	0.956	0.970	0.950	0.848	0.664
		P-SIS	7	9	0.844	0.970	0.986	0.972	0.870	0.716
		QASIS	525	700	0.044	0.072	0.100	0.088	0.070	0.000
	100	Fused KS-SIS	6	4	0.958	0.992	0.996	0.998	0.952	0.904
		CRIS	8	19	0.848	0.944	0.966	0.934	0.850	0.726
		FAST-SIS	5	0	0.996	1.000	1.000	1.000	0.998	0.994
		P-SIS	5	0	0.996	1.000	1.000	1.000	1.000	0.996
		QASIS	120	251	0.408	0.602	0.698	0.588	0.420	0.118
	200	Fused KS-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		CRIS	5	1	0.994	1.000	1.000	1.000	0.996	0.990
		FAST-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		P-SIS	5	0	1.000	1.000	1.000	1.000	1.000	1.000
		QASIS	7	9	0.944	0.994	0.992	0.988	0.936	0.874

To assess the performance of the screening procedures, we employed three criteria used in Li et al. (2012b). The first one is the minimum model size to include all active predictors, denoted by $\mathcal{S}$. Note that $\mathcal{S}$ can be used to measure the model complexity obtained from a screening procedure. The closer it is to the true minimum model size, the better the screening procedure is. We presented the median and interquartile range (IQR) of $\mathcal{S}$ out of 500 replications. The second one is the proportion that an individual active predictor is selected for a given model size over 500 replications, denoted by $\mathcal{P}_e$. The third one is the proportion that all active predictors are selected for a given model size over 500 replications, denoted by $\mathcal{P}_a$. An effective screening procedure is expected to yield $\mathcal{S}$ close to the true minimum model size and both $\mathcal{P}_e$ and $\mathcal{P}_a$ close to one. We chose the model size to be $d_n = \lceil n/\log n \rceil$.

The simulation results for $\mathcal{S}$, $\mathcal{P}_e$ and $\mathcal{P}_a$ are summarized in Tables 1–3. Tables 1 and 2 include the results obtained by the proposed fused KS-SIS, CRIS (Song et al., 2014), FAST-SIS (Gorst-Rasmussen and Scheike, 2013), P-SIS (Zhao and Li, 2012), and QASIS (He et al., 2013) for Examples 1 and 2, respectively. Table 3 displays the results by the proposed KS-SIS under a single slicing ($\Lambda_{kj} = 3, 4, 5$) and a fusion as well as the other four screening methods for Example 3 with $n = 200$.

In summary, we have the following findings:

(i) The proposed Kolmogorov–Smirnov statistic-based independence screening method through a single slicing works well and performs stably with the choice of number of slices, while the fused screening procedure tends to be more efficient.
(ii) The proposed fused KS-SIS procedure outperforms the other two model-free screening methods CRIS (Song et al., 2014) and QASIS (He et al., 2013) for all the situations considered here.
(iii) For the Cox model considered in Example 1, the two model-based screening procedures FAST-SIS (Gorst-Rasmussen and Scheike, 2013) and P-SIS (Zhao and Li, 2012) outperform the proposed fused KS-SIS and the other two model-free screening methods when the sample size is 50, as they take into account the Cox proportional model structure while the fully nonparametric screening methods do not rely on the specific model structure. When the sample size is increased to 100, the simulation results by the model-free screening methods are comparable to those by the two model-based screening methods.
(iv) For the nonlinear model with interaction considered in Example 2, the proposed screening procedure performs best among these five methods.

Table 2
The median and IQR of $\mathcal{S}$, the selection proportions $\mathcal{P}_e$ and $\mathcal{P}_a$ among 500 replications for Example 2 with the censoring rate CR = 20%, 40%.

CR	n	Method	Median	IQR	$\mathcal{P}_e$			$\mathcal{P}_a$
					X_1	X_5	X_{10}	
20%	50	Fused KS-SIS	189	543	1.000	0.692	0.148	0.096
		CRIS	1346	986	0.062	0.442	0.006	0.000
		FAST-SIS	575	952	0.254	0.948	0.094	0.022
		P-SIS	739	947	0.292	0.904	0.034	0.002
		QASIS	580	673	0.010	0.436	0.142	0.000
	100	Fused KS-SIS	10	31	1.000	1.000	0.696	0.696
		CRIS	1254	1092	0.186	0.712	0.016	0.008
		FAST-SIS	185	641	0.684	1.000	0.226	0.162
		P-SIS	317	806	0.692	1.000	0.116	0.082
		QASIS	241	411	0.060	0.888	0.516	0.026
	200	Fused KS-SIS	5	3	1.000	1.000	0.996	0.996
		CRIS	1053	1193	0.394	0.922	0.066	0.054
		FAST-SIS	35	176	0.968	1.000	0.536	0.514
		P-SIS	84	300	0.974	1.000	0.372	0.362
		QASIS	65	152	0.360	0.988	0.914	0.332
40%	50	Fused KS-SIS	283	664	1.000	0.584	0.109	0.079
		CRIS	1667	747	0.020	0.178	0.000	0.000
		FAST-SIS	828	978	0.436	0.871	0.020	0.000
		P-SIS	937	956	0.475	0.802	0.020	0.000
		QASIS	1732	872	0.040	0.366	0.119	0.000
	100	Fused KS-SIS	15	51	1.000	0.990	0.586	0.578
		CRIS	1734	636	0.080	0.278	0.004	0.000
		FAST-SIS	525	1001	0.846	1.000	0.046	0.042
		P-SIS	640	1017	0.854	0.998	0.026	0.026
		QASIS	1969	253	0.046	0.470	0.152	0.000
	200	Fused KS-SIS	5	4	1.000	1.000	0.976	0.976
		CRIS	1837	506	0.110	0.436	0.006	0.006
		FAST-SIS	278	709	0.996	1.000	0.224	0.222
		P-SIS	377	790	0.998	1.000	0.136	0.136
		QASIS	2000	4	0.040	0.572	0.156	0.004

(v) For the transformation model with homoscedastic normal error considered in Example 3, the simulation results show that our proposed procedure is comparable to the other four methods. But for the model with heteroscedastic error, the proposed method performs much better than the other four screening methods.
(vi) The proposed procedure also performs well when the censoring rate is increased to 40%, but the QASIS seems not to be applicable to the cases of high censoring.

Therefore, we can conclude that the proposed model-free screening method is more robust and it is a superior model-free screening technique which can be widely applicable to ultrahigh-dimensional survival analysis.

5. An application

As an illustration, we applied the proposed method to the mantle cell lymphoma (MCL) data, which was studied by Rosenwald et al. (2003). The data contain expression values of 8810 cDNA elements and can be downloaded from the web site http://llmpp.nih.gov/MCL. The primary goal of this study was to identify genes that have great influence on patients' survival risk. Among 101 untreated patients with no history of previous lymphoma, 92 were classified as having MCL based on the morphologic and immunophenotypic criteria. During the follow-up, 64 patients died of MCL and the other 28 patients were censored. The mean observed survival time was 2.8 years (ranging from 0.02 to 14.05 years). Given such small sample size and huge number of predictors, feature screening serves as a preliminary step prior to any other complicated statistical modeling techniques can be applied subsequently.

For comparison, we also applied CRIS, P-SIS, FAST-SIS, and QASIS to analyze this data. By excluding the genes with missing values, we first screened the important predictors among the 6312 genes and set the model size to be 20 ($= \lceil 92/\log(92) \rceil$). The gene unique identification (UNIQID) of these genes were summarized in Table 4. By observing the results, we can see that the gene with UNIQID 28872 was selected by all the five considered screening methods, indicating that this gene could be strongly associated with patients' survival risk; seven genes with UNIQID 32737, 24794, 30142, 17198, 27310, 32049, and 28805 were uniquely selected by our method. In particular, as pointed out by Rosenwald et al. (2003), the selected gene with UNIQID 30142 had an important effect on the survival time, but it was not selected by other screening methods.

To compare the predictive accuracy of these methods, we further adopted the C-statistic estimator proposed by Uno et al. (2011) to evaluate the fitted models using the 20 selected predictors by different screening methods. In practice, the true model is unknown. Here, we took the Cox model as a working model. Then risk scores were obtained from the fitted

Table 3
The median and IQR of $\mathcal{S}$, the selection proportion $\mathcal{P}_a$ among 500 replications for Example 3.

σ	Method	CR = 20%			CR = 40%		
		Median	IQR	$\mathcal{P}_a$	Median	IQR	$\mathcal{P}_a$
σ_1	KS-SIS						
	$\Lambda = 3$	5	2	1.000	5	2	1.000
	$\Lambda = 4$	5	2	1.000	5	2	1.000
	$\Lambda = 5$	5	3	1.000	5	3	1.000
	Fused	5	2	1.000	5	2	1.000
	CRIS	5	2	0.996	5	2	1.000
	FAST-SIS	5	1	1.000	4	1	1.000
	P-SIS	4	1	1.000	4	1	1.000
	QASIS	6	3	1.000	1861	650	0.022
σ_2	KS-SIS						
	$\Lambda = 3$	25	31	0.682	62	124	0.340
	$\Lambda = 4$	33	44	0.540	103	174	0.216
	$\Lambda = 5$	55	72	0.346	170	243	0.122
	Fused	21	16	0.804	55	99	0.388
	CRIS	1357	1062	0.008	636	996	0.050
	FAST-SIS	265	784	0.142	1487	897	0.000
	P-SIS	265	770	0.132	1478	915	0.004
	QASIS	8	6	0.962	1995	38	0.000

Table 4
The screened UNIQIDs of the first 20 selected genes of the five screening methods for the mantle cell lymphoma data.

Fused KS-SIS	CRIS	P-SIS	FAST-SIS	QASIS
28346	30334	30157	30157	30834
28990	28872	34771	27095	34546
30334	17326	27095	34771	24816
32737	28990	27019	34790	33457
30157	17370	27762	32699	28058
27762	34790	30282	29330	27884
17176	34771	16587	28346	31571
25234	31420	28872	24713	28872
24794	27049	28346	16587	31413
30142	25234	34790	27762	30951
17198	16528	24723	15936	27935
34771	32699	25234	30282	30040
27310	30157	34687	25234	26589
31420	30282	32699	24723	28150
34790	27095	24734	27049	16960
32049	32187	24656	27019	24448
28872	33549	16528	28872	17815
32187	24710	17343	29209	31573
24723	24404	27049	31420	32081
28805	17176	31420	17343	24493

Table 5
The C-statistic and the standard deviation (SD) under a working Cox model using the predictors selected by each of the five screening methods for the mantle cell lymphoma data.

	Fused KS-SIS	CRIS	P-SIS	FAST-SIS	QASIS
C-statistic	0.764	0.748	0.753	0.747	0.744
SD	0.036	0.043	0.037	0.047	0.044

models after screening and the corresponding C-statistic was computed. The standard deviation (SD) of the C-statistic was obtained from a perturbation resampling method with 200 times. Table 4 reports the values of C-statistic and SD obtained from different methods, respectively. According to Uno et al. (2011), the larger the C-statistic is, the stronger predictive power the method possesses. It can be seen from the results in Table 5 that our proposed method has the best performance under the working Cox model for the data after screening. This suggests that the set of active predictors selected by the new nonparametric screening method is the most reasonable one.

According to the referees' suggestions, we randomly split the full data into a training set ($n_1 = 55$) and a testing set ($n_2 = 37$) by the ratio of 3:2, while maintaining the censoring rate roughly the same in each set. For each split, we first applied the five screening methods to select top 13 ($= \lceil 55/\log(55) \rceil$) covariates using the training set, and then computed the C-statistic using these selected covariates in the testing set. A total of 200 splits was made and the average C-statistic and

Table 6
The average C-statistic and the standard deviation (SD) under a working Cox model and testing set using the predictors selected by training set for the mantle cell lymphoma data.

	Fused KS-SIS	CRIS	P-SIS	FAST-SIS	QASIS
C-statistic	0.789	0.780	0.785	0.793	0.754
SD	0.041	0.045	0.040	0.039	0.050

the SD were reported in Table 6. By observing the results, the performance of our proposed method and the two model-based methods P-SIS, FAST-SIS are comparable, and better than the CRIS and QASIS methods. Our proposed method does not show its distinctive superiority over the two model-based methods P-SIS and FAST-SIS. There are two possible reasons: (i) when we split the full data into training and testing sets, the sample sizes are very small in these two sets so that we do not have enough samples within each slice for our method; (ii) the Cox hazards model may be suitable for fitting such data; in this case, our analysis results are consistent with the simulation results.

6. Conclusion

To accommodate censoring in ultrahigh-dimensional survival data, we have considered replacing the conditional distribution of each covariate given a response variable in Mai and Zou's (2015) Kolmogorov filter with a conditional distribution of a response variable given each covariate, and then used the Kaplan–Meier estimator for estimation of unknown conditional distributions. The proposed Kolmogorov–Smirnov statistic-based independence screening method can deal with discrete, categorical or continuous covariates, and has several distinctive advantages over the existing screening procedures for ultrahigh-dimensional survival data. First, our procedure does not rely on any model assumption. Second, our approach is invariant under the monotone transformation of the response. Third, the new method enjoys the sure screening property under much weaker conditions without any moment conditions. Fourth, the simulation studies demonstrate that the proposed model-free screening method is more robust than the model-based screening approaches (Gorst-Rasmussen and Scheike, 2013; Zhao and Li, 2012), the rank-based (Song et al., 2014) and the quantile-adaptive-based screening procedures (He et al., 2013). All these advantages greatly facilitate its implementation in real applications.

Acknowledgments

The authors would like to thank the Editor, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. This research is partly supported by the Research Grant Council of Hong Kong (503513), the National Natural Science Foundation of China (No. 11371299), and The Hong Kong Polytechnic University.

Appendix. Proofs of theoretic results

We note that the Dvoretzky–Kiefer–Wolfowitz inequality plays a key role in proving the sure screening property of the fused Kolmogorov filter in Mai and Zou (2015). Bitouzé et al. (1999) established a Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator of the survival function in a right censored data model. We will use this exponential inequality to show the sure screening property of the proposed Kolmogorov–Smirnov statistic-based screening method for ultrahigh-dimensional censored data. As a preparation, we state this result as the following lemma.

Lemma 1. *Let $\widehat{S}_{KM}(\cdot)$ be the Kaplan–Meier estimator of the survival function $S(\cdot)$. There exists an absolute constant μ such that, for any positive λ,*

$$P\left(\sqrt{n}\sup_{0\le t\le\tau}\left|G(t)\left(\widehat{S}_{KM}(t)-S(t)\right)\right|>\lambda\right)\le 2.5\exp(-2\lambda^2+\mu\lambda),$$

where $S(\cdot)$ and $G(\cdot)$ represent the survival functions of the survival time and the censoring time, respectively, and τ denotes the largest observed time.

Next, we use this result to show the following lemma.

Lemma 2. *For any partition Λ_j with Z_j, define*

$$K_j(\Lambda_j; l_1, l_2)=\sup_{0\le t\le\tau}\left|S_j(t|I_j=l_1)-S_j(t|I_j=l_2)\right|.$$

Then for any $\epsilon > 0$, we have

$$Pr\left(\left|\widehat{K}_j(\mathbf{\Lambda}_j; l_1, l_2) - K_j(\mathbf{\Lambda}_j; l_1, l_2)\right| > \epsilon\right) \le \eta_j(\epsilon),$$

where

$$\eta_j(\epsilon) = 2.5\Bigg[\exp\left\{-\frac{P_{l_{j1}} n\epsilon^2}{4} G^2(\tau) + \mu_1 \frac{\epsilon\sqrt{n}}{2} G(\tau)\right\} + \exp\left\{-\frac{c_1 n P_{l_{j1}}^2}{4} + \mu_1 \frac{\epsilon\sqrt{n}}{2} G(\tau)\right\}$$
$$+ \exp\left\{-\frac{P_{l_{j2}} n\epsilon^2}{4} G^2(\tau) + \mu_2 \frac{\epsilon\sqrt{n}}{2} G(\tau)\right\} + \exp\left\{-\frac{c_2 n P_{l_{j2}}^2}{4} + \mu_2 \frac{\epsilon\sqrt{n}}{2} G(\tau)\right\}\Bigg]$$

with $P_{jl} = Pr(I_j = l)$ for positive constants c_1, c_2, μ_1, μ_2.

Proof. Note that

$$\begin{aligned} &Pr\left(\left|\widehat{K}_j(\mathbf{\Lambda}_j; l_1, l_2) - K_j(\mathbf{\Lambda}_j; l_1, l_2)\right| > \epsilon\right) \\ \le\ &Pr\left(\sup_{0\le t\le\tau}\left|\widehat{S}_j(t|I_j = l_1) - S_j(t|I_j = l_1)\right| \ge \frac{\epsilon}{2}\right) \\ &+ Pr\left(\sup_{0\le t\le\tau}\left|\widehat{S}_j(t|I_j = l_2) - S_j(t|I_j = l_2)\right| \ge \frac{\epsilon}{2}\right), \end{aligned}$$

and by Lemma 1,

$$\begin{aligned} &Pr\left(\sup_{0\le t\le\tau}\left|\widehat{S}_j(t|I_j = l) - S_j(t|I_j = l)\right| \ge \frac{\epsilon}{2}|Z_j\right) \\ \le\ &Pr\left(\sqrt{n_{jl}}\sup_{0\le t\le\tau}\left|\left(\widehat{S}_j(t|I_j = l) - S_j(t|I_j = l)\right)G(t)\right| \ge \frac{\epsilon}{2}\sqrt{n_{jl}}G(\tau)|Z_j\right) \\ \le\ &2.5\exp\left\{-\frac{n_{jl}\epsilon^2 G^2(\tau)}{2} + \mu\frac{\sqrt{n_{jl}}\epsilon G(\tau)}{2}\right\}, \end{aligned}$$

where n_{jl} is a sum of n independent and identically distributed Bernoulli random variables with the probability of success being P_{jl}. Thus, by the Chernoff bound, we have

$$Pr\left(n_{jl} < \frac{nP_{jl}}{2}\right) < \exp\left(-\frac{cnP_{jl}^2}{4}\right)$$

which completes the proof of the lemma. □

To prove Theorem 1, we need more lemmas.

Lemma 3. *Consider p pairs of random variables (T, Z_j) for $j = 1, 2, \ldots, p$, and let $f_j(z)$ denote the probability density function of Z_j and $S_j(t|Z_j) = Pr(T > t|Z_j)$. For any interval $[a, b)$ such that $f_j(z) > 0$ for $z \in [a, b)$, we have*

$$\inf_{z\in[a,b)} S_j\big(t|Z_j = z\big) \le S_j\big(t|Z_j \in [a, b)\big) \le \sup_{z\in[a,b)} S_j\big(t|Z_j = z\big)$$

for all t.

This lemma is similar to Proposition 1 of Mai and Zou (2015) and thus the proof is omitted here. Also by Lemma 4 of Mai and Zou (2015), we have the following lemma.

Lemma 4. *For Z_j, we define a partition*

$$\mathbf{\Lambda}_j = \left\{[a_l^j, a_{l+1}^j) :\ a_l^j < a_{l+1}^j,\ l = 0, \ldots, \Lambda_j - 1\right\},$$

where a_l^j is the $\frac{l}{\Lambda_j}$th sample quantile of Z_j. Then we have

$$Pr\left\{Pr(a_l^j \le Z_j < a_{l+1}^j|\mathbf{\Lambda}_j) < \frac{2}{\Lambda_j}\right\} \ge 1 - c\exp\left(-c\frac{n}{\Lambda_j^2}\right).$$

Moreover, a direct application of Lemma 2 to the special partition formed from the oracle uniform slicing yields the following lemma.

Lemma 5. *Under the conditions in Theorem 1, for any $\epsilon > 0$, we have*

$$Pr\left(|\widehat{K}_j^{(o)} - K_j^{(o)}| \geq N_j\epsilon\right) \leq cN(\log^2 n)\exp\left(-c\frac{n\epsilon^2}{\log n} + c\epsilon\sqrt{n}\right) + cN(\log^2 n)\exp\left(-c\frac{n}{\log^2 n} + c\epsilon\sqrt{n}\right).$$

Lemma 6. *Under the conditions in Theorem 1, for any $\epsilon > 0$, we have*

$$Pr\left(\left|\widehat{K}_j - K_j\right| \geq N_j\epsilon\right) \leq cN(\log^2 n)\exp\left(-c\frac{n\epsilon^2}{\log n} + c\epsilon\sqrt{n}\right).$$

Proof. The lemma can be shown by using Lemma 1 and the similar arguments as those used in the proof of Lemma 5 of Mai and Zou (2015), and thus the detailed proof is omitted here. □

Lemma 7. *Under the conditions in Theorem 1, we have*

$$Pr\left(\left|K_j - K_j^{(o)}\right| \geq N_j\Delta_{\mathcal{B}}/4\right) \leq cN\exp\left(-c\frac{n}{\log^2 n}\right),$$

where $\Delta_{\mathcal{B}}$ is defined in condition C1.

Proof. This lemma can be proved by using Lemmas 3 and 4 with condition 2 and the similar arguments used in the proof of Lemma 6 of Mai and Zou (2015), and so the detailed proof is omitted here. □

Proof of Theorem 1. In order to prove part (1), we first show that if

$$\left|\widehat{K}_j^{(o)} - K_j^{(o)}\right| < \frac{N_j\Delta_{\mathcal{B}}}{4}$$

for all j, then $\mathcal{A} \subset \widehat{\mathcal{A}}(oracle)$. Straightforward calculations entail that

$$-\frac{N_j\Delta_{\mathcal{B}}}{4} < \widehat{K}_j^{(o)} - K_j^{(o)} < \frac{N_j\Delta_{\mathcal{B}}}{4}.$$

For any $j \in \mathcal{B}$, we obtain that

$$\widehat{K}_j^{(o)} > K_j^{(o)} - \frac{N\Delta_{\mathcal{B}}}{4}.$$

By the definition of $\Delta_{\mathcal{B}}$, for any k, we have

$$\min_{j\in\mathcal{B}} K_j^{(o)}(\Lambda_{kj}) - \max_{j\notin\mathcal{B}} K_j^{(o)}(\Lambda_{kj}) \geq \Delta_{\mathcal{B}},$$

and so

$$K_j^{(o)} = \sum_{k=1}^{N_j} K_j^{(o)}(\Lambda_{kj}) \geq N_j\Delta_{\mathcal{B}} + \max_{j\notin\mathcal{B}} K_j^{(o)}.$$

Therefore

$$\widehat{K}_j^{(o)} > K_j^{(o)} - \frac{N_j\Delta_{\mathcal{B}}}{4} \geq \max_{j\notin\mathcal{B}} K_j^{(o)} + \frac{3N_j\Delta_{\mathcal{B}}}{4} > \max_{j\notin\mathcal{B}} K_j^{(o)} + \frac{N_j\Delta_{\mathcal{B}}}{4}.$$

On the other hand,

$$\widehat{K}_j^{(o)} < K_j^{(o)} + \frac{N_j\Delta_{\mathcal{B}}}{4} \leq \max_{j\notin\mathcal{B}} K_j^{(o)} + \frac{N_j\Delta_{\mathcal{B}}}{4}, \quad \text{for } j \notin \mathcal{B}.$$

Then we can conclude that

$$\max_{j\notin\mathcal{B}} \widehat{K}_j^{(o)} < \min_{j\in\mathcal{B}} \widehat{K}_j^{(o)}.$$

Because of $d_n \geq |\mathcal{B}|$, then we can obtain that $\mathcal{B} \subset \widehat{\mathcal{A}}(oracle)$ and so $\mathcal{A} \subset \widehat{\mathcal{A}}(oracle)$. By Lemma 5, we have

$$\begin{aligned} Pr(\mathcal{A} \subset \widehat{\mathcal{A}}(oracle)) &\geq Pr\left\{\bigcap_{j=1}^{p}\left(\left|\widehat{K}_j^{(o)} - K_j^{(o)}\right| < \frac{N\Delta_{\mathcal{B}}}{4}\right)\right\} \\ &= 1 - Pr\left\{\bigcup_{j=1}^{p}\left(\left|\widehat{K}_j^{(o)} - K_j^{(o)}\right| \geq \frac{N\Delta_{\mathcal{B}}}{4}\right)\right\} \\ &\geq 1 - \sum_{j=1}^{p} Pr\left(\left|\widehat{K}_j^{(o)} - K_j^{(o)}\right| \geq \frac{N\Delta_{\mathcal{B}}}{4}\right) \\ &\geq 1 - \eta_1, \end{aligned}$$

where η_1 is as defined in Theorem 1.

By a similar argument as above, we can show that if

$$\left|\widehat{K}_j - K_j^{(o)}\right| < \frac{N_j\Delta_{\mathcal{B}}}{4}$$

for all j, then $\mathcal{A} \subset \widehat{\mathcal{A}}(d_n)$. Combining Lemmas 6 and 7 yields that

$$\begin{aligned} &Pr\left(\left|\widehat{K}_j - K_j^{(o)}\right| \geq \frac{N_j\Delta_{\mathcal{B}}}{4}\right) \\ &\quad \leq cN(\log^2 n)\exp\left(-c\frac{n\Delta_{\mathcal{B}}^2}{\log n} + c\Delta_{\mathcal{B}}\sqrt{n}\right) + cN\exp\left(-c\frac{n}{\log^2 n}\right). \end{aligned}$$

Then we have

$$\begin{aligned} Pr\left(\mathcal{A} \subset \widehat{\mathcal{A}}(d_n)\right) &\geq Pr\left\{\bigcap_{j=1}^{p}\left(\left|\widehat{K}_j - K_j^{(o)}\right| < \frac{N\Delta_{\mathcal{B}}}{4}\right)\right\} \\ &= 1 - Pr\left\{\bigcup_{j=1}^{p}\left(\left|\widehat{K}_j - K_j^{(o)}\right| \geq \frac{N\Delta_{\mathcal{B}}}{4}\right)\right\} \\ &\geq 1 - \sum_{j=1}^{p} Pr\left(\left|\widehat{K}_j - K_j^{(o)}\right| \geq \frac{N\Delta_{\mathcal{B}}}{4}\right) \\ &\geq 1 - \eta_2, \end{aligned}$$

where η_2 is given in Theorem 1. It is easy to see that when $n \to \infty$, $\eta_1 \to 0$ and $\eta_2 \to 0$. Therefore both the oracle fused KS-SIS and the fused KS-SIS enjoy the sure screening property with probability going to 1.

References

Bitouzé, D., Laurent, B., Massart, P., 1999. A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator. Ann. Inst. Henri Poincare (B) Probab. Stat. 35, 735–763. http://dx.doi.org/10.1016/S0246-0203(99)00112-0.

Candes, E., Tao, T., 2007. The Dantzig selector: statistical estimation when p is much larger than n. Ann. Statist. 35, 2313–2351. http://dx.doi.org/10.1214/009053606000001523.

Cook, R.D., Zhang, X., 2014. Fused estimators of the central subspace in sufficient dimension reduction. J. Amer. Statist. Assoc. 109, 815–827. http://dx.doi.org/10.1080/01621459.2013.866563.

Fan, J., Feng, Y., Song, R., 2011. Nonparametric independence screening in sparse ultra-high-dimensional additive models. J. Amer. Statist. Assoc. 106, 544–557. http://dx.doi.org/10.1198/jasa.2011.tm09779.

Fan, J., Feng, Y., Wu, Y., 2010. High-dimensional variable selection for Cox's proportional hazards model. In: Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown, Vol. 6. Institute of Mathematical Statistics, pp. 70–86. http://dx.doi.org/10.1214/10-IMSCOLL606.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96, 1348–1360. http://dx.doi.org/10.1198/016214501753382273.

Fan, J., Lv, J., 2008. Sure independence screening for ultrahigh dimensional feature space. J. R. Stat. Soc. Ser. B Stat. Methodol. 70, 849–911. http://dx.doi.org/10.1111/j.1467-9868.2008.00674.x.

Fan, J., Samworth, R., Wu, Y., 2009. Ultrahigh dimensional feature selection: beyond the linear model. J. Mach. Learn. Res. 10, 2013–2038.

Fan, J., Song, R., 2010. Sure independence screening in generalized linear models with NP-dimensionality. J. Amer. Statist. Assoc. 38, 3567–3604. http://dx.doi.org/10.1214/10-AOS798.

Gorst-Rasmussen, A., Scheike, T., 2013. Independent screening for single-index hazard rate models with ultrahigh dimensional features. J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 217–245. http://dx.doi.org/10.1111/j.1467-9868.2012.01039.x.

He, X., Wang, L., Hong, H.G., 2013. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. Ann. Statist. 41, 342–369. http://dx.doi.org/10.1214/13-AOS1087.

Li, G., Peng, H., Zhang, J., Zhu, L., 2012a. Robust rank correlation based screening. Ann. Statist. 40, 1846–1877. http://dx.doi.org/10.1214/12-AOS1024.

Li, R., Zhong, W., Zhu, L., 2012b. Feature screening via distance correlation learning. J. Amer. Statist. Assoc. 107, 1129–1139. http://dx.doi.org/10.1080/01621459.2012.695654.
Mai, Q., Zou, H., 2013. The Kolmogorov filter for variable screening in high-dimensional binary classification. Biometrika 100, 229–234. http://dx.doi.org/10.1093/biomet/ass062.
Mai, Q., Zou, H., 2015. The fused Kolmogorov filter: A nonparametric model-free screening method. Ann. Statist. 43, 1471–1497. http://dx.doi.org/10.1214/14-AOS1303.
Rosenwald, A., Wright, G., Wiestner, A., Chan, W.C., et al., 2003. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell 3, 185–197. http://dx.doi.org/10.1016/S1535-6108(03)00028-X.
Song, R., Lu, W., Ma, S., Jeng, X.J., 2014. Censored rank independence screening for high-dimensional survival data. Biometrika 101, 799–814. http://dx.doi.org/10.1093/biomet/asu047.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58, 267–288.
Tibshirani, R., 2009. Univariate shrinkage in the Cox model for high dimensional data. Stat. Appl. Genet. Mol. Biol. 8, 1–18. http://dx.doi.org/10.2202/1544-6115.1438.
Uno, H., Cai, T., Pencina, M.J., D'Agostino, R.B., Wei, L.J., 2011. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat. Med. 30, 1105–1117. http://dx.doi.org/10.1002/sim.4154.
Van Houwelingen, H.C., Bruinsma, T., Hart, A.A.M., Van'T Veer, L.J., Wessels, L.F.A., 2006. Cross-validated Cox regression on microarray gene expression data. Stat. Med. 25, 3201–3216. http://dx.doi.org/10.1002/sim.2353.
Wu, Y., Yin, G., 2015. Conditional quantile screening in ultrahigh-dimensional heterogeneous data. Biometrika 102, 65–76. http://dx.doi.org/10.1093/biomet/asu068.
Zhang, C.H., 2010. Nearly unbiased variable selection under minimax concave penalty. Ann. Statist. 38, 894–942. http://dx.doi.org/10.1214/09-AOS729.
Zhao, S.D., Li, Y., 2012. Principled sure independence screening for Cox models with ultra-high-dimensional covariates. J. Multivariate Anal. 105, 397–411. http://dx.doi.org/10.1016/j.jmva.2011.08.002.
Zhu, L.P., Li, L., Li, R., Zhu, L.X., 2011. Model-free feature screening for ultrahigh dimensional data. J. Amer. Statist. Assoc. 106, 1464–1475. http://dx.doi.org/10.1198/jasa.2011.tm10563.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429. http://dx.doi.org/10.1198/016214506000000735.