Lifetime Data Anal (2018) 24:273–292
https://doi.org/10.1007/s10985-017-9395-2

Censored cumulative residual independent screening for ultrahigh-dimensional survival data

Jing Zhang[1] · **Guosheng Yin**[2] · **Yanyan Liu**[1] · **Yuanshan Wu**[1]

Received: 3 November 2015 / Accepted: 16 May 2017 / Published online: 26 May 2017
© Springer Science+Business Media New York 2017

Abstract For complete ultrahigh-dimensional data, sure independent screening methods can effectively reduce the dimensionality while retaining all the active variables with high probability. However, limited screening methods have been developed for ultrahigh-dimensional survival data subject to censoring. We propose a censored cumulative residual independent screening method that is model-free and enjoys the sure independent screening property. Active variables tend to be ranked above the inactive ones in terms of their association with the survival times. Compared with several existing methods, our model-free screening method works well with general survival models, and it is invariant to the monotone transformation of the responses, as well as requiring substantially weaker moment conditions. Numerical studies demonstrate the usefulness of the censored cumulative residual independent screening method, and the new approach is illustrated with a gene expression data set.

Keywords Cumulative residual · Model-free screening · Sure screening property · Survival data · Ultrahigh-dimensional data

✉ Yuanshan Wu
shan@whu.edu.cn

Jing Zhang
jing66@whu.edu.cn

Guosheng Yin
gyin@hku.hk

Yanyan Liu
liuyy@whu.edu.cn

[1] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, Hubei, China

[2] Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, China

1 Introduction

High-dimensional data often arise in real applications, where the number of variables p_n can be much larger than the sample size n. In the aspect of dimension reduction, numerous regularization methods have been proposed to select the active variables, such as the LASSO (Tibshirani 1996), smoothly clipped absolute deviation (Fan and Li 2001), adaptive LASSO (Zou 2006), the Dantzig selector (Candes and Tao 2007), and the minimax concave penalty (Zhang 2010). These penalized procedures generally work well for moderate to large p_n, which may increase with n at a polynomial rate. However, the performance of such regularization methods deteriorates and the associated computation is demanding for ultrahigh dimension p_n that increases at an exponential rate of n (Fan et al. 2009).

To overcome the difficulties associated with ultrahigh dimensionality, Fan and Lv (2008) proposed a sure independence screening (SIS) method to reduce the dimension in the context of linear regression models. Fan and Song (2010) extended SIS to ultrahigh-dimensional generalized linear models, and Fan et al. (2011) studied SIS for ultrahigh-dimensional additive models. Zhu et al. (2011) proposed a sure independent ranking and screening procedure for ultrahigh-dimensional general multi-index models, which avoids the specification of a particular model structure. Li et al. (2012) proposed a model-free SIS procedure based on the distance correlation.

For censored ultrahigh-dimensional data, Fan et al. (2010) investigated SIS methods for the Cox proportional hazards model via ranking variables according to their respective univariate partial log-likelihoods. Tibshirani (2009) suggested a soft thresholding procedure for the univariate Cox score statistics. Zhao and Li (2012) proposed a screening method based on the standardized marginal maximum partial likelihood estimator under 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 through inverse probability weighting of censoring, Song et al. (2014) proposed a censored rank independence screening method, which is robust to outliers and works for a general class of survival models. Wu and Yin (2015) developed a screening method which is constructed to identify the covariates that contribute to the conditional quantile of the response.

In a model-free fashion, we propose a censored cumulative residual independent screening (CCRIS) procedure for the ultrahigh-dimensional survival data. The Kaplan–Meier estimator is employed to handle censoring through the inverse probability weighting scheme. It is known that the martingale-based cumulative residuals play a central role in the diagnostics of survival models (Lin et al. 1993; Cook et al. 2007). Compared with the existing procedures, our approach enjoys several distinctive advantages. First, our procedure does not rely upon any model assumption and thus it works for various types of censored regression models. Second, our approach is invariant under the monotone transformation of the response. Third, we do not impose any moment conditions on the variables, while other screening procedures often require the existence of the exponential moments. The proposed CCRIS procedure does not involve any nonparametric approximation except for the Kaplan–Meier estimator, which greatly facilitates the SIS implementation in real applications.

The rest of the article is organized as follows. In Sect. 2, we propose the CCRIS procedure for the ultrahigh-dimensional survival data. We establish the theoretical properties of the proposed procedure in Sect. 3, and conduct simulation studies to evaluate its finite sample performance in Sect. 4. A real data example of the mantle cell lymphoma study is analyzed in Sect. 5, followed by some remarks in Sect. 6. All technical proofs are presented in the Appendix.

2 Screening procedures

For ultrahigh-dimensional survival data, suppose that the observations $\{X_i, \Delta_i, \mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{ip_n})^{\mathrm{T}} : i = 1, \ldots, n\}$ are independent copies of $(X, \Delta, \mathbf{Z})$, where $X = \min(T, C)$, $\Delta = I(T \leq C)$, T and C respectively denote the failure and censoring times, and $\mathbf{Z} = (Z_1, \ldots, Z_{p_n})^{\mathrm{T}}$ is the covariate vector. For ease of exposition, assume that the censoring mechanism is completely random in the sense that C is independent of $(T, \mathbf{Z})$.

In an ultrahigh-dimensional setting, consider the conditional survival function $S(t|\mathbf{Z}) = P(T > t|\mathbf{Z})$, where the dimensionality p_n, possibly depending on n, greatly exceeds sample size n, and can be allowed to increase at the exponential rate of n. To identify the active ones from p_n covariates, we define the active covariate set as

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

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

We observe that if the kth covariate Z_k does not contribute to the conditional survival function, it leads to $P(T > t|Z_k) = P(T > t)$ for any t. Further calculation yields that

$$E\left[\left\{\frac{\Delta I(X > t)}{G(X)} - H(t)\right\}\middle| Z_k\right] = 0,$$

where $G(t) = P(C > t)$ and $H(t) = P(T > t)$ are the survival functions of the censoring and failure times, respectively. Motivated by the definition of conditional expectation, we define

$$d_k(t, z) = E\left[\left\{\frac{\Delta I(X > t)}{G(X)} - H(t)\right\} I(Z_k < z)\right].$$

The empirical counterpart of $d_k(t, z)$ is given by

$$\widehat{d}_k(t, z) = n^{-1} \sum_{i=1}^{n} \left[\left\{\frac{\Delta_i I(X_i > t)}{\widehat{G}_n(X_i)} - \widehat{H}_n(t)\right\} I(Z_{ik} < z)\right],$$

where $\widehat{G}_n(t)$ and $\widehat{H}_n(t)$ are the Kaplan–Meier estimators for $G(t)$ and $H(t)$, respectively. We set $0/0 = 1$ so that $\widehat{d}_k(t, z)$ is well-defined. If the conditional survival

function of T given Z_k does not depend on Z_k, then $\widehat{d}_k(t, z)$ is expected to fluctuate around zero over the two-dimensional space spanned by t and z. Furthermore, $\widehat{d}_k(t, z)$ can also be viewed as a censored version of Hoeffding's independence test between T and Z_k (Hoeffding 1948).

Based on this rationale, we define the marginal screening utility for the kth predictor,

$$\|\widehat{d}_k\|_n^2 = n^{-1} \sum_{j=1}^{n} \widehat{d}_k \left(X_j, Z_{jk}\right)^2,$$

and those predictors with a large value of $\|\widehat{d}_k\|_n^2$ are considered important. As a result, we define the estimated active set as

$$\widehat{\mathcal{A}} = \left\{k : \|\widehat{d}_k\|_n^2 \geq cn^{-\alpha},\ k = 1, \ldots, p_n\right\},$$

where constants c and α are specified in the regularity condition.

An inherent issue with the marginal screening procedures is that predictors that are jointly related but marginally unrelated with the response may be screened out. An iterative procedure is developed to alleviate the negative effect arising from such subtle issue as well as to enhance the performance of the proposed CCRIS method. For ease of exposition, we denote $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$, $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_n)^{\mathrm{T}}$, and $\mathbb{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^{\mathrm{T}}$. Let $\mathbb{Z}_{\mathcal{T}}$ be the submatrix of $\mathbb{Z}$, whose columns consist of the index set $\mathcal{T} \subset \{1, \ldots, p_n\}$. The iterative version of CCRIS (ICCRIS) proceeds as follows.

1. We first apply the CCRIS procedure for the data $(\mathbf{X}, \mathbf{\Delta}, \mathbb{Z})$. The resulting selected active set is denoted by $\widehat{\mathcal{A}}_1$ whose size is assumed to be q_1.
2. Define the residual of predictors matrix $\mathbb{Z}$ after screening in the active set $\widehat{\mathcal{A}}_1$ as

$$\mathbb{Z}^{(1)} = \left\{\mathbf{I}_n - \mathbb{Z}_{\widehat{\mathcal{A}}_1} \left(\mathbb{Z}_{\widehat{\mathcal{A}}_1}^{\mathrm{T}} \mathbb{Z}_{\widehat{\mathcal{A}}_1}\right)^{-1} \mathbb{Z}_{\widehat{\mathcal{A}}_1}^{\mathrm{T}}\right\} \mathbb{Z}_{\widehat{\mathcal{A}}_1^c},$$

   where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Applying the CCRIS procedure for the data $(\mathbf{X}, \mathbf{\Delta}, \mathbb{Z}^{(1)})$ yields the selected active set $\widehat{\mathcal{A}}_2$ with size of q_2, which is the subset of $\widehat{\mathcal{A}}_1^c$.
3. Update the selected active set by $\widehat{\mathcal{A}}_1 \cup \widehat{\mathcal{A}}_2$.
4. Repeat steps 2 and 3 $N - 1$ times until the number of totally selected active predictors $q_1 + \cdots + q_N$ exceeds a prespecified number. Finally, the selected active predictor set is $\widehat{\mathcal{A}}_1 \cup \ldots \cup \widehat{\mathcal{A}}_N$.

3 Theoretical properties

We show that the CCRIS procedure enjoys the sure independent screening and the ranking consistency properties. We impose two 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. It holds that

$$\min_{k \in \mathcal{A}} \|d_k\|_n^2 \geq 2cn^{-\alpha}$$

for some constants $c > 0$ and $\alpha \in [0, 1/2)$, where $\|d_k\|_n^2 = E\{d_k(X, Z_k)^2\}$.

Condition C1 is a common assumption in survival analysis with τ being the end time of the study. Condition C2 requires that the screening utilities carrying information for the active predictors should not decay too fast. It is interesting to observe that we do not impose any moment conditions for predictors Z_i's. Compared with the exponential moment condition C1 in Li et al. (2012) and condition C3 in Zhu et al. (2011), ours is weaker. In the next theorem, we present the sure independent screening property of the CCRIS procedure for ultrahigh-dimensional survival data.

Theorem 1 *Under condition C1, there exists a constant $\eta > 0$ such that*

$$P\left(\max_{1 \leq k \leq p_n} \left| \left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2 \right| \geq cn^{-\alpha}\right) \leq O\left\{p_n \exp\left(-\eta n^{1-2\alpha}\right)\right\}.$$

Under conditions C1 and C2, it holds that

$$P\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}\right) \geq 1 - O\left\{a_n \exp\left(-\eta n^{1-2\alpha}\right)\right\},$$

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 CCRIS procedure is summarized as follows.

Theorem 2 *Assume that (i) T 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 C2, we have that*

$$\max_{k \notin \mathcal{A}} \|d_k\|_n^2 < \min_{k \in \mathcal{A}} \|d_k\|_n^2,$$

and $\|d_k\|_n^2 = 0$ if and only if $k \notin \mathcal{A}$. Furthermore, under conditions C1 and C2, there exists a constant $\eta > 0$ such that

$$P\left(\max_{k \notin \mathcal{A}} \left\|\widehat{d}_k\right\|_n^2 < \min_{k \in \mathcal{A}} \left\|\widehat{d}_k\right\|_n^2\right) \geq 1 - O\left\{p_n \exp\left(-\eta n^{1-2\alpha}\right)\right\}.$$

This lays out the theoretical foundation that our CCRIS 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 compare it with existing methods via simulation studies. For brevity, we refer to 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, and the censored rank independence screening of Song et al. (2014) as CRIS.

Example 1 We generated survival time T_i from the Cox proportional hazards model with the conditional hazard function given by

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

where the baseline hazard function is set to be $\lambda_0(t) = (t - 0.5)^2$ and 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 the true parameters $\boldsymbol{\beta}_0 = (0.35, 0.35, 0.35, 0.35, 0.35, 0, \ldots, 0)^{\mathrm{T}}$, 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$(0, \tau + 2)$. The study duration τ was chosen to yield the desirable censoring rate of 20%. We took the sample size $n = 50$, 100 and 200, coupled with a large 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. 2012). 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 simultaneously 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. In general, the performances of the FAST-SIS and P-SIS procedures are comparable. Both of them outperform the CCRIS and CRIS procedures as they take into account the Cox proportional model structure while the latter two screening procedures does not rely on the specific model structure. Nevertheless, the proposed CCRIS method exhibits more satisfactory results than the CRIS method. All of the four methods perform equally well when the sample size is increased to 200.

To gain more insight into the proposed CCRIS procedure, we further plot in Fig. 1 the scatter points of $\widehat{d}_k(t, z)$ for $k = 1$ and 10 against t and z based on one simulated dataset from Example 1. For $k = 1$, the first covariate Z_1 is an active one, and thus

Table 1 Five quantiles of $\mathcal{S}$ (the minimum model size needed to include all active predictors) among 500 replications in Example 1 with the true model size $p_0 = 5$

n	Method	5%	25%	50%	75%	95%
50	CCRIS	7	21	72	248	1055
	CRIS	6	18	72	253	967
	FAST-SIS	5	5	6	10	65
	P-SIS	5	5	5	8	41
100	CCRIS	5	5	8	20	109
	CRIS	5	5	7	13	114
	FAST-SIS	5	5	5	5	6
	P-SIS	5	5	5	5	6
200	CCRIS	5	5	5	6	9
	CRIS	5	5	5	5	7
	FAST-SIS	5	5	5	5	5
	P-SIS	5	5	5	5	5

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

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

n	Method	$\mathcal{P}_e$					$\mathcal{P}_a$
		X_1	X_2	X_3	X_4	X_5	
50	CCRIS	0.408	0.572	0.628	0.548	0.340	0.148
	CRIS	0.464	0.652	0.660	0.592	0.398	0.172
	FAST-SIS	0.892	0.976	0.990	0.990	0.894	0.798
	P-SIS	0.918	0.990	1.000	0.998	0.934	0.864
100	CCRIS	0.860	0.954	0.982	0.960	0.868	0.754
	CRIS	0.896	0.970	0.978	0.962	0.884	0.816
	FAST-SIS	0.998	1.000	1.000	1.000	0.998	0.996
	P-SIS	0.998	1.000	1.000	1.000	1.000	0.998
200	CCRIS	0.996	0.998	0.998	1.000	0.992	0.990
	CRIS	0.996	1.000	1.000	1.000	0.998	0.994
	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

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

the scatter points of $\widehat{d}_1(t, z)$ substantially deviate from the zero surface. For $k = 10$, the tenth covariate Z_{10} is not an active predictor, and thus those of $\widehat{d}_{10}(t, z)$ for the inactive covariate undulate around the zero surface.

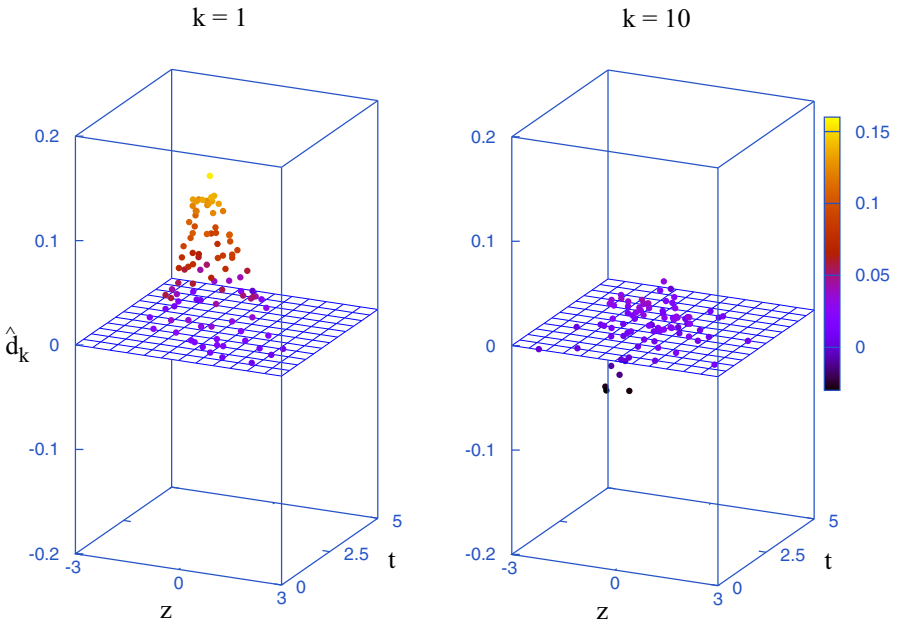

Fig. 1 The scatter points of $\widehat{d}_k(t, z)$ based on one simulated dataset with the sample size $n = 100$ in Example 1

Example 2 To examine the performance of the proposed screening procedure for censored nonlinear survival models with interactions, we generated the log survival times from the model

$$\log T = (2 + \sin Z_1)^2 + 0.5(1 + Z_5)^{-3} + 3(Z_{10}^2 + Z_{10}) + 0.5 Z_1 Z_{10} + \epsilon,$$

where the error ϵ was generated from the standard normal distribution. The remaining setups were kept the same as those in Example 1. The corresponding results are summarized in Tables 3 and 4, from which we can see that the CCRIS method is able to capture the nonlinear covariate effects with interactions and delivers favorable screening results. Specifically, the CCRIS method outperforms the other three screening procedures, especially the model-dependent FAST-SIS and P-SIS procedures, in terms of the minimum model size required to cover all the active covariates and the proportion that all active predictors are selected for a given model size.

We further evaluate the performance of the proposed screening method when some predictors are categorical variables. To be specific, predictor Z_1 was generated from the Bernoulli distribution with success probability 0.5, or 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. The remaining $(p_n - 1)$ covariates $(Z_2, \ldots, Z_{p_n})$ still follows a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\mathbf{\Sigma} = (0.8^{|i-j|})$ for $i, j = 1, \ldots, (p_n - 1)$. Simulation results are summarized in Table 5, from which

Table 3 Five quantiles of $\mathcal{S}$ (the minimum model size needed to include all active predictors) among 500 replications in Example 2 with the true model size $p_0 = 3$

n	Method	5%	25%	50%	75%	95%
50	CCRIS	21	106	327	821	1521
	CRIS	125	583	1157	1695	1944
	FAST-SIS	110	596	1130	1577	1933
	P-SIS	120	550	1140	1581	1933
100	CCRIS	6	18	71	253	1129
	CRIS	61	402	1019	1602	1929
	FAST-SIS	83	493	1017	1528	1920
	P-SIS	64	467	972	1514	1928
200	CCRIS	4	7	12	35	287
	CRIS	40	283	797	1412	1930
	FAST-SIS	62	349	848	1380	1866
	P-SIS	54	321	780	1369	1859

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

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_5	X_{10}	
50	CCRIS	0.370	0.104	0.366	0.026
	CRIS	0.156	0.022	0.180	0.000
	FAST-SIS	0.142	0.004	0.250	0.000
	P-SIS	0.166	0.010	0.174	0.000
100	CCRIS	0.796	0.358	0.796	0.284
	CRIS	0.412	0.038	0.452	0.022
	FAST-SIS	0.372	0.016	0.602	0.014
	P-SIS	0.372	0.022	0.444	0.018
200	CCRIS	0.984	0.770	0.980	0.760
	CRIS	0.770	0.052	0.740	0.046
	FAST-SIS	0.776	0.032	0.912	0.030
	P-SIS	0.754	0.034	0.842	0.026

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

we can see that the CCRIS method is able to capture the discrete predictor and performs overwhelmingly superior to the other three screening procedures.

We also consider the covariate-dependent censoring scheme where the completely random censoring mechanism does not hold. In particular, the censoring times C were generated from a uniform distribution if $Z_1 > 1$, otherwise from an exponential distribution. With sample size $n = 100$, Table 6 shows that the proposed method

Table 5 The median of the minimum model size $\mathcal{S}$, $\mathcal{S}_{50\%}$, and the selection proportions $\mathcal{P}_e$ and $\mathcal{P}_a$ in Example 2 with Z_1 being a discrete predictor

Z_1	n	Method	$\mathcal{S}_{50\%}$	$\mathcal{P}_e$			$\mathcal{P}_a$
				X_1	X_5	X_{10}	
Bernoulli	100	CCRIS	210	0.648	0.160	0.848	0.102
		CRIS	1098	0.654	0.010	0.478	0.002
		FAST-SIS	1139	0.318	0.020	0.560	0.002
		P-SIS	1125	0.320	0.022	0.414	0.002
	200	CCRIS	54	0.930	0.446	0.982	0.434
		CRIS	953	0.948	0.022	0.784	0.012
		FAST-SIS	1023	0.664	0.006	0.916	0.002
		P-SIS	984	0.670	0.030	0.818	0.012
Categorical	100	CCRIS	265	0.750	0.152	0.706	0.112
		CRIS	1302	0.112	0.026	0.344	0.000
		FAST-SIS	1039	0.492	0.012	0.502	0.004
		P-SIS	1016	0.522	0.024	0.350	0.006
	200	CCRIS	64	0.950	0.438	0.962	0.414
		CRIS	1227	0.266	0.026	0.650	0.002
		FAST-SIS	1061	0.892	0.012	0.824	0.008
		P-SIS	1009	0.916	0.014	0.710	0.004

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

Table 6 Five quantiles of the minimum model size $\mathcal{S}$, and selection proportions $\mathcal{P}_e$ and $\mathcal{P}_a$ in Example 2 with sample size $n = 100$ under the covariate-dependent censoring scheme

Method	5%	25%	50%	75%	95%
CCRIS	7	26	92	398	1254
CRIS	57	379	981	1581	1950
FAST-SIS	119	569	997	1527	1906
P-SIS	93	553	971	1519	1896

	$\mathcal{P}_e$			$\mathcal{P}_a$
	X_1	X_5	X_{10}	
CCRIS	0.740	0.288	0.710	0.226
CRIS	0.442	0.018	0.428	0.014
FAST-SIS	0.368	0.010	0.556	0.008
P-SIS	0.382	0.020	0.424	0.012

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

Table 7 Five quantiles of the minimum model size $\mathcal{S}$ and selection proportion $\mathcal{P}_a$ for all active predictors among 500 replications in Example 3

p_0	Method	5%	25%	50%	75%	95%	$\mathcal{P}_a$
1	CCRIS	1	1	1	11	177	0.822
	CRIS	1	2	15	166	1286	0.538
	FAST-SIS	1	1	2	20	240	0.752
	P-SIS	1	1	3	20	272	0.756
5	CCRIS	5	5	7	24	278	0.734
	CRIS	5	7	28	233	1614	0.452
	FAST-SIS	5	6	18	76	494	0.556
	P-SIS	5	7	21	100	556	0.504
10	CCRIS	11	28	90	291	1232	0.216
	CRIS	16	76	310	938	1933	0.086
	FAST-SIS	13	55	185	546	1627	0.114
	P-SIS	14	59	191	562	1636	0.106
15	CCRIS	32	159	424	942	1717	0.014
	CRIS	77	365	1009	1722	1972	0.012
	FAST-SIS	51	246	618	1335	1883	0.012
	P-SIS	48	260	651	1342	1884	0.016

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

delivers satisfactory results even when the completely random censoring assumption does not hold.

Example 3 To assess the performance of the model size selection rule $\lceil n/\log n \rceil$, we generated the survival times from the nonlinear model,

$$\log T = \sum_{i=1}^{p_0} \left\{ (1 + Z_i)^{-1} + 1.5 Z_i \right\} + \epsilon,$$

and considered the true model size $p_0 = 1, 5, 10$, and 15, respectively, where the error ϵ was generated from the standard normal distribution. The remaining setups were kept the same as those in Example 1. Simulation results under sample size $n = 100$ are summarized in Table 7. It shows that if the true model size p_0 is increased while keeping the model size selection rule $[n/\log n]$ fixed, the performances of these four methods would deteriorate. Nevertheless, the proposed CCRIS method outperforms the others.

Example 4 To assess the performance of ICCRIS procedure, we generated the survival times T from the following transformation model adapted from Song et al. (2014),

$$H(T) = -\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \epsilon,$$

where $\boldsymbol{\beta} = (5, 5, 5, -15\rho^{1/2}, 0, \ldots, 0)^{\mathrm{T}}$, $H(t) = \log\{0.5(e^{2t} - 1)\}$ and ϵ follows the standard normal distribution. We generated the ultrahigh-dimensional covariate $\mathbf{Z}$ from a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{p_n \times p_n}$, where $\sigma_{ii} = 1$ for $i = 1, \ldots, p_n$, $\sigma_{i4} = \sigma_{4i} = \rho^{1/2}$ for $i \neq 4$, and $\sigma_{ij} = \rho$ for

Table 8 The selection proportions $\mathcal{P}_e$ for each active predictors and $\mathcal{P}_a$ for all active predictors among 500 replications in Example 4

ρ	Method	$\mathcal{P}_e$				$\mathcal{P}_a$
		X_1	X_2	X_3	X_4	
0	ICCRIS	0.926	0.930	0.912	0.020	0.848
	CCRIS	0.936	0.936	0.930	0.030	0.858
0.2	ICCRIS	0.960	0.960	0.960	0.704	0.610
	CCRIS	0.890	0.882	0.864	0.000	0.000
0.4	ICCRIS	0.966	0.960	0.948	0.774	0.680
	CCRIS	0.804	0.808	0.802	0.000	0.000
0.6	ICCRIS	0.956	0.958	0.944	0.852	0.734
	CCRIS	0.738	0.702	0.712	0.000	0.000
0.8	ICCRIS	0.970	0.960	0.956	0.902	0.814
	CCRIS	0.608	0.592	0.590	0.000	0.000

ICCRIS denotes the iterative version of CCRIS

$i \neq j, i \neq 4, j \neq 4$. We vary the value of ρ to be 0, 0.2, 0.4, 0.6 and 0.8, with a larger ρ yielding a higher collinearity. Based on such well-designed correlation matrix, we observe that, when $\rho \neq 0$, the active predictor Z_4 is jointly related but marginally unrelated with the transformed survival time $H(T)$. On the other hand, Z_4 is an inactive predictor when $\rho = 0$. For the ICCRIS procedure, we choose $N = 2$ and set $q_1 = \lceil d/2 \rceil$ and $q_2 = d - q_1$. Simulation results of the proposed CCRIS procedure and its iterative counterpart with sample size $n = 200$ and censoring rate 15% are summarized in Table 8. We can see that the ICCRIS and CCRIS procedures perform equally well when Z_4 is an indeed inactive predictor for $\rho = 0$. Whereas the ICCRIS procedure exhibits superior performances over the CCRIS procedure for ρ varying from 0.2 to 0.8, showing that the ICCRIS procedure can retain those predictors which are jointly but not marginally importation.

5 A real example

As an illustration, we applied the proposed CCRIS 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 is available from website of 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 to 14.05 years). We took the survival times as the response and excluded the genes with missing values. We applied the CCRIS, CRIS, P-SIS and FAST-SIS approaches to screen the important ones among the 6312 genes, 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 9. It shows that the gene UNIQIDs 31420 (i.e., Aurora kinase

Table 9 The screened UNIQIDs of genes for the mantle cell lymphoma study

CCRIS	CRIS	P-SIS	FAST-SIS
27116	30334	30157	30157
30917	28872	34771	27095
15936	17326	27095	34771
17174	28990	27019	34790
17198	17370	27762	32699
24404	34790	30282	29330
16312	34771	16587	28346
17176	31420	28872	24713
31420	27049	28346	16587
19325	25234	34790	27762
17917	16528	24723	15936
34790	32699	25234	30282
30142	30157	34687	25234
16020	30282	32699	24723
24758	27095	24734	27049
24656	32187	24656	27019
30334	33549	16528	28872
17326	24710	17343	29209
28148	24404	27049	31420
23887	17176	31420	17343

CCRIS, the proposed censored cumulative residual independent screening procedure; CRIS, the censored rank independence screening procedure of Song et al. (2014); FAST-SIS, the feature aberration at survival times screening procedure of Gorst-Rasmussen and Scheike (2013); P-SIS, the principled sure independent screening procedure of Zhao and Li (2012)

B) and 34790 (i.e., Thymidine kinase 1, soluble) were selected by all the considered four screening methods, indicating that these genes could be strongly associated with patients' survival times. Moreover, six of the top 20 genes selected by the mode-free screening method CRIS were also selected by the proposed CCRIS method, which indicates that the results of the two model-free screening methods largely coincide with each other. As a result, we conclude that the proposed method offers a reliable selection result for the MCL data.

6 Conclusion

We propose the censored cumulative residual independent screening procedure for variable selection with the ultrahigh-dimensional survival data. Its sure independent screening properties are established when the number of covariates diverges at an exponential rate of the sample size. Numerical studies demonstrate that the performances of the proposed method are competitive with the existing model-dependent procedures such as the FAST-SIS and P-SIS procedures. However, for the complicated nonlinear models which could be more feasible to capture the characteristic of the ultrahigh-dimensional survival data, the proposed model-free screening procedure delivers its distinctive advantages over the existing ones. An iterative procedure is also developed to enhance the performance of the proposed CCRIS method in the situation where covariates are jointly but not marginally important.

Acknowledgements We would like to thank the editor, the associate editor, and the referees for their insightful comments, which immensely improved the work. This research was supported in part by grants (11371299, 11571263, 11671311) from the National Science Foundation of China and a grant (17125814) from the Research Grants Council of Hong Kong.

Appendix: Theoretic Proofs

Proof of Theorem 1 Let

$$\widetilde{d}_k(t,z) = n^{-1}\sum_{i=1}^{n}\left[\left\{\frac{\Delta_i I(X_i > t)}{G(X_i)} - H(t)\right\} I(Z_{ik} < z)\right]$$

and define

$$\left\|\widetilde{d}_k\right\|_n^2 = n^{-1}\sum_{j=1}^{n}\widetilde{d}_k\left(X_j, Z_{jk}\right)^2.$$

Straightforward calculations entail that

$$\left\|\widetilde{d}_k\right\|_n^2 = \frac{(n-1)(n-2)}{n^2}\left(\frac{1}{n-2}\widetilde{D}_{k1} + \widetilde{D}_{k2}\right), \tag{A.1}$$

where

$$\begin{aligned}\widetilde{D}_{k1} &= \frac{2}{n(n-1)}\sum_{i<j}\frac{1}{2}\left[\left\{\frac{\Delta_i I(X_i > X_j)}{G(X_i)} - H(X_j)\right\}^2 I(Z_{ik} < Z_{jk})\right.\\ &\quad\left. + \left\{\frac{\Delta_j I(X_j > X_i)}{G(X_j)} - H(X_i)\right\}^2 I\left(Z_{jk} < Z_{ik}\right)\right]\\ &\equiv \frac{2}{n(n-1)}\sum_{i<j} h_1\left(\mathcal{O}_{ik};\mathcal{O}_{jk};G,H\right),\end{aligned}$$

$$\begin{aligned}\widetilde{D}_{k2} &= \frac{6}{n(n-1)(n-2)}\sum_{i<j<l}\frac{1}{3}\left[\left\{\frac{\Delta_i I(X_i > X_j)}{G(X_i)} - H(X_j)\right\}\left\{\frac{\Delta_l I(X_l > X_j)}{G(X_l)} - H(X_j)\right\}\right.\\ &\quad\times I(Z_{ik} < Z_{jk})I(Z_{lk} < Z_{jk}) + \left\{\frac{\Delta_i I(X_i > X_l)}{G(X_i)} - H(X_l)\right\}\left\{\frac{\Delta_j I(X_j > X_l)}{G(X_j)} - H(X_l)\right\}\\ &\quad\times I(Z_{ik} < Z_{lk})I(Z_{jk} < Z_{lk}) + \left\{\frac{\Delta_j I(X_j > X_i)}{G(X_j)} - H(X_i)\right\}\left\{\frac{\Delta_l I(X_l > X_i)}{G(X_l)} - H(X_i)\right\}\\ &\quad\left.\times I(Z_{jk} < Z_{ik})I(Z_{lk} < Z_{ik})\right]\\ &\equiv \frac{6}{n(n-1)(n-2)}\sum_{i<j<l} h_2\left(\mathcal{O}_{ik};\mathcal{O}_{jk};\mathcal{O}_{lk};G,H\right),\end{aligned}$$

$\mathcal{O}_{ik} = (X_i, \Delta_i, Z_{ik})$, and the definitions of kernels $h_1(\mathcal{O}_{ik}; \mathcal{O}_{jk}; G, H)$ and $h_2(\mathcal{O}_{ik}; \mathcal{O}_{jk}; \mathcal{O}_{lk}; G, H)$ in the U-statistics are clear from the context. Likewise, we have

$$\left\|\widehat{d}_k\right\|_n^2 = \frac{(n-1)(n-2)}{n^2}\left(\frac{1}{n-2}\widehat{D}_{k1} + \widehat{D}_{k2}\right), \tag{A.2}$$

where $\widehat{D}_{ks}, s = 1, 2$, are obtained by replacing G and H in $\widetilde{D}_{ks}$ with $\widehat{G}_n$ and $\widehat{H}_n$ respectively.

First, we derive the exponential tail probability bound of $P\left(\left|\|\widehat{d}_k\|_n^2 - \|\widetilde{d}_k\|_n^2\right| \geq \upsilon n^{-\alpha}\right)$ for any positive constants υ and $\alpha \in [0, 1/2)$. Consider $P(|\widehat{D}_{k1} - \widetilde{D}_{k1}| \geq \upsilon n^{-\alpha}/2)$ and note that

$$\begin{aligned}
|\widehat{D}_{k1} - \widetilde{D}_{k1}| &\leq \frac{2}{n(n-1)}\sum_{i<j}\left|h_1\left(\mathcal{O}_{ik}; \mathcal{O}_{jk}; \widehat{G}_n, \widehat{H}_n\right) - h_1\left(\mathcal{O}_{ik}; \mathcal{O}_{jk}; G, H\right)\right| \\
&\leq \frac{1}{n(n-1)}\sum_{i<j}\left[\left|\left\{\frac{\Delta_i I(X_i > X_j)}{\widehat{G}_n(X_i)} - \widehat{H}_n(X_j)\right\}^2 - \left\{\frac{\Delta_i I(X_i > X_j)}{G(X_i)} - H(X_j)\right\}^2\right|\right. \\
&\quad \left. + \left|\left\{\frac{\Delta_j I(X_j > X_i)}{\widehat{G}_n(X_j)} - \widehat{H}_n(X_i)\right\}^2 - \left\{\frac{\Delta_j I(X_j > X_i)}{G(X_j)} - H(X_i)\right\}^2\right|\right].
\end{aligned}$$

By condition C1 and the boundness of the indicator function, there exists a constant c_1 such that

$$\begin{aligned}
&\left|\left\{\frac{\Delta_i I(X_i > X_j)}{\widehat{G}_n(X_i)} - \widehat{H}_n(X_j)\right\}^2 - \left\{\frac{\Delta_i I(X_i > X_j)}{G(X_i)} - H(X_j)\right\}^2\right| \\
&\quad \leq c_1\left\{\left|\widehat{G}_n(X_i) - G(X_i)\right| + \left|\widehat{H}_n(X_j) - H(X_j)\right|\right\}.
\end{aligned}$$

Denoting $c_2 = \min\{G(\tau), H(\tau)\}$, we immediately have

$$\begin{aligned}
|\widehat{D}_{k1} - \widetilde{D}_{k1}| &\leq \frac{c_1}{n}\sum_{i=1}^{n}\left[\left|\widehat{G}_n(X_i) - G(X_i)\right| + \left|\widehat{H}_n(X_i) - H(X_i)\right|\right] \\
&\leq \frac{c_1}{c_2 n}\sum_{i=1}^{n}\left[\left|H(X_i)\left\{\widehat{G}_n(X_i) - G(X_i)\right\}\right| + \left|G(X_i)\left\{\widehat{H}_n(X_i) - H(X_i)\right\}\right|\right] \\
&\leq c_3 \sup_{0\leq t\leq\tau}\left|H(t)\left\{\widehat{G}_n(t) - G(t)\right\}\right| + c_3 \sup_{0\leq t\leq\tau}\left|G(t)\left\{\widehat{H}_n(t) - H(t)\right\}\right|,
\end{aligned} \tag{A.3}$$

where $c_3 = c_1/c_2$.

Using the similar argument, along with some tedious calculation, we also have

$$\left|\widehat{D}_{k2} - \widetilde{D}_{k2}\right| \leq c_4 \sup_{0\leq t\leq\tau}\left|H(t)\left\{\widehat{G}_n(t) - G(t)\right\}\right| + c_4 \sup_{0\leq t\leq\tau}\left|G(t)\left\{\widehat{H}_n(t) - H(t)\right\}\right|,$$

where c_4 is a constant. It follows from (A.3) and Theorem 1 of Bitouzé et al. (1999) that

$$P\left(\left|\widehat{D}_{k1}-\widetilde{D}_{k1}\right|\geq 2\upsilon n^{-\alpha}\right)\leq P\left(c_3\sup_{0\leq t\leq\tau}\left|H(t)\left\{\widehat{G}_n(t)-G(t)\right\}\right|\geq\upsilon n^{-\alpha}\right)$$
$$+P\left(c_3\sup_{0\leq t\leq\tau}\left|G(t)\left\{\widehat{H}_n(t)-H(t)\right\}\right|\geq\upsilon n^{-\alpha}\right)$$
$$\leq 5\exp\left(-2c_3^{-2}\upsilon^2 n^{1-2\alpha}+\mu_1 c_3^{-1}\upsilon n^{1/2-\alpha}\right), \quad \text{(A.4)}$$

where μ_1 is a constant. Similarly, we also have

$$P\left(\left|\widehat{D}_{k2}-\widetilde{D}_{k2}\right|\geq 2\upsilon n^{-\alpha}\right)\leq 5\exp\left(-2c_4^{-2}\upsilon^2 n^{1-2\alpha}+\mu_2 c_4^{-1}\upsilon n^{1/2-\alpha}\right), \quad \text{(A.5)}$$

where μ_2 is a constant. Combining (A.1), (A.2), (A.4) and (A.5), we have

$$\begin{aligned}
&P\left(\left|\|\widehat{d}_k\|_n^2-\|\widetilde{d}_k\|_n^2\right|\geq 4\upsilon n^{-\alpha}\right)\\
&=P\left\{\left|\frac{n-1}{n^2}\left(\widehat{D}_{k1}-\widetilde{D}_{k1}\right)+\frac{(n-1)(n-2)}{n^2}\left(\widehat{D}_{k2}-\widetilde{D}_{k2}\right)\right|\geq 4\upsilon n^{-\alpha}\right\}\\
&\leq P\left\{\left|\widehat{D}_{k1}-\widetilde{D}_{k1}\right|\geq 2\upsilon n^{1-\alpha}\right\}+P\left\{\left|\widehat{D}_{k2}-\widetilde{D}_{k2}\right|\geq 2\upsilon n^{-\alpha}\right\}\\
&\leq 5\exp\left(-2c_3^{-2}\upsilon^2 n^{3-2\alpha}+\mu_1 c_3^{-1}\upsilon n^{3/2-\alpha}\right)\\
&\quad+5\exp\left(-2c_4^{-2}\upsilon^2 n^{1-2\alpha}+\mu_2 c_4^{-1}\upsilon n^{1/2-\alpha}\right). \qquad \text{(A.6)}
\end{aligned}$$

Second, we derive the exponential tail probability bound of $P\left(\left|\|\widetilde{d}_k\|_n^2-\|d_k\|_n^2\right|\geq\upsilon n^{-\alpha}\right)$ for any positive constants υ and $0\leq\alpha<1/2$.

Note that $\|d_k\|_n^2=E\{h_2(\mathcal{O}_{ik};\mathcal{O}_{jk};\mathcal{O}_{lk};G;H)\}=E(\widetilde{D}_{k2})$. Employing the Markov inequality, we obtain that, for any $\epsilon>0$ and $\xi>0$,

$$P\left(\widetilde{D}_{k2}-\|d_k\|_n^2\geq\epsilon\right)\leq\exp(-\xi\epsilon)\exp\left(-\xi\|d_k\|_n^2\right)E\left\{\exp(\xi\widetilde{D}_{k2})\right\}.$$

Serfling (1980, Section 5.1.6) showed that any U-statistic can be represented as an average of averages of i.i.d. random variables. We can rewrite

$$\widetilde{D}_{k2}=(n!)^{-1}\Sigma_{n!}D_2\left(\mathcal{O}_{1k};\cdots;\mathcal{O}_{nk};G,H\right),$$

where $\Sigma_{n!}$ denotes the summation over all possible permutations of $(1,\ldots,n)$, and each $D_2(\mathcal{O}_{1k};\cdots;\mathcal{O}_{nk};G,H)$ is an average of $m\equiv[n/3]$ i.i.d. random variables. Denote $\psi(\xi)=E[\exp\{\xi h_2(\mathcal{O}_{ik};\mathcal{O}_{jk};\mathcal{O}_{lk};G,H)\}]$. Jensen's inequality yields that

$$
\begin{aligned}
E\left\{\exp\left(\xi\widetilde{D}_{k2}\right)\right\} &= E\left[\exp\left\{\xi(n!)^{-1}\Sigma_{n!}D_2\left(\mathcal{O}_{1k};\cdots;\mathcal{O}_{nk};G,H\right)\right\}\right] \\
&\le (n!)^{-1}\Sigma_{n!}E\left[\exp\left\{\xi D_2\left(\mathcal{O}_{1k};\cdots;\mathcal{O}_{nk};G,H\right)\right\}\right] \\
&= \psi^m(\xi/m).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
&P\left(\widetilde{D}_{k2} - \|d_k\|_n^2 \ge \epsilon\right) \le \exp(-\xi\epsilon)\exp\left(-\xi\,\|d_k\|_n^2\right)\psi^m(\xi/m) \\
&= \exp(-\xi\epsilon)\left\{E\left(\exp\left[m^{-1}\xi\left\{h_2\left(\mathcal{O}_{ik};\mathcal{O}_{jk};\mathcal{O}_{lk};G,H\right) - \|d_k\|_n^2\right\}\right]\right)\right\}^m.
\end{aligned}
$$

Under condition C1, there exists a positive constant c_5 such that $P(|h_2| < c_5) = 1$. It follows from Lemma 1 in Li et al. (2012) that

$$
E\left\{\exp\left[m^{-1}\xi\left\{h_2\left(\mathcal{O}_{ik};\mathcal{O}_{jk};\mathcal{O}_{lk};G,H\right) - \|d_k\|_n^2\right\}\right]\right\} \le \exp\left\{c_5^2\xi^2/(2m^2)\right\},
$$

which immediately entails that

$$
P\left(\widetilde{D}_{k2} - \|d_k\|_n^2 \ge \epsilon\right) \le \exp\left(-\frac{\epsilon^2 m}{2c_5^2}\right),
$$

by choosing $\xi = \epsilon m/c_5^2$. It further follows from the symmetry of the U-statistic that

$$
P\left(\left|\widetilde{D}_{k2} - \|d_k\|_n^2\right| \ge \epsilon\right) \le 2\exp\left(-\frac{\epsilon^2 m}{2c_5^2}\right).
$$

Using the similar argument, we also have

$$
P\left(\left|\widetilde{D}_{k1} - E(\widetilde{D}_{k1})\right| \ge \epsilon\right) \le 2\exp\left(-\frac{\epsilon^2 m^*}{2c_6^2}\right),
$$

where c_6 is a positive constant such that $P(|h_1| < c_6) = 1$ and $m^* = [n/2]$. Obviously, under condition C1, there exist constants c_7 and c_8 such that $0 \le \|d_k\|_n^2 = E(\widetilde{D}_{k2}) \le E|\widetilde{D}_{k2}| \le c_7$ and $0 \le E(\widetilde{D}_{k1}) \le E|\widetilde{D}_{k1}| \le c_8$ for any $1 \le k \le p_n$. Taking $\epsilon = \upsilon n^{-\alpha}$ and n large enough such that $(3n-2)n^{-2}E(\widetilde{D}_{k2}) < \upsilon n^{-\alpha}$ and $(n-1)n^{-2}E(\widetilde{D}_{k1}) < \upsilon n^{-\alpha}$, we have

$$
\begin{aligned}
&P\left(\left|\left\|\widetilde{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge 4\upsilon n^{-\alpha}\right) \\
&= P\left\{\left|\frac{(n-1)(n-2)}{n^2}\left(\widetilde{D}_{k2} - \|d_k\|_n^2\right) - \frac{3n-2}{n^2}E(\widetilde{D}_{k2})\right.\right. \\
&\quad \left.\left. + \frac{n-1}{n^2}\left\{\widetilde{D}_{k1} - E\left(\widetilde{D}_{k1}\right)\right\} + \frac{n-1}{n^2}E(\widetilde{D}_{k1})\right| \ge 4\upsilon n^{-\alpha}\right\}
\end{aligned}
$$

$$\begin{aligned}
&\le P\left(\left|\widetilde{D}_{k1} - E(\widetilde{D}_{k1})\right| \ge \upsilon n^{1-\alpha}\right) + P\left(\left|\widetilde{D}_{k2} - \|d_k\|_n^2\right| \ge \upsilon n^{-\alpha}\right) \\
&\le 2\exp\left(-\frac{\upsilon^2 n^{2-2\alpha} m^*}{2c_6^2}\right) + 2\exp\left(-\frac{\upsilon^2 n^{-2\alpha} m}{2c_5^2}\right) \\
&\le 2\exp\left(-c_9\upsilon^2 n^{3-2\alpha}\right) + 2\exp\left(-c_{10}\upsilon^2 n^{1-2\alpha}\right), \qquad \text{(A.7)}
\end{aligned}$$

by noting that $m^* \ge m \ge n/4$, where $c_9 = 1/(8c_6^2)$ and $c_{10} = 1/(8c_5^2)$. It follows from (A.6) and (A.7) that

$$\begin{aligned}
&P\left(\left|\left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge 8\upsilon n^{-\alpha}\right) \\
&\le P\left(\left|\left\|\widehat{d}_k\right\|_n^2 - \left\|\widetilde{d}_k\right\|_n^2\right| \ge 4\upsilon n^{-\alpha}\right) + P\left(\left|\left\|\widetilde{d}_k\right\|_n^2 - \left\|d_k\right\|_n^2\right| \ge 4\upsilon n^{-\alpha}\right) \\
&\le 5\exp\left(-2c_3^{-2}\upsilon^2 n^{3-2\alpha} + \mu_1 c_3^{-1}\upsilon n^{3/2-\alpha}\right) \\
&\quad +5\exp\left(-2c_4^{-2}\upsilon^2 n^{1-2\alpha} + \mu_2 c_4^{-1}\upsilon n^{1/2-\alpha}\right) \\
&\quad +2\exp\left(-c_9\upsilon^2 n^{3-2\alpha}\right) + 2\exp\left(-c_{10}\upsilon^2 n^{1-2\alpha}\right) \\
&\le O\left\{\exp\left(-\eta n^{1-2\alpha}\right)\right\}, \qquad \text{(A.8)}
\end{aligned}$$

where $\eta = \min\{2c_4^{-2}\upsilon^2, c_{10}\upsilon^2\}$. Immediately, we have

$$P\left(\max_{1\le k\le p_n}\left|\left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge 8\upsilon n^{-\alpha}\right) \le O\left\{p_n\exp\left(-\eta n^{1-2\alpha}\right)\right\}, \qquad \text{(A.9)}$$

which proves the first part of Theorem 1 by taking $c = 8\upsilon$.

If $\mathcal{A} \not\subseteq \widehat{\mathcal{A}}$, then there must exist some $k \in \mathcal{A}$ such that $\|\widehat{d}_k\|_n^2 < cn^{-\alpha}$. It follows from condition C2 that $|\|\widehat{d}_k\|_n^2 - \|d_k\|_n^2| > cn^{-\alpha}$ for some $k \in \mathcal{A}$, which implies that $\{\mathcal{A} \not\subseteq \widehat{\mathcal{A}}\} \subseteq \{|\|\widehat{d}_k\|_n^2 - \|d_k\|_n^2| > cn^{-\alpha}$ for some $k \in \mathcal{A}\}$. As a result, $\{\max_{k\in\mathcal{A}}|\|\widehat{d}_k\|_n^2 - \|d_k\|_n^2| \le cn^{-\alpha}\} \subseteq \{\mathcal{A} \subseteq \widehat{\mathcal{A}}\}$. Using (A.8), we have

$$\begin{aligned}
P\big(\mathcal{A} \subseteq \widehat{\mathcal{A}}\big) &\ge P\Big(\max_{k\in\mathcal{A}}\big|\big\|\widehat{d}_k\big\|_n^2 - \big\|d_k\big\|_n^2\big| \le cn^{-\alpha}\Big) \\
&\ge 1 - O\big\{a_n\exp\big(-\eta n^{1-2\alpha}\big)\big\},
\end{aligned}$$

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

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

$$\begin{aligned}
d_k(t,z) &= E\left[\left\{\frac{\Delta I(X>t)}{G(X)} - P(T>t)\right\} I(Z_k < z)\right] \\
&= E\left\{E\left[\left\{\frac{\Delta I(X>t)}{G(X)} - P(T>t)\right\} I(Z_k < z)\,\middle|\,\mathbf{Z}\right]\right\}
\end{aligned}$$

$$= E\left\{I(Z_k < z)E\left[\left\{\frac{\Delta I(X > t)}{G(X)} - P(T > t)\right\}\middle|\mathbf{Z}\right]\right\}$$
$$= E\left\{I(Z_k < z)\left[E\left\{\frac{\Delta I(X > t)}{G(X)}\middle|\mathbf{Z}_{\mathcal{A}}\right\} - P(T > t)\right]\right\}.$$

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

$$d_k(t, z) = E\{I(Z_k < z)\}E\left[E\left\{\frac{\Delta I(X > t)}{G(X)}\middle|\mathbf{Z}_{\mathcal{A}}\right\} - P(T > t)\right] = 0,$$

for any t and z. As a result, $\|d_k\|_n^2 = E\{d_k(X, Z_k)^2\} = 0$. It follows from condition C2 that $\max_{k\notin\mathcal{A}} \|d_k\|_n^2 < \min_{k\in\mathcal{A}} \|d_k\|_n^2$. On the other hand, $\|d_k\|_n^2 = 0$ directly implies that $k \notin \mathcal{A}$ under condition C2. Thus, the first part of Theorem 2 is proved.

Under condition C2 and assumptions (i) and (ii), coupled with (A.9), we have

$$\begin{aligned}
&P\left(\min_{k\in\mathcal{A}} \left\|\widehat{d}_k\right\|_n^2 \le \max_{k\notin\mathcal{A}} \left\|\widehat{d}_k\right\|_n^2\right)\\
&= P\left(\max_{k\notin\mathcal{A}} \left\|\widehat{d}_k\right\|_n^2 - \max_{k\notin\mathcal{A}} \|d_k\|_n^2 - \min_{k\in\mathcal{A}} \left\|\widehat{d}_k\right\|_n^2 + \min_{k\in\mathcal{A}} \|d_k\|_n^2 \ge \min_{k\in\mathcal{A}} \|d_k\|_n^2\right)\\
&\le P\left(\max_{k\notin\mathcal{A}} \left|\left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge cn^{-\alpha}\right) + P\left(\max_{k\in\mathcal{A}} \left|\left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge cn^{-\alpha}\right)\\
&\le 2P\left(\max_{1\le k\le p_n} \left|\left\|\widehat{d}_k\right\|_n^2 - \|d_k\|_n^2\right| \ge cn^{-\alpha}\right)\\
&\le O\left\{p_n \exp\left(-\eta n^{1-2\alpha}\right)\right\},
\end{aligned}$$

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. Annales de l'Institut Henri Poincare (B) Probab Stat 35:735–763

Candes E, Tao T (2007) The Dantzig selector: statistical estimation when p is much larger than n. Ann Stat 35:2313–2351

Cook AJ, Gold DR, Li Y (2007) Spatial cluster detection for censored outcome data. Biometrics 63:540–549

Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96:1348–1360

Fan J, Lv J (2008) Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Ser B 70:849–911

Fan J, Song R (2010) Sure independence screening in generalized linear models with NP-dimensionality. J Am Stat Assoc 38:3567–3604

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

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, Institute of Mathematical Statistics 6:70–86

Fan J, Feng Y, Song R (2011) Nonparametric independence screening in sparse ultra-high-dimensional additive models. J Am Stat Assoc 106:544–557

Gorst-Rasmussen A, Scheike T (2013) Independent screening for single-index hazard rate models with ultrahigh dimensional features. J R Stat Soc Ser B 75:217–245
Hoeffding W (1948) A non-parametric test of independence. Ann Math Stat 19:546–557
Lin DY, Wei LJ, Ying Z (1993) Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 80:557–572
Li R, Zhong W, Zhu L (2012) Feature screening via distance correlation learning. J Am Stat Assoc 107:1129–1139
Rosenwald A, Wright G, Wiestner A, Chan WC 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
Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
Song R, Lu W, Ma S, Jeng XJ (2014) Censored rank independence screening for high-dimensional survival data. Biometrika 101:799–814
Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B 58:267–288
Tibshirani R (2009) Univariate shrinkage in the Cox model for high dimensional data. Stat Appl Genet Mol Biol 8:1–18
Wu Y, Yin G (2015) Conditional quantile screening in ultrahigh-dimensional heterogeneous data. Biometrika 102:65–76
Zhang CH (2010) Nearly unbiased variable selection under minimax concave penalty. Ann Stat 38:894–942
Zhao SD, Li Y (2012) Principled sure independence screening for Cox models with ultra-high-dimensional covariates. J Multivar Anal 105:397–411
Zhu LP, Li L, Li R, Zhu LX (2011) Model-free feature screening for ultrahigh dimensional data. J Am Stat Assoc 106:1464–1475
Zou H (2006) The adaptive Lasso and its oracle properties. J Am Stat Assoc 101:1418–1429