Computational Statistics and Data Analysis 108 (2017) 121–132

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Correlation rank screening for ultrahigh-dimensional survival data

Jing Zhang, Yanyan Liu, Yuanshan Wu *

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China

ARTICLE INFO

Article history:
Received 17 November 2015
Received in revised form 24 August 2016
Accepted 6 November 2016
Available online 15 November 2016

Keywords:
Censored data
Correlation rank
Model-free screening
Sure independent screening
Survival data
Ultrahigh-dimensional data

ABSTRACT

With the recent explosion of ultrahigh-dimensional data, extensive work has been carried out for screening methods which can effectively reduce the dimensionality. However, censored survival data which often arise in clinical trials and genetic studies have been left greatly unexplored for ultrahigh-dimensional scenarios. A novel feature screening procedure is proposed for ultrahigh-dimensional survival data. Also established are the ranking consistency and the sure independent screening properties. Compared with the existing methods, the proposed screening procedure is invariant to the monotone transformation, known or unknown, of the response. Moreover, it can be readily applied to ultrahigh-dimensional complete data when the censoring rate is zero. Simulation studies demonstrate that the proposed procedure exhibits favorably in comparisons with the existing ones. As an illustration, the proposed method is applied to the mantle cell lymphoma study.

1. Introduction

With the rapid advance of technology, ultrahigh-dimensional data could be collected at a relatively low cost and have appeared in various fields such as genomics, imaging and economics. Because the dimensionality p_n increases very rapidly with sample size n, existing penalized variable selection methods, such as the least absolute shrinkage and selection operator (LASSO, Tibshirani, 1996), the smoothly clipped absolute deviation (SCAD, Fan and Li, 2001), the adaptive LASSO (Zou, 2006), the Dantzig selector (Candes and Tao, 2007) and the minimax concave penalty (MCP, Zhang, 2010) may not perform well (Fan et al., 2009).

To overcome ultrahigh dimensionality, Fan and Lv (2008) proposed a sure independence screening (SIS) method to reduce the dimension in the context of linear regression models, so that penalized variable selection methods are applicable. Such screening procedures have been extensively studied in various ultrahigh-dimensional contexts, such as generalized linear models (Fan and Song, 2010) and additive models (Fan et al., 2011). Furthermore, in order to avoid the specification of a particular model structure, Zhu et al. (2011) proposed a sure independent ranking and screening (SIRS) procedure for ultrahigh-dimensional data in the framework of the general multi-index models. Li et al. (2012b) proposed a model-free SIS procedure based on the distance correlation. Using the Kendall τ, Li et al. (2012a) proposed a robust screening procedure in the framework of the transformation models.

For censored ultrahigh-dimensional data, Fan et al. (2010) investigated the SIS method for the Cox proportional hazards model via ranking variables according to their respective univariate partial log-likelihoods. Zhao and Li (2012) proposed a

* Corresponding author.
E-mail address: shan@whu.edu.cn (Y. Wu).

http://dx.doi.org/10.1016/j.csda.2016.11.005
0167-9473/

screening method based on the standardized marginal maximum partial likelihood estimators of the Cox model, and they also provided theoretical justification for the sure independent screening property. To relax the Cox model assumption, Gorst-Rasmussen and Scheike (2013) proposed a screening procedure for a general class of single-index hazard rate models. Based on Kendall's τ and via the inverse-probability-of-censoring weighting, Song et al. (2014) proposed a censored rank independence screening method which is shown to be robust against the potential outliers and to work for a general class of survival models. Wu and Yin (2015) developed a screening method which is designed to identify the covariates that contribute to the conditional quantile of the response. Recently, Zhou and Zhu (in press) proposed a censored version of the SIRS method by incorporating the weight of the inverse probability of censoring.

In a model-free fashion, we propose a novel correlation rank sure independent screening procedure (CR-SIS), which can naturally handle ultrahigh-dimensional survival data without any nonparametric approximation except for the Kaplan–Meier estimator. Compared with the existing procedures, our approach enjoys several distinctive advantages. Our procedure does not rely on any model assumption and works generally for nonlinear survival regression models. On the other hand, our approach is invariant under the monotone transformation of the response. These advantages greatly facilitate the implementation of the proposed method in real applications.

The rest of the article is organized as follows. In Section 2, we propose the CR-SIS procedure for both ultrahigh-dimensional complete and censored data. In Section 3, we establish the theoretical properties of the proposed procedure. Its finite-sample performances are evaluated in Section 4 via extensive simulation studies. In Section 5, we apply the proposed method to a recent study on mantle cell lymphoma. Section 6 concludes some remarks. All technical proofs are presented in the Appendix.

2. Screening procedures

Consider the conditional distribution function,

$$F(y|\mathbf{Z}) = P(Y \le y|\mathbf{Z}),$$

where Y denote the response variable and $\mathbf{Z} = (Z_1, \ldots, Z_{p_n})^{\mathrm{T}}$ the covariate vector. In an ultrahigh-dimensional setting, the dimensionality p_n, possibly depending on and greatly exceeding the sample size n, might increase with n at an exponential rate. To identify which covariates among the p_n ones contribute to the conditional distribution function of Y given $\mathbf{Z}$, we define the active covariate set as

$$\mathcal{A} = \{k : F(y|\mathbf{Z}) \text{ depends on } Z_k,\ k = 1, \ldots, p_n\}.$$

Without loss of generality, we assume throughout this article that $E(Z_k) = 0$ for $k = 1, \ldots, p_n$. Let $G(y) = P(Y \le y)$ denotes the unconditional distribution function of Y. Define $\mathbf{R}(Y) = E\{\mathbf{Z}G(Y)\}$, let $R_k(Y)$ be the kth element of $\mathbf{R}(Y)$, then $R_k(Y) = E\{Z_kG(Y)\} = \mathrm{cov}\{Z_k, G(Y)\}$, where Z_k denotes the kth element of $\mathbf{Z}$. Define

$$r_k = [R_k(Y)]^2,$$

where $k = 1, \ldots, p_n$, then r_k serves as the population version of our proposed marginal utility measure for the kth covariate.

Intuitively, the unconditional distribution function of Y, $G(y)$, compositing with Y, is expected to contain the whole information of Y. Consequently, r_k, which measures the correlation between $G(Y)$ and Z_k, could reflect the relationship between Y and Z_k. If Y and Z_k are independent, $G(Y)$ and Z_k should be independent; hence $r_k = 0$. On the other hand, it is reasonable to expect $r_k > 0$ if Y and Z_k are dependent. Under the framework of semiparametric regression, Zhu and Zhu (2009) proposed a distribution-weighted least squares estimator which can be deduced from the variant of $\mathrm{cov}\{Z_k, G(Y)\}$. Our proposed marginal utility r_k shares the spirit of their method. The SIRS method proposed by Zhu et al. (2011) adopted the dichotomous $I(Y < y)$ variable, while we use $G(y)$, which is continuous and thus expected to contain the whole information of Y. The correlation between $G(Y)$ and Z_k could be consequently appropriate to reflect the relationship between Y and Z_k. Furthermore, our method can naturally handle ultrahigh-dimensional survival data without any nonparametric approximation except for the routine Kaplan–Meier estimator. These remarkable properties motivate us to use r_k for feature screening in ultrahigh-dimensional data. We can see in the sequel that the proposed method indeed enjoys the ranking consistency property and also performs well in different scenarios.

Given a random sample $\{Y_i, \mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{ip_n})^{\mathrm{T}}\}, i = 1, \ldots, n$, from the population $\{Y, \mathbf{Z} = (Z_1, \ldots, Z_{p_n})^{\mathrm{T}}\}$. It is desirable to derive an estimator of r_k based on the n independent and identical observations. For ease of presentation, we assume that the sample predictors are all centralized, that is, $n^{-1}\sum_{i=1}^{n} Z_{ik} = 0$ for $k = 1, \ldots, p_n$, where Z_{ik} is the kth element of $\mathbf{Z}_i$. Obviously, we can use the empirical distribution function, which is given by

$$\widehat{G}_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le y),$$

to estimate $G(y)$. Therefore, we propose an estimator for r_k, which takes the form of

$$\widehat{r}_k = \left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widehat{G}_n(Y_i)\right\}^2.$$

Intuitively, we can see that, if Z_k and Y are independent, then $\widehat{r}_k$ is expected to fluctuate around zero. On the contrary, those predictors with a large value of $\widehat{r}_k$ are considered important. As a result, we define the estimated active set as

$$\widehat{\mathcal{A}} = \{k : \widehat{r}_k \geq cn^{-\alpha},\ k = 1, \ldots, p_n\},$$

where constants c and α are specified in the regularity conditions of Section 3.

We extend the screening procedure to ultrahigh-dimensional survival data while taking the censoring into account. Suppose that we observe the data $\{(X_i, \Delta_i, \mathbf{Z}_i) : i = 1, \ldots, n\}$, independent copies from $(X, \Delta, \mathbf{Z})$, where $X = \min(Y, C)$, $\Delta = I(Y \leq C)$, with C denoting the censoring variable. For ease of exposition, we assume that the censoring mechanism is completely random, i.e., the censoring variable C is independent of response variable Y and covariate $\mathbf{Z}$.

In the scenario of censored response, the estimator for the cumulative distribution function of Y, $G(y)$, can be deduced from the Kaplan–Meier estimator, which is given by

$$\widetilde{G}_n(y) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{1}{\sum_{j=1}^{n} I(X_j \geq X_i)} \right\}^{\Delta_i I(X_i \leq y)}.$$

Accordingly, the estimator for r_k is obtained as follows,

$$\widetilde{r}_k = \left\{ \frac{1}{n} \sum_{i=1}^{n} Z_{ik} \widetilde{G}_n(X_i) \right\}^2.$$

We propose to rank the $\widetilde{r}_k$ from the largest to smallest and select the top ones as the active predictors via defining estimated active set,

$$\widetilde{\mathcal{A}} = \{k : \widetilde{r}_k \geq cn^{-\alpha},\ k = 1, \ldots, p_n\}.$$

3. Theoretical properties

We show that the CR-SIS procedure possesses sure independent screening and ranking consistency properties for the censored response case. For the complete response case, these properties can be considered as a trivial extension. We impose the following regularity conditions throughout our discussion.

C1. There exist constants $\delta > 0$ and $\tau > 0$ such that

$$P(C \geq \tau) = P(C = \tau) \geq \delta.$$

C2. There exists a positive constant ξ such that

$$\max_{1 \leq k \leq p_n} E(Z_k^2) < \xi.$$

C3. It holds that

$$\min_{k \in \mathcal{A}} r_k \geq 2cn^{-\alpha}$$

for some constants $c > 0$ and $\alpha \in [0, 1/2)$.

Condition C1 is a common assumption in survival analysis with τ being the end time of the study. It means that at least some subjects do not fail at the end time τ and by definition they are considered to be right-censored at τ. Condition C2 is regarding to the second-order moments of predictors and it holds for a large variety of distributions. Condition C3 requires that the marginal utilities carrying information for the active predictors should not decay too fast. We state the sure independent screening property of the CR-SIS procedure for ultrahigh-dimensional survival data.

Theorem 1. *Under conditions* C1 *and* C2*, there exists a constant* $\eta > 0$ *such that*

$$P\Big(\max_{1 \leq k \leq p_n} |\widetilde{r}_k - r_k| \geq cn^{-\alpha} \Big) \leq O\Big[p_n \exp\Big\{ -\eta \Big(\frac{n^{1-2\alpha}}{\log \log n} \Big)^{1/2} \Big\} \Big].$$

Under conditions C1, C2 *and* C3 *, it holds that*

$$P\Big(\mathcal{A} \subseteq \widetilde{\mathcal{A}} \Big) \geq 1 - O\Big[a_n \exp\Big\{ -\eta \Big(\frac{n^{1-2\alpha}}{\log \log n} \Big)^{1/2} \Big\} \Big],$$

where $a_n = |\mathcal{A}|$ *is the cardinality of* $\mathcal{A}$.

Denote $\mathbf{Z}_{\mathcal{A}} = \{Z_j : j \in \mathcal{A}\}$ and $\mathbf{Z}_{\mathcal{A}^c} = \{Z_j : j \notin \mathcal{A}\}$. The ranking consistency of the CR-SIS procedure is summarized as follows.

Theorem 2. *Assume that* (i) Y *and* $\mathbf{Z}_{\mathcal{A}^c}$ *are conditionally independent given* $\mathbf{Z}_{\mathcal{A}}$ *and* (ii) $\mathbf{Z}_{\mathcal{A}}$ *is independent of* $\mathbf{Z}_{\mathcal{A}^c}$. *Under condition* C3, *we have that*

$$\max_{k\notin\mathcal{A}} r_k < \min_{k\in\mathcal{A}} r_k,$$

and $r_k = 0$ *if and only if* $k \notin \mathcal{A}$. *Furthermore, under conditions* C1, C2 *and* C3, *there exists a constant* $\eta > 0$ *such that*

$$P\Big(\max_{k\notin\mathcal{A}} \widetilde{r}_k < \min_{k\in\mathcal{A}} \widetilde{r}_k\Big) \geq 1 - O\Big[p_n \exp\Big\{-\eta\Big(\frac{n^{1-2\alpha}}{\log\log n}\Big)^{1/2}\Big\}\Big].$$

This lays out the theoretical foundation that our CR-SIS procedure tends to rank the active predictors above the inactive ones with high probability.

4. Simulation studies

We examine the finite-sample performance of the proposed method and make comparisons with existing methods via simulation studies. For brevity, we refer to the sure independence screening method proposed by Fan and Lv (2008) as SIS, the sure independent ranking and screening procedure of Zhu et al. (2011) as SIRS, the distance correlation screening procedure of Li et al. (2012b) as DC-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 censored sure independent ranking and screening of Zhou and Zhu (in press) as CSIRS.

Example 1. We first considered the performance of the CR-SIS procedure for complete data and compared it with the SIS, SIRS and DC-SIS methods. The simulation setups are the same as Example 1 of Zhu et al. (2011). It is a classical linear model with varying squared multiple correlation coefficient R^2 and error distribution:

$$Y = c\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \sigma\epsilon,$$

where $\boldsymbol{\beta} = (1, 0.8, 0.6, 0.4, 0.2, 0, \ldots, 0)^{\mathrm{T}}$, i.e., only the first five predictors are active. The ultrahigh-dimensional covariate $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip_n})$ follows a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{\Sigma} = (0.8^{|i-j|})$ for $i, j = 1, \ldots, p_n$. We set $\sigma = 6.83$ and considered two error ϵ distributions, a standard normal $N(0, 1)$ and a t-distribution with one degree of freedom that has a heavy tail. We varied the constant c in front of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}$ to control the signal-to-noise ratio. We chose $c = 0.5$, 1 and 2, with the corresponding $R^2 = 20\%$, 50% and 80%. Set sample size $n = 200$, coupled with the number of covariates $p_n = 2000$. For each configuration, we repeated 500 simulations.

To assess the performance of the screening procedures, we employed three evaluation criteria (Li et al., 2012b). First, we compare the minimum model size, denoted by $\mathcal{S}$, which includes all the active predictors. Obviously, $\mathcal{S}$ can be used to measure the resulting model complexity for each screening procedure. The closer to the true minimum model size, the better the screening procedure. We present the 5%, 25%, 50%, 75% and 95% quantiles of $\mathcal{S}$ out of 500 replications. Second, for each individual active predictor, we report its selection proportion, denoted by $\mathcal{P}_e$, for a given model size among the 500 replications. Third, we exhibit the proportion that all active predictors are selected for a given model size in the 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 estimated model size to be $d = \lceil n/\log n\rceil$, where $\lceil x\rceil$ denotes the integer part of x.

The simulation results for $\mathcal{S}$, $\mathcal{P}_e$ and $\mathcal{P}_a$ are summarized in Tables 1 and 2. We can see that the proposed CR-SIS procedure is comparable to the SIS method for the normal error. However, it is consistently superior for the heavy-tailed error distribution even Condition C2 is violated. Compared with the SIRS method, the performances of the CR-SIS procedure are equally good in all the considered scenarios; both of them deliver more satisfactory results than the DC-SIS procedure.

Example 2. We further considered the performance of the CR-SIS procedure for ultrahigh-dimensional censored data and compared it with the existing approaches, including the CRIS, FAST-SIS and P-SIS procedures. We generated survival times T_i, associated with covariate $\mathbf{Z}_i$, from the Cox proportional hazards regression model

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

where the baseline hazard function was set to be $\lambda_0(t) = (t - 0.5)^2$ and the ultrahigh-dimensional covariate $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip_n})$ was generated in the same way as that in Example 1. 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 τ was the study duration time, chosen to yield a censoring rate of 20%. We took the sample size $n = 50$, 100 and 200, coupled with $p_n = 2000$. For each configuration, we repeated 500 simulations.

Table 1
Five quantiles of the minimum model size $\mathcal{S}$ among 500 replications in Example 1 with the true model size $p_0 = 5$.

ϵ	c	Method	5%	25%	50%	75%	95%
$N(0, 1)$	0.5	CR-SIS	5	5	5	7	28
		SIS	5	5	5	6	16
		DC-SIS	5	5	5	7	35
		SIRS	5	5	5	7	36
	1	CR-SIS	5	5	5	5	5
		SIS	5	5	5	5	5
		DC-SIS	5	5	5	5	5
		SIRS	5	5	5	5	5
	2	CR-SIS	5	5	5	5	5
		SIS	5	5	5	5	5
		DC-SIS	5	5	5	5	5
		SIRS	5	5	5	5	5
$t(1)$	0.5	CR-SIS	6	18	81	260	1127
		SIS	224	741	1233	1668	1925
		DC-SIS	36	245	621	1051	1650
		SIRS	5	11	50	179	884
	1	CR-SIS	5	5	5	7	44
		SIS	80	443	960	1485	1881
		DC-SIS	5	7	34	190	746
		SIRS	5	5	5	6	26
	2	CR-SIS	5	5	5	5	5
		SIS	7	107	549	1119	1802
		DC-SIS	5	5	5	6	187
		SIRS	5	5	5	5	5

Table 2
Selection proportions $\mathcal{P}_e$ for each active predictor and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 1.

ϵ	c	Method	$\mathcal{P}_e$					$\mathcal{P}_a$
			X_1	X_2	X_3	X_4	X_5	
$N(0, 1)$	0.5	CR-SIS	1.000	1.000	0.998	0.998	0.964	0.960
		SIS	1.000	1.000	1.000	0.998	0.984	0.982
		DC-SIS	1.000	1.000	1.000	1.000	0.954	0.954
		SIRS	0.998	1.000	0.998	0.998	0.954	0.950
	1	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		SIS	1.000	1.000	1.000	1.000	1.000	1.000
		DC-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		SIRS	1.000	1.000	1.000	1.000	1.000	1.000
	2	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		SIS	1.000	1.000	1.000	1.000	1.000	1.000
		DC-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		SIRS	1.000	1.000	1.000	1.000	1.000	1.000
$t(1)$	0.5	CR-SIS	0.772	0.836	0.800	0.648	0.470	0.372
		SIS	0.052	0.056	0.050	0.046	0.040	0.002
		DC-SIS	0.278	0.360	0.288	0.222	0.138	0.052
		SIRS	0.848	0.884	0.858	0.708	0.530	0.446
	1	CR-SIS	0.994	0.998	0.996	0.980	0.944	0.940
		SIS	0.138	0.158	0.126	0.116	0.070	0.024
		DC-SIS	0.826	0.854	0.826	0.742	0.538	0.514
		SIRS	1.000	1.000	0.998	0.990	0.960	0.960
	2	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		SIS	0.356	0.392	0.362	0.286	0.242	0.160
		DC-SIS	0.946	0.956	0.954	0.938	0.906	0.896
		SIRS	1.000	1.000	1.000	1.000	1.000	1.000

We presented the simulation results for $\mathcal{S}$, $\mathcal{P}_e$ and $\mathcal{P}_a$ in Tables 3 and 4. In general, the performances of the FAST-SIS and P-SIS procedures are comparable. Both of them outperform the CR-SIS and CRIS procedures when the sample size is 50. This is mainly due to that the FAST-SIS and P-SIS procedures are carried out via effectively utilizing the preassumed Cox proportional hazards regression model structure, although it is always unjustifiable in practice. On the contrary, the model-free CR-SIS and CRIS procedures obviously do not fully use the information of the model structure. Nevertheless, when the sample size n is increased to 100, four screening methods perform equally well.

Table 3
Five quantiles of the minimum model size $\mathcal{S}$ among 500 replications in Example 2 with the true model size $p_0 = 5$.

n	Method	5%	25%	50%	75%	95%
50	CR-SIS	5	6	10	35	283
	CRIS	5	19	76	238	1022
	FAST-SIS	5	5	6	10	62
	P-SIS	5	5	6	8	39
100	CR-SIS	5	5	5	6	11
	CRIS	5	5	6	13	105
	FAST-SIS	5	5	5	5	6
	P-SIS	5	5	5	5	6
200	CR-SIS	5	5	5	5	6
	CRIS	5	5	5	5	7
	FAST-SIS	5	5	5	5	6
	P-SIS	5	5	5	5	5

Table 4
Selection proportions $\mathcal{P}_e$ for each active predictor and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 2.

n	Method	$\mathcal{P}_e$					$\mathcal{P}_a$
		X_1	X_2	X_3	X_4	X_5	
50	CR-SIS	0.730	0.876	0.914	0.876	0.736	0.544
	CRIS	0.410	0.638	0.668	0.578	0.408	0.162
	FAST-SIS	0.876	0.968	0.990	0.984	0.916	0.800
	P-SIS	0.908	0.988	0.998	0.998	0.938	0.854
100	CR-SIS	0.988	0.998	1.000	1.000	0.984	0.972
	CRIS	0.888	0.974	0.980	0.962	0.886	0.810
	FAST-SIS	0.998	1.000	1.000	1.000	1.000	0.998
	P-SIS	0.998	1.000	1.000	1.000	1.000	0.998
200	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
	CRIS	1.000	1.000	1.000	1.000	0.998	0.998
	FAST-SIS	1.000	1.000	1.000	1.000	1.000	1.000
	P-SIS	1.000	1.000	1.000	1.000	1.000	1.000

Table 5
Five quantiles of the minimum model size $\mathcal{S}$ among 500 replications in Example 3 with the true model size $p_0 = 5$.

n	Method	5%	25%	50%	75%	95%
50	CR-SIS	11	79	270	825	1867
	CRIS	55	494	1309	1794	1972
	FAST-SIS	120	703	1405	1776	1972
	P-SIS	114	705	1395	1788	1973
100	CR-SIS	5	10	43	187	975
	CRIS	21	223	801	1617	1966
	FAST-SIS	81	442	989	1703	1960
	P-SIS	83	448	1016	1712	1959
200	CR-SIS	5	5	6	9	57
	CRIS	6	48	286	941	1872
	FAST-SIS	17	152	496	1163	1880
	P-SIS	19	153	502	1165	1875

Example 3. To examine the performance of the screening procedures for the generally nonlinear survival models, we independently generated the survival times T_i from the model

$$\log T_i = Z_{i1}^2 + (2 + \sin Z_{i2})^2 + (1 + Z_{i3})^{-3} + (Z_{i4}^2 + Z_{i4} - 1)^{-1} + Z_{i5} + \epsilon_i,$$

where the error ϵ_i was generated from the standard normal distribution. The remaining setups were kept the same as those in Example 2. The corresponding results are summarized in Tables 5 and 6, from which we can see the CR-SIS method is able to capture the nonlinear covariate effects and thus produces acceptable screening results. Obviously, the CR-SIS method outperforms the other three screening procedures, especially overwhelmingly superior to the model-dependent FAST-SIS and P-SIS procedures, in terms of either the minimum model size required to cover all the active covariates or the proportion that all active predictors are selected for a given model size.

Table 6
Selection proportions $\mathcal{P}_e$ for each active predictor and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 3.

n	Method	$\mathcal{P}_e$					$\mathcal{P}_a$
		X_1	X_2	X_3	X_4	X_5	
50	CR-SIS	0.272	0.462	0.264	0.268	0.196	0.062
	CRIS	0.102	0.214	0.038	0.072	0.058	0.006
	FAST-SIS	0.074	0.100	0.026	0.036	0.036	0.010
	P-SIS	0.064	0.112	0.024	0.036	0.050	0.008
100	CR-SIS	0.662	0.842	0.642	0.678	0.608	0.368
	CRIS	0.284	0.460	0.098	0.212	0.226	0.052
	FAST-SIS	0.168	0.204	0.042	0.096	0.104	0.014
	P-SIS	0.152	0.202	0.050	0.106	0.116	0.014
200	CR-SIS	0.972	0.994	0.974	0.978	0.964	0.922
	CRIS	0.626	0.820	0.250	0.516	0.526	0.214
	FAST-SIS	0.484	0.568	0.132	0.318	0.378	0.102
	P-SIS	0.440	0.548	0.150	0.334	0.376	0.104

Table 7
Selection proportions $\mathcal{P}_e$ for each active predictor and the inactive predictor Z_6, and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 3.

$\mathbf{Z}_{\mathcal{A}} \perp \mathbf{Z}_{\mathcal{A}^c}$	Method	$\mathcal{P}_e$						$\mathcal{P}_a$
		X_1	X_2	X_3	X_4	X_5	X_6	
No	CR-SIS	0.662	0.842	0.642	0.678	0.608	0.354	0.368
	CRIS	0.284	0.460	0.098	0.212	0.226	0.130	0.052
	FAST-SIS	0.168	0.204	0.042	0.096	0.104	0.080	0.014
	P-SIS	0.152	0.202	0.050	0.106	0.116	0.090	0.014
Yes	CR-SIS	0.676	0.876	0.656	0.684	0.612	0.008	0.396
	CRIS	0.290	0.474	0.098	0.224	0.182	0.006	0.042
	FAST-SIS	0.160	0.208	0.050	0.100	0.112	0.014	0.016
	P-SIS	0.136	0.226	0.058	0.104	0.124	0.012	0.016

Till now the covariance matrix for generating the high-dimensional covariate $\mathbf{Z}$ is assumed to follow the Toeplitz structure, which implies that in Theorem 2 the assumption of $\mathbf{Z}_{\mathcal{A}}$ being independent of $\mathbf{Z}_{\mathcal{A}^c}$ does not hold. Nevertheless, the proposed method delivers favorable results over the existing ones in variant scenarios. Furthermore, we consider another mechanism of generating the covariates to guarantee the independence assumption holds. In particular, the first five covariates $(Z_{i1}, \ldots, Z_{i5})$ still follow a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\mathbf{\Sigma} = (0.8^{|i-j|})$ for $i, j = 1, \ldots, 5$ while the remaining $(p_n - 5)$ covariates independently follow a standard norm distribution.
Table 7 reports the selection proportions $\mathcal{P}_e$ for $Z_1, \ldots, Z_6$ and $\mathcal{P}_a$ for all active predictors under sample size $n = 100$. It can be seen that the proposed method exhibits reasonable results and outperforms the existing methods whether the independence assumption holds or not.

Example 4. We further compare the performances of the proposed CR-SIS method and the CSIRS method. This example is adapted from Zhou and Zhu (in press). In particular, the survival times T_i were generated from the model

$$\log(T_i) = c_0(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}) + \epsilon_i,$$

where the error ϵ_i was generated independently from the type-I extreme value distribution EV(0,1), i.e., the cumulative distribution function of ϵ_i has the form of $F_{\epsilon_i}(u) = \exp\{-\exp(-u)\}$. We set $\boldsymbol{\beta} = (1, 0.8, 0.6, 0.4, 0.2, 0, \ldots, 0)^{\mathrm{T}}$, i.e., only the first five predictors are active. The constant c_0 is used to control the signal-to-noise ratio. We set $c_0 = 1$ and $c_0 = 0.25$ to represent the high and the low signal-to-noise ratio, respectively. The remaining setups were kept the same as those in Example 2. The corresponding results are summarized in Tables 8 and 9, from which we can see that the performances of the CR-SIS and CSIRS procedures are comparable in all the considered scenarios; both of them deliver favorable results.

5. A real example

As an illustration, we applied the proposed screening method to the mantle cell lymphoma (MCL) data, which was studied by Rosenwald et al. (2003). The gene expression data set contains expression values of 8810 cDNA elements, which can be downloaded from 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, which led to a censoring rate of 30.4%. The mean survival time was 2.8 years (ranging from 0.02

Table 8
Five quantiles of the minimum model size $\mathcal{S}$ among 500 replications in Example 4 with the true model size $p_0 = 5$.

c_0	n	Method	5%	25%	50%	75%	95%
0.25	50	CR-SIS	6	37	143	406	1385
		CSIRS	8	38	125	384	1298
	100	CR-SIS	5	6	16	68	407
		CSIRS	5	6	13	55	488
	200	CR-SIS	5	5	5	6	23
		CSIRS	5	5	5	6	21
1	50	CR-SIS	5	5	5	6	18
		CSIRS	5	5	5	6	24
	100	CR-SIS	5	5	5	5	5
		CSIRS	5	5	5	5	5
	200	CR-SIS	5	5	5	5	5
		CSIRS	5	5	5	5	5

Table 9
Selection proportions $\mathcal{P}_e$ for each active predictor and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 4.

c_0	n	Method	$\mathcal{P}_e$					$\mathcal{P}_a$
			X_1	X_2	X_3	X_4	X_5	
0.25	50	CR-SIS	0.508	0.558	0.502	0.402	0.208	0.108
		CRIS	0.490	0.572	0.552	0.390	0.200	0.094
	100	CR-SIS	0.916	0.950	0.928	0.820	0.618	0.544
		CRIS	0.940	0.966	0.940	0.836	0.624	0.580
	200	CR-SIS	1.000	1.000	1.000	0.998	0.974	0.974
		CRIS	1.000	1.000	1.000	0.996	0.974	0.974
1	50	CR-SIS	1.000	1.000	1.000	0.994	0.918	0.918
		CRIS	0.998	1.000	1.000	0.992	0.912	0.912
	100	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		CRIS	1.000	1.000	1.000	1.000	1.000	1.000
	200	CR-SIS	1.000	1.000	1.000	1.000	1.000	1.000
		CRIS	1.000	1.000	1.000	1.000	1.000	1.000

Table 10
The screened UNIQIDs of genes for the Mantle cell lymphoma data.

CR-SIS	CSIRS	CRIS	P-SIS	FAST-SIS
30157	27095	30334	30157	30157
28346	30157	28872	34771	27095
27762	25234	17326	27095	34771
15936	32187	28990	27019	34790
24723	34790	17370	27762	32699
17198	28346	34790	30282	29330
27116	24794	34771	16587	28346
16312	34771	31420	28872	24713
34771	31420	27049	28346	16587
34790	16528	25234	34790	27762
27095	17326	16528	24723	15936
30334	28872	32699	25234	30282
31420	28990	30157	34687	25234
25234	32699	30282	32699	24723
24610	17343	27095	24734	27049
17326	27049	32187	24656	27019
17434	34687	33549	16528	28872
24656	26950	24710	17343	29209
30917	24723	24404	27049	31420
17174	24610	17176	31420	17343

to 14.05 years). Taking the survival times as the response and excluding the genes with missing values, we screened the important ones among the 6312 genes using the CR-SIS, CSIRS, CRIS, FAST-SIS and P-SIS approaches, respectively.

We set the model size to be $\lceil 92/\log(92) \rceil = 20$ and summarized the first 20 screened gene unique identifications (UNIQIDs) in Table 10. We can see that six genes whose UNIQIDs are 25234 (i.e., Antigen identified by monoclonal antibody Ki-67), 27095 (i.e., Topoisomerase (DNA) II alpha 170kDa), 30157 (i.e., Centromere protein F, 350/400ka (mitosin)), 31420 (i.e., Aurora kinase B), 34771 (i.e., Tubulin, alpha, ubiquitous) and 34790 (i.e., Thymidine kinase 1, soluble) were all selected by the considered five screening methods, indicating that these genes could be strongly associated with patients' survival

Table 11
The results of selected important genes for the Mantle cell lymphoma data using the regularization methods.

LASSO		SCAD		MCP	
UNIQID	EST.	UNIQID	EST.	UNIQID	EST.
30157	0.233	30157	0.826	30157	0.814
34687	0.124	34687	0.391	34687	0.433
27095	0.116				
29330	0.059				
34790	0.025				
34771	0.020				

risk. Moreover, the FAST-SIS, P-SIS and CR-SIS methods all rank gene 30157 at the top position and the CSIRS method ranks it at the second top position. However, the CRIS method throws such informative gene in less attractive corner.

The screening methods are usually considered as an initial step to reduce the dimensionality and then followed with some model-based regularization methods. In particular, we first applied the P-SIS procedure of Zhao and Li (2012) to reduce the dimension from $p_n = 6312$ to $d_n = 3\lceil n/\log(n)\rceil = 60$ and then utilized different regularization methods such as the LASSO, SCAD and MCP penalties to select the significant ones among these 60 genes under the framework of the Cox proportional hazards regression. The selection results for UNIQID and the estimated value of the coefficient of selected predictors are summarized in Table 11. We can see that six genes whose UNIQIDs are 30157, 34687, 27095, 29330, 34790, 34771 were selected by LASSO method; both the SCAD and MCP methods selected 30157 and 34687 simultaneously. On the other hand, the selection results of the proposed screening method and the regularization methods coincide with each other to much extent, which further implies that the proposed screening method can offer acceptable results.

6. Conclusion

We propose a novel model-free feature screening procedure for the ultrahigh-dimensional data, including complete data and censored data. Its theoretical properties are established when the number of covariates diverges at an exponential rate of the sample size. Numerical studies demonstrate that the performance of the proposed method is competitive with the existing model-dependent procedures such as the SIS, FAST-SIS and P-SIS procedures. However, for the complicated nonlinear models, the proposed model-free screening procedure delivers favorable performance over the existing ones. As common to all existing sure independent screening procedures based on marginal utilities, our method also suffers from the situations where covariates are jointly but not marginally important. It is warranted for further exploration of correlations among predictors using iterative approaches.

Acknowledgment

This work is supported in part by grants from the National Science Foundation of China (11571263, 11671311).

Appendix. Theoretic proofs

Proof of Theorem 1. Let

$$r_k^* = \left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right\}^2.$$

We prove this theorem via two steps.

First, we derive the exponential tail probability bound of $P(|\widetilde{r}_k - r_k^*| \geq \upsilon n^{-\alpha})$ for any positive constants υ and $0 \leq \alpha < 1/2$. Straightforward calculations entail that

$$\begin{aligned}\left|\widetilde{r}_k - r_k^*\right| &= \left|\left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widetilde{G}_n(X_i)\right\}^2 - \left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right\}^2\right| \\ &= \left|\left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widetilde{G}_n(X_i) + \frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right\}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widetilde{G}_n(X_i) - \frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right\}\right|.\end{aligned}$$

By the strong law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ik}^2 \xrightarrow{\text{a.s.}} E(Z_k^2).$$

Combining it with condition C2, there exists a positive constant c_1 such that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ik}^2 \le c_1^2 \tag{A.1}$$

holds almost surely when n is sufficiently large. Without loss of generality, assume that (A.1) holds for the total probability space as the set with measure zero does not affect the derivations. Using the Cauchy–Schwarz inequality and the boundedness of $\widetilde{G}_n(t)$ and $G(t)$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widetilde{G}_n(X_i)\right| \le c_1 \quad \text{and} \quad \left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right| \le c_1. \tag{A.2}$$

Using condition C1, along with (A.1) and (A.2), we have

$$\begin{aligned}
\left|\widetilde{r}_k - r_k^*\right| &\le c_3\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\widetilde{G}_n(X_i) - \frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i)\right| \\
&\le \frac{c_3}{\delta}\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}\left\{\widetilde{G}_n(X_i) - G(X_i)\right\}S(X_i)\right| \\
&\le \frac{c_3 c_1}{\delta}\left(\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\widetilde{G}_n(X_i) - G(X_i)\right\}S(X_i)\right]^2\right)^{1/2} \\
&\le c_4 \sup_{0\le t\le\tau}\left|S(t)\left\{\widetilde{G}_n(t) - G(t)\right\}\right|,
\end{aligned}$$

where $c_3 = 2c_1$, $c_4 = c_3c_1/\delta$ and $S(t) = P(C > t)$ represents the survival function of the censoring variable. It follows from Theorem 1 in Bitouzé et al. (1999) that

$$\begin{aligned}
P\left(|\widetilde{r}_k - r_k^*| \ge \upsilon n^{-\alpha}\right) &\le P\left(c_4 \sup_{0\le t\le\tau}\left|S(t)\left\{\widetilde{G}_n(t) - G(t)\right\}\right| \ge \upsilon n^{-\alpha}\right) \\
&= P\left(n^{1/2}\sup_{0\le t\le\tau}\left|S(t)\left\{\widetilde{G}_n(t) - G(t)\right\}\right| \ge c_4^{-1}\upsilon n^{1/2-\alpha}\right) \\
&\le 2.5\exp(-2c_4^{-2}\upsilon^2 n^{1-2\alpha} + c_5c_4^{-1}\upsilon n^{1/2-\alpha}),
\end{aligned} \tag{A.3}$$

where c_5 is a constant.

Second, we derive the exponential tail probability bound of $P(|r_k^* - r_k| \ge \upsilon n^{-\alpha})$ for any positive constants υ and $0 \le \alpha < 1/2$. Using the similar arguments, we also have

$$|r_k^* - r_k| \le c_3\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i) - E\left\{Z_kG(X)\right\}\right|.$$

By the exponential Chebyshev inequality, for any $\zeta > 0$, we have

$$\begin{aligned}
P\left(\left|r_k^* - r_k\right| \ge \upsilon n^{-\alpha}\right) &\le P\left(c_3\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i) - E\left\{Z_kG(X)\right\}\right| \ge \upsilon n^{-\alpha}\right) \\
&= P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i) - E\left\{Z_kG(X)\right\}\right| \ge c_3^{-1}\upsilon n^{-\alpha}\right) \\
&\le \exp(-\zeta c_3^{-1}\upsilon n^{-\alpha})\cdot E\left(\exp\left[\zeta\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i) - E\left\{Z_kG(X)\right\}\right|\right]\right).
\end{aligned} \tag{A.4}$$

Using the law of the iterated logarithm, we have

$$\limsup_{n\to\infty}\frac{\sum_{i=1}^{n} Z_{ik}G(X_i) - nE\left\{Z_kG(X)\right\}}{\left[n\log\log n\cdot \mathrm{Var}\left\{Z_kG(X)\right\}\right]^{1/2}} = \sqrt{2}, \quad \text{a.s.}$$

Without loss of generality, when n is large enough and removing a zero measure set, under condition C2, there exists a positive constant c_6 such that

$$\left(\frac{n}{\log\log n}\right)^{1/2} \cdot \left[\frac{1}{n}\sum_{i=1}^{n} Z_{ik}G(X_i) - E\left\{Z_k G(X)\right\}\right] \le c_6. \tag{A.5}$$

We chose $\zeta = \left(\frac{n}{\log\log n}\right)^{1/2}$, then it follows from (A.4) and (A.5) that

$$P\left(\left|r_k^* - r_k\right| \ge \upsilon n^{-\alpha}\right) \le \exp\left\{-\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2} c_3^{-1}\upsilon + c_6\right\}. \tag{A.6}$$

Combining (A.3) and (A.6), we have

$$\begin{aligned} P\left(\left|\widetilde{r}_k - r_k\right| \ge 2\upsilon n^{-\alpha}\right) &\le P\left(\left|\widetilde{r}_k - r_k^*\right| \ge \upsilon n^{-\alpha}\right) + P\left(\left|r_k^* - r_k\right| \ge \upsilon n^{-\alpha}\right) \\ &\le 2.5\exp(-2c_4^{-2}\upsilon^2 n^{1-2\alpha} + c_5 c_4^{-1}\upsilon n^{1/2-\alpha}) + \exp\left\{-\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2} c_3^{-1}\upsilon + c_6\right\} \\ &\le O\left[\exp\left\{-\eta\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2}\right\}\right], \end{aligned} \tag{A.7}$$

where $\eta = c_3^{-1}\upsilon$. Immediately, we have

$$P\left(\max_{1\le k\le p_n}\left|\widetilde{r}_k - r_k\right| \ge 2\upsilon n^{-\alpha}\right) \le O\left[p_n \exp\left\{-\eta\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2}\right\}\right], \tag{A.8}$$

which proves the first part of Theorem 1 by taking $c = 2\upsilon$.

If $\mathcal{A} \not\subseteq \widetilde{\mathcal{A}}$, then there must exist some $k \in \mathcal{A}$ such that $\widetilde{r}_k < cn^{-\alpha}$. It follows from condition C3 that $|\widetilde{r}_k - r_k| > cn^{-\alpha}$ for some $k \in \mathcal{A}$, which implies that $\{\mathcal{A} \not\subseteq \widetilde{\mathcal{A}}\} \subseteq \{|\widetilde{r}_k - r_k| > cn^{-\alpha}$ for some $k \in \mathcal{A}\}$. As a result, $\{\max_{k\in\mathcal{A}}|\widetilde{r}_k - r_k| \le cn^{-\alpha}\} \subseteq \{\mathcal{A} \subseteq \widetilde{\mathcal{A}}\}$. Using (A.7), we have

$$\begin{aligned} P\left(\mathcal{A} \subseteq \widetilde{\mathcal{A}}\right) &\ge P\left(\max_{k\in\mathcal{A}}|\widetilde{r}_k - r_k| \le cn^{-\alpha}\right) \\ &\ge 1 - O\left[a_n \exp\left\{-\eta\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2}\right\}\right], \end{aligned}$$

where $a_n = |\mathcal{A}|$. Thus, the proof of Theorem 1 is completed. □

Proof of Theorem 2. Under assumption (i), we rewrite

$$R_k(Y) = E\left[Z_k E\left\{G(Y)|\mathbf{Z}\right\}\right] = E\left[Z_k E\left\{G(Y)|\mathbf{Z}_{\mathcal{A}}\right\}\right].$$

If $k \notin \mathcal{A}$, then assumption (ii) implies that

$$R_k(Y) = E(Z_k)E\left[E\left\{G(Y)|\mathbf{Z}_{\mathcal{A}}\right\}\right] = 0.$$

As a result, $r_k = [R_k(Y)]^2 = 0$. It follows from condition C3 that $\max_{k\notin\mathcal{A}} r_k < \min_{k\in\mathcal{A}} r_k$. On the other hand, $r_k = 0$ directly implies that $k \notin \mathcal{A}$ under condition C3. Thus, the first part of Theorem 2 is proved.

Under condition C3 and assumptions (i) and (ii), coupled with (A.8), we have

$$\begin{aligned} P\left(\min_{k\in\mathcal{A}}\widetilde{r}_k \le \max_{k\notin\mathcal{A}}\widetilde{r}_k\right) &= P\left(\max_{k\notin\mathcal{A}}\widetilde{r}_k - \max_{k\notin\mathcal{A}} r_k - \min_{k\in\mathcal{A}}\widetilde{r}_k + \min_{k\in\mathcal{A}} r_k \ge \min_{k\in\mathcal{A}} r_k\right) \\ &\le P\left(\max_{k\notin\mathcal{A}}|\widetilde{r}_k - r_k| \ge cn^{-\alpha}\right) + P\left(\max_{k\in\mathcal{A}}|\widetilde{r}_k - r_k| \ge cn^{-\alpha}\right) \\ &\le 2P\left(\max_{1\le k\le p_n}|\widetilde{r}_k - r_k| \ge cn^{-\alpha}\right) \\ &\le O\left[p_n \exp\left\{-\eta\left(\frac{n^{1-2\alpha}}{\log\log n}\right)^{1/2}\right\}\right], \end{aligned}$$

which completes the proof of Theorem 2. □

References

Bitouzé, D., Laurent, B., Massart, P., 1999. A dvoretzky–kiefer–wolfowitz type inequality for the kaplan–meier estimator. Ann. Inst. H. Poincaré Probab. Statist. 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.

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.

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.

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.

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.

Zhou, T.Y., Zhu, L.P., 2016. Model-free feature screening for ultrahigh dimensional censored regression. Stat. Comput. http://dx.doi.org/10.1007/s11222-016-9664-z, (in press).

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.

Zhu, L.P., Zhu, L.X., 2009. On distribution-weighted partial least squares with diverging number of highly correlated predictors. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 525–548. http://dx.doi.org/10.1111/j.1467-9868.2008.00697.x.

Zou, H., 2006. The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429. http://dx.doi.org/10.1198/016214506000000735.