Censored Rank Independence Screening for High-dimensional Survival Data

Rui Song,
Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA

Wenbin Lu,
Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA

Shuangge Ma, and
Division of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut 06510, USA

X. Jessie Jeng
Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA

Rui Song: rsong@ncsu.edu; Wenbin Lu: lu@stat.ncsu.edu; Shuangge Ma: shuangge.ma@yale.edu; X. Jessie Jeng: xjjeng@ncsu.edu

Summary

In modern statistical applications, the dimension of covariates can be much larger than the sample size. In the context of linear models, correlation screening (Fan and Lv, 2008) has been shown to reduce the dimension of such data effectively while achieving the sure screening property, i.e., all of the active variables can be retained with high probability. However, screening based on the Pearson correlation does not perform well when applied to contaminated covariates and/or censored outcomes. In this paper, we study censored rank independence screening of high-dimensional survival data. The proposed method is robust to predictors that contain outliers, works for a general class of survival models, and enjoys the sure screening property. Simulations and an analysis of real data demonstrate that the proposed method performs competitively on survival data sets of moderate size and high-dimensional predictors, even when these are contaminated.

Some key words

High-dimensional survival data; Rank independence screening; Sure screening property

1. Introduction

Our study was motivated by a breast cancer data (van Houwelingen et al., 2006) that contains expression profiles of 24885 candidate genes for 295 patients with breast cancer. The primary interest was to find genes that are predictive for the overall survival time of breast cancer patients. In addition to their dimensionality being large, some predictors are not normally distributed and contain outliers; see the Supplementary Material. These

phenomena are common in microarray data. High dimensionality and the existence of outliers make variable selection for censored survival data challenging.

There are numerous studies in the literature regarding variable selection for regression problems with and without censoring. Recently, many studies have focused on penalized methods, such as the lasso (Tibshirani, 1996), the smoothly clipped absolute deviation (Fan & Li, 2001), the Dantzig selector (Candes & Tao, 2007), and their variants. These methods have been thoroughly studied for variable selection with high-dimensional data (e.g., Bickel et al., 2009; Meinshausen & Yu, 2009; van de Geer, 2008). Studies of variable selection for survival outcomes include penalized partial likelihood (Fan & Li, 2002; Tibshirani, 1997; Zhang & Lu, 2007), penalized estimating equations (Johnson, 2008; Johnson et al., 2008), and other approaches that can be used for simultaneous variable selection and estimation. Generally, the associated optimization problems may be solved quickly for moderate to large p, such as p being hundreds or thousands. However, for very large p, such as is encountered in microarray data, these methods remain computationally demanding.

A computationally simple method for very high-dimensional data that can work well in practice is sure independence screening, which was demonstrated in the classical regression context in Fan & Lv (2008). In this method, the outcome variable is regressed on each covariate separately. Sure independence screening recruits the features that have the best marginal utility. In the context of least-squares regression for a linear model, this corresponds to the largest marginal absolute Pearson correlation between the response and the predictor. Fan & Lv (2008) showed that this method has a sure screening property: with probability very close to 1, the method can retain all of the important features in the model. It can also be derived from an empirical likelihood point of view (Hall et al., 2009). Correlation screening is a crude yet effective way to decrease the dimensionality of data. However, the Pearson correlation might not work well for censored survival data because it cannot be estimated reliably, especially when the censoring rate is high. In addition, its performance can be ruined by outliers in predictors because correlation is not a robust measure for association. Such outliers cause trouble for theoretical studies of screening methods, most of which require tail probability conditions for the covariates.

Variable screening methods for high-dimensional survival data are mostly based on the partial-likelihood of the Cox model. For example, Tibshirani (2009) used a lasso penalization approach for pre-screening. Zhao & Li (2012) proposed a screening method based on standardized marginal maximum partial likelihood estimators. However, in practice, the true models often remain unknown, and it is unclear if these methods will work well under model misspecification. Gorst-Rasmussen & Scheike (2013) proposed a model-free screening statistic: the feature aberration at survival times. For each covariate, this new statistic is equivalent to the numerator of the marginal log-rank test. These screening methods might be influenced by outliers in predictors.

In this paper, we propose a censored rank independence screening method for high-dimensional survival data. The rank statistic we consider can be viewed as an inverse probability-of-censoring weighted Kendall's τ (Kendall, 1962). Our proposed method has several advantages. First, it is robust against the existence of outliers. This robustness is

inherited from Kendall's τ coefficient (Sen, 1968). Second, it is a non model-based method, so it works for a wide class of survival models. In contrast to Pearson's correlation, Kendall's τ is invariant under monotonic transformations of responses and predictors. This invariance allows our method to discover any nonlinear relationships between the response and predictors. Third, the proposed method has technical improvements over some other high-dimensional methods, as the proposed screening utility is a U-statistic with a bounded kernel function, which enables us to obtain the sure screening property without requiring tail probability conditions.

2. Censored Rank Independence Screening

Let T denote the event time of interest, C denote the censoring time, and $X = (X_{(1)}, \ldots, X_{(p)})'$ denote the p-dimensional vector of covariates. Further, define $V = \min(T, C)$ and $\Delta = I(T \leq C)$. where $I(\cdot)$ denotes the indicator function. The observed data are independent and identically distributed copies of $W = (X, V, \Delta)$ and are denoted by $W_i = (X_i, V_i, \Delta_i)$ for $i = 1, \ldots, n$, where $X_i = (X_{1i}, \ldots, X_{pi})'$. Throughout the paper, it is assumed that the censoring time, C, is independent of the event time, T, and the covariates, X. Let $\mathscr{A} = \{1, \ldots, p\}$ and $X_{\mathscr{B}} = \{X_{(j)} : j \in \mathscr{B}\}$ for a set $\mathscr{B} \subset \mathscr{A}$. Let $\mathscr{M}_\star$ denote the index set of the active variables:

$$\mathscr{M}_\star = \{k : \mathrm{pr}(T > t \mid X) \text{ functionally depends on } X_{(k)}\}.$$

Our goal is to select the set of active variables, $X_{\mathscr{M}_\star}$, where $\mathscr{M}_\star \subset \mathscr{A}$.

We consider the following inverse probability-of-censoring weighted marginal rank correlation utility,

$$\hat{\tau}_k = \binom{n}{2}^{-1} \sum_{i<j} \frac{\Delta_j}{\hat{S}^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) - 1/4, \quad k = 1, \ldots, p.$$

where $\hat{S}(\cdot)$ is the Kaplan–Meier estimator of $S(t) = \mathrm{pr}(C \geq t)$. We define $0/0 = 0$ to make $\hat{\tau}_k$ well-defined. For a prespecified γ_n, we select the set

$$\hat{\mathscr{M}} = \{k : |\hat{\tau}_k| \geq \gamma_n\}$$

as active variables. In this way, the dimension of the covariates used in the model can be reduced to a value much smaller than n.

Let $\tau_k = \mathrm{pr}(X_{ki} > X_{kj}, T_i > T_j) - 1/4$. It can be shown that

$$E\left\{\frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j)\right\} = \mathrm{pr}(X_{ki} > X_{kj}, T_i > T_j).$$

and it follows that $\hat{\tau}_k$ provides a consistent estimate of τ_k. Without assuming any particular model structure, such as the proportional hazards model, the set selected by the proposed

censored rank independence screening comprises the variables that have strong marginal rank correlation with the failure time. In the next section we show that the proposed method enjoys the sure screening property under general conditions.

3. Sure Screening Property

Let $\{C_i, V_i, X_{ki}\}$ $(i = 1, 2)$ be independent and identically distributed copies of $\{C, V, X_{(k)}\}$, for $k = 1, \ldots, p$. The following conditions are required:

A. there exists a $\nu > 0$ such that $\mathrm{pr}(C = \nu) > 0$ and $\mathrm{pr}(C > \nu) = 0$;

B. $\min_{k \in \mathcal{M}_\star} |\mathrm{pr}(X_{k1} > X_{k2}, T_1 > T_2) - 1/4| \geq c_0 n^{-\kappa}$ for some $0 < \kappa < 1/2$ and $c_0 > 0$.

Condition A, adopted from Peng & Fine (2009), is a technical condition that simplifies the derivation of asymptotic properties. Because Condition A is satisfied in many clinical settings, it is widely used in the literature. Condition B is a key assumption to ensure the sure screening property, even without assuming specific model forms. This indicates that to ensure the sure screening property, the minimal marginal rank correlation between the active variables and the response variable should exceed a certain threshold.

Theorem 1. *Under Condition A, for any positive constants $c_5 \leq c_6$, when*

$$n > \max\left\{ D^2(1-\delta)^{-2}\{(3/2)^{1/2} - 1\}^{-2}\|S\|_\infty^{-2},\, 25D^2 c_5^2 n^{2\kappa}(1-\delta)^{-2}\|S\|_\infty^{-2},\, (c_5/1.12)^{1/\kappa}\right\},$$

there exist constants c_1, c_2, and c_4 such that

$$\mathrm{pr}\left(\max_{1 \leq k \leq p} |\hat{\tau}_k - \tau_k| > c_6 n^{-\kappa} \right) \leq p\left\{ 2.5n\exp(-c_1 n) + 4\exp(-c_4 n^{1-2\kappa}) + 2.5n\exp(-c_2 n^{1-2\kappa}) \right\},$$

where $\|\cdot\|_\infty$ is the L_∞ norm, and D is a constant introduced in Lemma 1 of the Appendix. *Moreover, when Condition B holds, taking $\gamma_n = c_7 n^{-\kappa}$ with $c_7 \leq c_0/2$ leads to*

$$\mathrm{pr}\left(\mathcal{M}_\star \subset \hat{\mathcal{M}} \right) \geq 1 - s\left\{ 2.5n\exp(-c_1 n) + 4\exp(-c_4 n^{1-2\kappa}) + 2.5n\exp(-c_2 n^{1-2\kappa}) \right\},$$

where s is the number of variables in $\mathcal{M}_\star$.

The first result of Theorem 1 leads to the conditions whereby the sure screening property of our method is ensured. Specifically, as n goes to infinity, the maximum dimensionality is $p = o\{\exp(n^{1-2\kappa})\}$, for $\kappa \in (0, 1/2)$. This limit is of the same order as that obtained in Fan and Lv (2008) for correlation learning in the linear model set-up, and it is stronger than the result obtained in Fan et al. (2010). Because no tail probability conditions for covariates are needed, the conditions for the sure screening property are more relaxed than those in Fan & Lv (2008) and Fan et al. (2010). Therefore, our method generally allows heavy tailed covariates. Moreover, the method is robust to model misspecification because no model assumptions are required for the sure screening property to hold. In the next section, we apply the proposed method to a general class of transformation models, under which a set of sufficient conditions for showing Condition B will be provided and the size of the set $\hat{\mathcal{M}}$ can be controlled.

The threshold, γ_n, controls how many covariates pass screening. To ensure the sure screening property, γ_n can be taken as any value that is smaller than the minimum signal, provided the minimum signal is distinguishable from the estimation noise. Model selection consistency, i.e., $\text{pr}(\hat{\mathcal{M}} = \mathcal{M}_\star) = 1 - o(1)$, can be achieved if there is a gap between signal variables and noise variables. In our case, a sufficient condition for model selection consistency is that $X_{\mathcal{M}_\star}$ and $X_{\mathcal{M}_\star^c}$ are independent. Then

$$\text{pr}(T_1 > T_2, X_{j1} > X_{j2}) = E\big[E\{1(T_1 > T_2, X_{j1} > X_{j2})|X_{\mathcal{M}_\star}\}\big] = 1/4, j \in \mathcal{M}_\star^c.$$

4. Selection of the important set

In applications, it is common practice to select a prefixed number of top-ranked variables for follow-up study. The prefixed number may reflect researchers' prior knowledge of the number of susceptible predictors, or budget limitations (Kuo & Zaykin, 2011; Skol et al., 2006). Another commonly used procedure is to set the size of $\hat{M}$ to a number less than the sample size, so that follow-up regression analysis can be performed in a $p < n$ scenario (Fan and Lv, 2008). Data-driven procedures for selecting the size of the important set based on screening statistics are appealing but relatively limited. Zhao and Li (2012) proposed a principled selection method based on controlling false positive rate, but it can be conservative for screening purposes because controlling the false positive rate at a low level can lead to large false negative error.

We propose to estimate the size of the important set using a technique developed in the multiple testing literature. Specifically, consider the hypotheses $H_{0k} : \tau_k = 0$ and $H_{ak} : \tau_k \neq 0$ ($k = 1, \ldots, p$). Under H_{0k}, we can show that $n^{1/2}\hat{\tau}_k$ converges in distribution to a mean-zero normal random variable, and its asymptotic variance can be consistently estimated using U-statistic techniques similar to those studied in Fine et al. (1998). Let $\hat{\sigma}_k^2$ denote the estimated asymptotic variance of $\hat{\tau}_k$. Then, the p-value for testing H_{0k} can be computed as $q_k = 2\{1 - \Phi(|\hat{\tau}_k|/\hat{\sigma}_k)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Order the p-values as $q_{(1)} \leq \cdots \leq q_{(p)}$. Let $|\mathscr{A}|$ denote the size of a set $\mathscr{A}$. The proportion of true signals is $\pi = |\mathcal{M}_*|/p$. For a large number of independently tested hypotheses, Meinshausen et al. (2006) showed that π can be consistently estimated by

$$\hat{\pi} = \max_{i=1,\ldots,p/2} \frac{i/p - q_{(i)} - (2\log\ \log p/p)^{1/2}\{q_{(i)}(1 - q_{(i)})\}^{1/2}}{1 - q_{(i)}}. \quad (1)$$

However, for general dependent test statistics, such as our proposed censored rank screening statistics $\hat{\tau}_k$'s, the consistency of $\hat{\pi}$ is usually unclear. In this paper, we use $\hat{\pi}$ as an estimator of π and set $|\hat{\mathcal{M}}| = \hat{\pi}p$. We study the empirical performance of this estimator in Section 6.

5. Application to a general class of transformation models

Although the sure screening property of our method does not depend on the specific modeling form, the active set, $\mathcal{M}_\star$, is not easily specified without assuming a model structure. To benefit from aspects of both the model-based and the model-free approaches, it is helpful to consider a wide class of models that contains the underlying true model. Here,

we consider a general class of transformation models, under which the active set, $\mathcal{M}_\star$, can be easily specified, and the sure screening property will hold.

Specifically, the general class of transformation models is given by

$$H(T)=m(X)+\varepsilon, \quad (2)$$

where $H(\cdot)$ is an increasing transformation function, $m(\cdot)$ is monotone in each element of X, and ε is independent of X and has a continuous distribution function. Under model (2), the conditional survival function takes the form

$$\mathrm{pr}(T>t|X)=S_\varepsilon\{H(t)-m(X)\}, \quad (3)$$

where $S_\varepsilon(\cdot)$ is the survival function of ε. This class of transformation models includes many popular survival models as special cases. For example, when $H(\cdot)$ is unknown, $S_\varepsilon(\cdot)$ is specified, and $m(X) = \beta'X$, model (2) becomes the linear transformation model (Clayton & Cuzick, 1985). This model includes the proportional hazards and proportional odds models as special cases. When H is the log transformation, $S_\varepsilon(\cdot)$ is unspecified, and $m(X) = \beta'X$, model (2) becomes the accelerated failure time model (Kalbfleisch & Prentice, 2002). Other examples of transformation models include the odds-rate, inverse Gaussian and log-normal families (Kosorok et al., 2004; Scharfstein et al., 1998).

For transformation models with $m(X) = \beta'X$, it is clear that $\mathcal{M}_\star = \{j \in \mathcal{A} : \beta_j \neq 0\}$, where $\beta = (\beta_1, \ldots, \beta_p)'$. In general, $\mathcal{M}_\star$ can be defined as the smallest subset of $\mathcal{A}$ such that $m(\cdot)$ is only a function of covariates in $\mathcal{M}_\star$, i.e., the transformation models can be equivalently written as $H(T) = m(X_{\mathcal{M}_\star}) + \varepsilon$. Define $s_n = |\mathcal{M}_\star|$, the number of active variables in $\mathcal{M}_\star$, and $\mathcal{M}_\star^c = \mathcal{A} / \mathcal{M}_\star$.

For $k \in \mathcal{M}_\star$, define $m_k(x) = E\{m(x, X_{\mathcal{M}_\star/k})\}$, where the expectation is taken with respect to the joint distribution of covariates in $\mathcal{M}_\star/k$ with $X_{(k)}$ fixed at x. Without loss of generality, $X_{(k)}$ is assumed to have mean 0 and variance 1. The following conditions are sufficient to show Condition B for a general class of transformation models.

C1 For any $k \in \mathcal{M}_\star$, the conditional density function of $H(T_1) - H(T_2) - \{m_k(X_{k1}) - m_k(X_{k2})\}$ given $m_k(X_{k1}) - m_k(X_{k2})$ is unimodal and symmetric around zero.

C2 For any $k \in \mathcal{M}_\star$, there exist positive constants σ_1 and σ_2 such that the variance of $m_k(X_{k1})$ is uniformly bounded above by σ_1^2, and the conditional variance of $H(T_1) - H(T_2) - \{m_k(X_{k1}) - m_k(X_{k2})\}$ given $m_k(X_{k1}) - m_k(X_{k2})$ is uniformly bounded above by σ_2^2.

C3 For any $k \in \mathcal{M}_\star$, there exists a positive constant, d_0, that is independent of p such that $\min_k E\{|m_k(X_{(k)}) - Em_k(X_{(k)})|\} \geq d_0 n^{-\kappa/2}$ for $0 < \kappa < 1/2$.

Proposition 1. *If Conditions C1–C3 hold, then Condition B holds for some $c_0 > 0$.*

As $m(\cdot)$ is monotone in each element of $\mathcal{M}_\star$, $m_k(x)$ is monotone in x. As a marginal projection of $m(\cdot)$ onto the univariate dimension of X_k, $m_k(x)$ is utilized as a parsimonious way to pass along the monotonicity from the joint model to the marginal model. As

technical conditions, C1 and C2 can be checked empirically. Condition C3 states that the least absolute deviation of $m_k(X_{(k)})$ in the active set, $\mathcal{M}_\star$, can serve as the measurement of detectable signals for transformation models.

Next, we show that the size of set $\hat{\mathcal{M}}$ can be controlled for linear transformation models with $m(X) = \beta'X$. This result is similar in rationale to the result of Theorem 5 in Fan and Song (2010). The following condition, C4, is needed.

C4   For $k \in \mathcal{M}_\star$, the conditional density function of $H(T_1) - H(T_2) - (X_{k1} - X_{k2})EX_{(k)}Y$ given $X_{k1} - X_{k2}$ is unimodal and symmetric around zero.

Theorem 2. *Under Conditions A and C1–C4, when $var\{H(T)\} = O(1)$, for $\gamma_n = c_7 n^{-\kappa}$, there exist positive constants c_1, c_2, and c_4 such that*

$$\mathrm{pr}[|\hat{\mathcal{M}}| \leq O\{\gamma_n^{-2}\lambda_{\max}(\Sigma)\}] \geq 1 - p\left\{2.5n\exp(-c_1 n) + 4\exp(-c_4 n^{1-2\kappa}) + 2.5n\exp(-c_2 n^{1-2\kappa})\right\},$$

where Σ is the covariance matrix of X and $\lambda_{\max}(\Sigma)$ is its largest eigenvalue.

Taking the choice of γ_n as in Theorem 1, if the largest eigenvalue of the covariance matrix of $X = O(n^\tau)$ is of polynomial order for some $\tau > 0$, then the size of $\hat{\mathcal{M}}$ is also of polynomial order $O(n^{2\kappa+\tau})$ according to Theorem 2. This indicates that the size of the selected set can indeed be effectively controlled.

6. Simulation Studies

We conducted simulations to evaluate the empirical performance of the proposed censored rank independence screening method. For comparison, we considered three alternative methods: feature aberration at survival times screening (Gorst-Rasmussen & Scheike, 2013), partial likelihood ratio screening, and correlation screening. For partial likelihood ratio screening, we fit a marginal Cox model for each covariate and constructed the corresponding partial likelihood ratio statistic versus the no covariates model. This method is asymptotically equivalent to the screening method proposed by Zhao & Li (2012). For correlation screening, we used uncensored data to compute the marginal correlation between event time and covariate using an inverse probability-of-censoring weighted method. This generalizes standard sure independence screening for linear regression to survival data.

The failure time, T_i, was generated from the class of linear transformation models

$$H(T_i) = -\beta'X_i + \varepsilon_i, i = 1, \ldots, n,$$

where $H(t) = \log\{0.5(e^{2t} - 1)\}$, and X_i is a p-dimensional vector of covariates. We set $n = 100$, 300 and $p = 5000$, 10000. The covariates, X_i, were generated from a multivariate normal distribution with mean 0, variance 1, and a first order autoregressive structure, i.e., $corr(X_{ij}, X_{ik}) = 0.5^{|j-k|}$ ($j, k = 1, \ldots, p$). We considered two scenarios for the true regression coefficients: Scenario 1, $\beta = (-1, -0.9, 0_6', 0.8, 1.0, 0_{p-10}')'$ and Scenario 2,

$$\beta = (-1, -0.9, 0.5, 0.8, 0.6, 0_5', 0.3, 0.7, -0.8, -0.5, -1.0, 0_5', -0.3, 0.6, 0.8, -0.7, -0.9, 0_{p-25}')'.$$

Scenario 2 is more challenging than Scenario 1 because there are several active variables with relatively small effects. We considered three error distributions: the standard extreme value distribution, which corresponds to a proportional hazards model; the standard logistic distribution, which corresponds to a proportional odds model; and the standard normal distribution, which corresponds to a normal transformation model. The censoring time was generated from a uniform distribution on $[0, c]$, where the constant, c, was chosen to achieve censoring proportions of 15% and 40%. For each setting, we conducted 100 simulation runs. The simulation results for normal covariates are given in the Supplementary Material. Based on the results, the selection performances of the proposed method and partial likelihood ratio screening were comparable for the Scenario 1 settings and for the 15% censoring rate setting in Scenario 2. The performance of partial likelihood ratio screening was slightly better than the proposed method for Scenario 2 when the censoring rate was high, i.e., 40%. Generally, correlation screening performed poorly relative to both the proposed method and partial likelihood ratio screening. Correlation screening became very poor for Scenario 2 when the censoring rate was high.

To study the performances when the covariates might be contaminated by outliers, we added outliers to the covariates. All other settings were unchanged. Specifically, with a probability of 0.1, each covariate was replaced by a random variable generated from a t distribution. Again, we conducted 100 simulation runs. For Scenario 1, we report the average number of active variables contained in the top 4, 10, 20, 30, 40, and 50 selected variables, denoted by true positive; the true number was 4. For Scenario 2, we report the corresponding number in the top 15, 30, 45, 60, 75, 90, 120, and 150 selected variables, and the true number was 15. The results for the proportional hazards model under Scenarios 1 and 2 are summarized in Figures 1 and 2, respectively. Results for the proportional odds model are similar, and are provided in the Supplementary Material.

Under all settings, the selection performance of the proposed method was very similar to that when the covariates did not contain outliers, so it is robust to outliers in covariates. In addition, the selection performance improved when the sample size increased and the censoring rate decreased, though the performances when $p = 5000$ and $p = 10000$ were similar. For most settings, the performance of the proposed method was superior to both partial likelihood ratio screening and the feature aberration at survival times statistic. However, these three methods had comparable performances when $n = 100$ and the censoring rate was 40% in Scenario 2. Generally, partial likelihood ratio screening and feature aberration at survival times screening had comparable performances. In addition, the censoring proportion had less effect on their performances than on the performance of the proposed method. This behavior is expected because the proposed method uses the inverse probability-of-censoring weighted technique to address censoring, which might lose some efficiency when the censoring proportion is high. As was the case for normal covariates, correlation screening had the poorest selection performance.

Next, we examine the performance of the method proposed in Section 4 for selecting the number of important predictors. We considered $n = 100$ and 300, and $p = 5000$. The average numbers of selected predictors over 100 simulations are given in Table 1. The average numbers of selected predictors are much larger for Scenario 2 and censoring rate of 40%, which is expected because the signal is much weaker under Scenario 2 and the censoring rate is higher. In addition, with $n = 300$, the selected sets cover all the true signals in almost all the simulation runs for Scenario 1 under both models and censoring rates, while for Scenario 2, they cover almost half of the true signals on average. For smaller sample size of 100, the average numbers of selected true signals decrease, and the magnitude of decrease is relatively large for 40% censoring rate and Scenario 2.

Although the independent censoring condition was imposed for theoretical development, it can be relaxed in practice. We conducted simulations using a censoring distribution that depended on covariates. Specifically, the censoring times were generated from an exponential distribution with mean $c \exp(X_1 - X_8)$, with c chosen to achieve censoring rates of 15% and 40%. All other settings were unchanged from the previous simulations. Here, we only considered the proportional hazards model under Scenario 2 with $n = 100, 300$ and $p = 5000$. The simulation results are given in the Supplementary Material. Although the Kaplan–Meier estimator is not consistent for the survival distribution of the censoring time, the proposed method continued to perform competitively in this limited simulation study. In addition, we conducted simulations to examine the performance of the proposed and competing methods under a censoring rate of 70%. The simulation results are given in the Supplementary Material. In summary, the proposed method showed comparable performance under the heavy censoring case. The performance of the proposed method became slightly worse than partial likelihood ratio screening and the feature aberration at survival times screening. This behavior is expected because the proposed method uses the inverse probability-of-censoring weighted technique to address censoring, which might lose some efficiency when the censoring proportion is high.

7. Application to breast cancer data

We applied our proposed rank independence screening method to the analysis of survival from a breast cancer study (van Houwelingen et al., 2006), with 295 female patients with primary invasive breast carcinoma. For each patient, the expressions of 24885 genes were profiled on cDNA arrays from all tumors. A set of 4919 candidate genes were selected after initial screening using the Rosetta error model (van't Veer et al., 2002). The primary endpoint of interest was the overall survival time. Of the 295 patients, 216 had censored responses, giving a 73% censoring rate. A main goal of the study was to identify genes that are associated with the overall survival of breast cancer patients. As discussed in §1, gene expression profiles commonly contain outliers. For the breast cancer data, we identified potential outliers in the gene expressions. Specifically, for the jth gene of subject i, we calculated a modified z-statistic: $z_{ij} = 0.6745|X_{ji} - m_j|/v_j$, where m_j and v_j are the median and median absolute deviation of the jth gene expression profiles over the 295 subjects. If $z_{ij} > 3.5$, the data point was claimed to be an outlier. This criterion was suggested by Iglewicz & Hoaglin (1993) for outlier detection and has been widely used in the literature. Based on this

rule, of the 4919 genes, 3488 contained at least 1 outlier, 582 contained at least 10 outliers, and 58 contained at least 30 outliers.

To check the covariate-independent censoring assumption, we fit a marginal Cox proportional hazards model for each individual predictor. Out of 4919 genes, 368 genes have significant coefficients, with p-values less than 0.05. After Bonferroni correction, only one gene is significantly related to censoring times. For the 368 significant genes, we replaced the Kaplan–Meier estimator of the censoring survival function by the estimated conditional survival function from the fitted proportional hazards model and recalculated the corresponding screening statistics. We found that the rankings of the screening statistics are nearly unchanged. Therefore, in our data application, we used the Kaplan–Meier estimator for the censoring survival function.

We used the proposed method to analyze both the original data and the data with outliers removed. For comparison, we also include the selection results obtained using partial likelihood ratio screening and correlation screening. We used our method to estimate the number of important predictors based on the original data. The estimated size of the important set is 492. We report the symbols of the top 20 selected genes in Table 2. For the original data, none of the top 20 genes selected by correlation screening were selected by either partial likelihood ratio screening or the proposed method. However, several genes were selected by both partial likelihood ratio screening and the proposed method: 3 of the top 10 genes and 13 of the top 20 genes. For the data with outliers removed, only 1 of the top 20 genes selected by correlation screening was also selected by partial likelihood ratio screening. For each method, we compared the top 20 genes selected from the original data to the top 20 selected genes obtained from the data with outliers removed. The two sets of genes selected using partial likelihood ratio screening were completely different. However, the two sets selected by the proposed method had 18 genes in common, with a similar order. These results imply that partial likelihood ratio screening and the proposed method might give more reliable selection results than correlation screening; and the proposed method is robust to outliers in covariates, but partial likelihood ratio screening is not.

8. Discussion

In our method, censoring times are assumed to be independent of failure times and predictors, which may be too restrictive in some applications. This assumption can be relaxed to a certain extent. For example, as considered in He et al. (2013) for quantile-adaptive variable screening, we may assume that T_i and C_i are conditionally independent given a single predictor. Then, the survival function, $S(t)$, of censoring times can be replaced by the conditional survival function $S(t \mid X_{ij}) = \text{pr}(C_i \geq t \mid X_{ij})$, which can be consistently estimated by the local Kaplan–Meier estimator (Gonzalez-Manteiga & Cadarso-Suarez, 1994). Alternatively, we may build a semi-parametric survival model for censoring times, for example, a proportional hazards model with lasso selection of important predictors, and compute the model-based conditional survival function for censoring times. The sure screening property of the associated statistics needs to be further investigated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Appendix

Proof of the sure screening property

The following Lemmas are used to prove the sure screening property of $\hat{\tau}_k$.

Lemma 1. (Bitouze et al., 1999, *Theorem 1) Let $\{T_i\}_{i=1}^n$ and $\{C_i\}_{i=1}^n$ be independent sequences of independently identically distributed nonnegative random variables with distribution functions F and G, respectively. Let $\hat{F_n}$ be the Kaplan–Meier estimator of the distribution function F. There exists a positive constant, D, such that for any positive constant* λ,

$$\mathrm{pr}\{n^{1/2}\|(1-G)(\hat{F}_n - F)\|_\infty > \lambda\} \leq 2.5e^{-2\lambda^2 + D\lambda}.$$

Lemma 2. (Hoeffding, 1963) *Let $g = g(x_1, \ldots, x_m)$ be a symmetric kernel of the U-statistic, U, with $a \leq h(x_1, \ldots, x_m) \leq b$. For any $t > 0$ and $m \leq n$, we have*

$$\mathrm{pr}\{|U - E(U)| > t\} \leq 2\exp\left\{\frac{-2\lfloor n/m\rfloor t^2}{(b-a)^2}\right\}.$$

Lemma 3. *For any $c > 0$, when $n \geq D^2 c_1^{-1}$, where $c_1 = (1-\delta)^2 \{(2c+1)^{1/2}(c+1)^{-1/2} - 1\}^2$*

$c_1 = (1-\delta)^2\left\{(2c+1)^{1/2}(c+1)^{-1/2} - 1\right\}^2 \|S\|_\infty^2,$

$$\mathrm{pr}\left\{\max_i \left|\frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1\right| \geq c\right\} \leq 2.5ne^{-c_1 n}. \quad \text{(A1)}$$

Moreover, for any $0 < l < 1.12$, when $n > 25D^2 l^{-2}(1-\delta)^{-2}\|S\|_\infty^{-2}$,

$$\mathrm{pr}\left\{\max_i \left|\frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1\right| \geq l\right\} \leq 2.5ne^{-c_2 nl^2}, \quad \text{(A2)}$$

where $c_2 = 0.04(1-\delta)^2\|S\|_\infty^{-2}$.

Proof of Lemma 3. To show Lemma 3, we claim the following result, giving its proof at the end: for $t \in (0, 2^{1/2} - 1)$, $|\hat{S}^{-2}(V_i) - S^{-2}(V_i)| \geq cS^{-2}(V_i)$ implies $|\hat{S}(V_i) - S(V_i)| \geq tS(V_i)$, where

$c = \{(t + 1)^2 - 1\}/\{2 - (t + 1)^2\}$. This claim further implies that $\|\hat{S} - S\|_\infty \geq t\|S\|_\infty$. Therefore,

$$
\begin{aligned}
\mathrm{pr}\left\{\max_i \left|\frac{S^2(V_i)}{\hat{S}^2(V_i)}\right.\right.\\
\left.\left. - 1\right| \geq c\right\} \leq n\mathrm{pr}\left\{\left|\frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1\right| \geq c\right\} \leq n\mathrm{pr}\{|\hat{S}(V_i) \qquad \text{(A3)}\\
\left. - S(V_i)| \geq tS(V_i)\right\} \leq n\mathrm{pr}\left(\left\|\hat{S} - S\right\|_\infty \geq t\|S\|_\infty\right).
\end{aligned}
$$

Let $G(\cdot)$ denote the cumulative distribution function of T, i.e., $G(t) = \mathrm{pr}(T \leq t)$. By Condition A, $1 - \delta < |1 - G(V_i)| < 1$, for $i = 1, \ldots, n$. Therefore, by Lemma 1, we have

$$
\begin{aligned}
\mathrm{pr}\left(\|\hat{S} - S\|_\infty \geq t\|S\|_\infty\right) \leq \mathrm{pr}\left\{n^{1/2}\|(1 - G)(\hat{S} - S)\|_\infty > n^{1/2}(1 - \delta)\|S\|_\infty t\right\} \leq 2.5\exp\{\\
- 2n(1 - \delta)^2\|S\|_\infty^2 t^2 \qquad\qquad\qquad \text{(A4)}\\
+ Dn^{1/2}(1\\
- \delta)\|S\|_\infty t\}.
\end{aligned}
$$

When $n^{1/2} > Dt^{-1}(1 - \delta)^{-1}\|S\|_\infty^{-1}$, (A4) is further bounded by 2.5 $\exp\{-n(1 - \delta)^2\|S\|_\infty^2 t^2\}$.

For $0 < c < 1$, i.e., $0 < t < (3/2)^{1/2} - 1$, because

$$
c = \frac{1}{2 - (t+1)^2} - 1 < 5t,
$$

from the above calculations, for $n^{1/2} > Dt^{-1}(1 - \delta)^{-1}\|S\|_\infty^{-1}$,

$$
\mathrm{pr}\left\{\left|\frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1\right| \geq 5t\right\} \leq \mathrm{pr}\left\{\left|\frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1\right| \geq c\right\} \leq 2.5\exp\{-n(1 - \delta)^2\|S\|_\infty^2 t^2\}.
$$

The desired result (A2) now follows from the union bound of probability and by setting $l = 5t$.

We now show the result given at the beginning of the proof. Take $A = \hat{S}(V_i)$ and $B = S(V_i)$. Let $a = (t + 1)^2 - 1$. Therefore, $c = 1/(1 - a) - 1$. Because

$$
|A^{-2} - B^{-2}| \geq \{1/(1 - a) - 1\}B^{-2},
$$

we have $A^{-2} - B^{-2} \leq -\{1/(1 - a) - 1\}B^{-2}$, or $\geq \{1/(1 - a) - 1\}B^{-2}$. For $t \in (0, 2^{1/2} - 1)$, i.e., $a \in (0, 1)$, we have $1 - 1/(1 + a) < 1/(1 - a) - 1$. It follows that $A^{-2} - B^{-2} \leq -\{1 - 1/(1 + a)\}B^{-2}$, or $\geq \{1/(1 - a) - 1\}B^{-2}$, which is equivalent to $|A - B| \geq tB$.

Proof of Theorem 1. Rewrite $\hat{\tau_k} = 2\{n(n-1)\}^{-1} \sum_{i<j} g(W_i, W_j) - 1/4$, where

$$g(W_i, W_j) = \left\{ \frac{\Delta_j}{\hat{S}^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) + \frac{\Delta_j}{\hat{S}^2(V_i)} I(X_{kj} > X_{ki}, V_j > V_i) \right\} / 2$$

is the symmetric kernel of $\hat{\tau_k}$. Therefore, $\hat{\tau_k}$ is a U-statistic.

Let $U_n f = 2\{n(n-1)\}^{-1} \sum_{i<j} f(X_i, X_j)$ denote the empirical function for U-statistics. After some algebra, we have $\hat{\tau_k} - \tau_k = I_{k1} + I_{k2}$, where

$$I_{k1} = \binom{n}{2}^{-1} \sum_{i<j} \frac{\Delta_j}{\hat{S}^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j)$$
$$- \binom{n}{2}^{-1} \sum_{i<j} \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j)$$
$$= U_n \left[\left\{ \frac{1}{\hat{S}^2(V_j)} - \frac{1}{S^2(V_j)} \right\} \Delta_i I(X_{ki} > X_{kj}, V_i > V_j) \right],$$

and

$$I_{k2} = \binom{n}{2}^{-1} \sum_{i<j} \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) - \mathrm{pr}(X_{ki} > X_{kj}, T_i > T_j) = (U_n - E) \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j).$$

We bound I_{k1} and I_{k2} piece by piece. In particular, I_{k1} can be bounded from above as

$$I_{k1} \le \binom{n}{2}^{-1} \max_j \left| \frac{S^2(V_j)}{\hat{S}^2(V_j)} - 1 \right| \sum_{i<j} \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j). \quad (A5)$$

By the triangle inequality, (A5) can be further bounded above as

$$\max_j \left| \frac{S^2(V_j)}{\hat{S}^2(V_j)} \right.$$
$$\left. - 1 \right| \left| (U_n \right.$$
$$\left. - E) \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) \right|$$
$$+ \max_j \left| \frac{S^2(V_j)}{\hat{S}^2(V_j)} \right.$$
$$\left. - 1 \right| \left| E \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) \right| \equiv I_{k11} + I_{k12}.$$

To bound I_{k11}, recall that

$$g(W_i, W_j) = (1/2)\left\{ \frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) + \frac{\Delta_i}{S^2(V_i)} I(X_{ki} > X_{kj}, V_i > V_j) \right\} - 1/4.$$

By Condition A, we have

$$-\delta^{-2} - 1/4 \leq h(W_i, W_j) \leq \delta^{-2} - 1/4.$$

For any $c_3 > 0$, by Lemma 2, there exist c_3 and $c_4 = 4c_3^2$ such that

$$\mathrm{pr}\left\{ \left| (U_n - E)\frac{\Delta_j}{S^2(V_j)} I(X_{ki} > X_{kj}, V_i > V_j) \right| > c_3 n^{-\kappa} \right\} \leq 2\exp(-c_4 n^{1-2\kappa}). \quad \text{(A6)}$$

By Lemma 3, letting $c = 1$ in (A1), when $n \geq D^2(1-\delta)^{-2}\{(3/2)^{1/2} - 1\}^{-2}\|S\|_\infty^{-2}$,

$$\mathrm{pr}\left\{ \max_i \left| \frac{S^2(V_i)}{\hat{S}^2(V_i)} - 1 \right| \geq 1 \right\} \leq 2.5 n\exp(-c_1 n). \quad \text{(A7)}$$

It follows from (A6) and (A7) that for any $c_3 > 0$, when
$n \geq D^2(1-\delta)^{-2}\{(3/2)^{1/2} - 1\}^{-2}\|S\|_\infty^{-2}$, there exist c_1 and c_4 such that

$$\mathrm{pr}(|I_{k11}| \geq c_3 n^{-\kappa}) \leq 2.5 n\exp(-c_1 n) + 2\exp(-c_4 n^{1-2\kappa}). \quad \text{(A8)}$$

Because $|E\{\Delta_j S(V_j)^{-2} I(X_{ki} > X_{kj}, V_i > V_j)\}| \leq 1$, it follows from (A2) that for any $c_5 > 0$ and
$c_5 n^{-\kappa} < 1.12$, when $n > 25 D^2 c_5^2 n^{2\kappa}(1-\delta)^{-2}\|S\|_\infty^{-2}$, there exists $c_2 > 0$ such that

$$\mathrm{pr}(|I_{k12}| \geq c_5 n^{-\kappa}) \leq 2.5 n\exp(-c_2 n^{1-2\kappa}). \quad \text{(A9)}$$

By the triangle inequality, it now follows from (A6), (A8), and (A9) that for any $c_3, c_5 > 0$,
when

$$n > \max\left[D^2(1-\delta)^{-2}\{(3/2)^{1/2} - 1\}^{-2}\|S\|_\infty^{-2}, 25 D^2 c_5^2 n^{2\kappa}(1-\delta)^{-2}\|S\|_\infty^{-2}, (c_5/1.12)^{1/\kappa} \right],$$

there exist c_1, c_2, and c_4 such that

$$\mathrm{pr}\left\{ \max_{1 \leq k \leq p} |\hat{\tau}_k - \tau_k| > (2c_3 + c_5) n^{-\kappa} \right\} \leq p\left\{ 2.5 n\exp(-c_1 n) + 4\exp(-c_4 n^{1-2\kappa}) + 2.5 n\exp(-c_2 n^{1-2\kappa}) \right\}.$$

The first result follows by letting $c_6 = 2c_3 + c_5$.

For the second part, note that on the event

$$A_n \equiv \left\{ \max_{k \in \mathscr{M}_\star} |\hat{\tau}_k - \tau_k| \leq c_0 n^{-2\kappa}/2 \right\},$$

by Condition B, we have $|\hat{\tau_k}| \geq c_0 n^{-\kappa}/2$, for all $k \in \mathscr{M}_\star$. Therefore, by the choice of ν_n, we have $\mathscr{M}_\star \subset \hat{\mathscr{M}}_{\nu_n}$. The result now follows from a simple union bound:

$$\mathrm{pr}(A_n^c) \leq s \left\{ 2.5n\exp(-c_1 n) + 4\exp(-c_4 n^{1-2\kappa}) + 2.5n\exp(-c_2 n^{1-2\kappa}) \right\}.$$

This completes the proof.

Verification of Condition B for a general class of transformation models

Proof of Proposition 1. Recall that $\tau_k = \mathrm{pr}(X_{k1} > X_{k2}, T_1 > T_2) - 1/4$. Next we will show that $|\tau_k| \geq c_0 n^{-\kappa}$ for some $c_0 > 0$, if $k \in \mathscr{M}_\star$. For $k \in \mathscr{M}_\star$, we have

$$\begin{aligned}
\tau_k = E\left\{ I(X_{k1} > X_{k2}, T_1 > T_2) \right\} \\
- \frac{1}{4} = E\left(I(X_{k1} > X_{k2}) I\left[H(T_1) \right.\right. \\
- H(T_2) \\
- \{m_k(X_{k1}) - m_k(X_{k2})\} > m_k(X_{k2}) \\
- m_k(X_{k1}) - \frac{1}{4} = E\left(I(X_{k1} > X_{k2}) [1\right. \\
- F_{\Delta\varepsilon_k | \Delta m_k} \{m_k(X_{k2}) \\
\left.\left.\left. - m_k(X_{k1})\} \right] \right) - \frac{1}{4},
\end{aligned}$$

where $F_{\Delta\varepsilon_k | \Delta m_k}(\cdot)$ is the conditional cumulative distribution function of $\Delta\varepsilon_k = H(T_1) - H(T_2) - \{m_k(X_{k1}) - m_k(X_{k2})\}$ given $\Delta m_k = m_k(X_{k1}) - m_k(X_{k2})$. Because $m_k(\cdot)$ is a monotone function, $m_k(X_{k2}) - m_k(X_{k1})$ is either greater than or less than zero for all $X_{k1} > X_{k2}$. This implies that $1 - F_{\Delta\varepsilon_k | \Delta m_k} \{m_k(X_{k2}) - m_k(X_{k1})\}$ is either greater or less than $1/2$ due to Condition C1. Therefore, τ_k is either greater or less than zero for $k \in \mathscr{M}_\star$. In the following, we further establish the lower bound of $|\tau_k|$.

Without loss of generality, assume $m_k(\cdot)$ is monotone increasing. Note that τ_k can be equivalently written as

$$\tau_k = E\left(I(X_{k1} > X_{k2}) \left[F_{\Delta\varepsilon_k | \Delta m_k} \{m_k(X_{k1}) - m_k(X_{k2})\} - F_{\Delta\varepsilon_k | \Delta m_k}(0) \right] \right) = E\left\{ I(X_{k1} > X_{k2}) \int_0^{m_k(X_{k1}) - m_k(X_{k2})} f_{\Delta\varepsilon_k | \Delta m_k}(t)dt \right\}.$$

According to Corollary 3 in Sellke & Sellke (1997), for a random variable X with mean zero, variance σ^2, and unimodal symmetric distribution, $\mathrm{pr}(|X| \geq t) \leq 3^{1/2}\sigma/(t + 3^{1/2}\sigma)$. By Condition C1, we have

$$\int_0^{m_k(X_{k1})-m_k(X_{k2})} f_{\Delta\varepsilon_k|\Delta m_k}(t)dt \geq \frac{1}{2}\left\{1 - \frac{3^{1/2}\sigma_2}{m_k(X_{k1})-m_k(X_{k2})+3^{1/2}\sigma_2}\right\} = \frac{m_k(X_{k1})-m_k(X_{k2})}{2\{m_k(X_{k1})-m_k(X_{k2})+3^{1/2}\sigma_2\}}.$$

This expression leads to

$$\tau_k \geq E\left[I\{m_k(X_{k1})>m_k(X_{k2})\}\frac{m_k(X_{k1})-m_k(X_{k2})}{2\{m_k(X_{k1})-m_k(X_{k2})+3^{1/2}\sigma_2\}}\right] \geq \frac{1}{2M+2\times 3^{1/2}\sigma_2}E\left[\{m_k(X_{k1})-m_k(X_{k2})\}I\{0<m_k(\cdot\right.$$

$$=\frac{1}{2M+2\times 3^{1/2}\sigma_2}E\left[\{m_k(X_{k1})-m_k(X_{k2})\}I\{m_k(X_{k1})\right.$$

$$-m_k(X_{k2})>0\}$$

$$-\{m_k(X_{k1})$$

$$-m_k(X_{k2})\}1\{m_k(X_{k1})$$

$$\left.-m_k(X_{k2})>M\}\right],$$

for $M > 0$. By Chebyshev's inequality and C2, the second term can be further bounded below as

$$\frac{1}{2M+2\times 3^{1/2}\sigma_2}\left(\frac{1}{2}E|m_k(X_{k1})\right.$$

$$\left.-m_k(X_{k2})| - [2\sigma_1^2\mathrm{pr}\{m_k(X_{k1})-m_k(X_{k2})>M\}]^{1/2}\right) \geq \frac{1}{2M+2\times 3^{1/2}\sigma_2}\left\{\frac{1}{2}E|m_k(X_{k1})-Em_k(X_{k1})|-\frac{2\sigma_1^2}{M}\right\}.$$

Let $M=8\sigma_1^2 d_0^{-1}n^{\kappa/2}$. Then, $M \geq 3^{1/2}\sigma_2$ when n is large. We have $\tau_k \geq c_0 n^{-\kappa}$, where $c_0=d_0^2/(128\sigma_1^2)$. Similarly, if $m_k(\cdot)$ is monotone decreasing, we can show that $\tau_k \leq -c_0 n^{-\kappa}$. Therefore, $|\tau_k| \geq c_0 n^{-\kappa}$ for any $k \in \mathcal{M}_\star$. Condition B is then proved.

Proof of Theorem 2. To show Theorem 2, we note that for any $c_8 > 0$, if $|EX_{(k)}Y| > c_8 n^{-\kappa}$, then $|\tau_k| > c_0 n^{-\kappa}$ for some $c_0 > 0$. The proof of this statement is similar to that in Proposition 1, hence we omit the details.

Because $\mathrm{var}\{H(T)\} = O(1)$, we have that

$$\sum_{k=1}^p \{EX_{(k)}H(T)\}^2 = \|EX'H(T)\|^2 = \|EX'(X'\beta+\varepsilon)\|^2 = O(\|\Sigma\beta\|^2) = O\{\lambda_{\max}(\Sigma)\}.$$

Therefore, the number of $\{k : \tau_k > c_0 n^{-\kappa}\} = \{k : |EX_{(k)}H(T)| > c_8 n^{-\kappa}\} = O\{n^{2\kappa}\lambda_{\max}(\Sigma)\}$. Because the number of $\{k : |\hat{\tau}_k| > 2c_0 n^{-\kappa}\}$ is no bigger than the number of $\{k : |\tau_k| > c_0 n^{-\kappa}\}$ on the set, $\{\max_{1 \leq k \leq p}|\hat{\tau}_k - \tau_k| \leq c_0 n^{-\kappa}\}$. By taking $c_8 = c_7/2$,

$$\mathrm{pr}[|\hat{\mathcal{M}}_{\nu_n}| \leq O\{n^{2\kappa}\lambda_{\max}(\Sigma)\}] \geq \mathrm{pr}(\max_{1\leq k\leq p}|\hat{\tau}_k - \tau_k| \leq c_0 n^{-\kappa}).$$

The conclusion follows from the tail probability in Theorem 1. This completes the proof.

References

Bickel P, Ritov Y, Tsybakov A. Simultaneous analysis of lasso and Dantzig selector. The Annals of Statistics. 2009; 37:1705–1732.

Bitouze, Laurent; Massart. A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator. Annales de l'Institut Henri Poincare (B) Probability and Statistics. 1999; 35:735–763.

Candes E, Tao T. The Dantzig selector: statistical estimation when p is much larger than n (with discussion). The Annals of Statistics. 2007; 35:2313–2404.

Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). Journal of the Royal Statistical Society, Series A. 1985; 148:82–117.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96:1348–1360.

Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. The Annals of Statistics. 2002; 30:74–99.

Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space (with discussion). Journal of the Royal Statistical Society, Series B. 2008; 70:849–911.

Fan J, Song R, et al. Sure independence screening in generalized linear models with NP– dimensionality. The Annals of Statistics. 2010; 38:3567–3604.

Gonzalez-Manteiga W, Cadarso-Suarez C. Asymptotic properties of a generalized Kaplan– Meier estimator with some applications. Journal of Nonparametric Statisics. 1994; 4:65–78.

Gorst-Rasmussen A, Scheike T. Independent screening for single-index hazard rate models with ultrahigh dimensional features. Journal of the Royal Statistical Society, Series B. 2013; 75:217– 245.

Hall P, Titterington DM, Xue J. Tilting methods for assessing the influence of components in a classifier. Journal of the Royal Statistical Society, Series B. 2009; 71:783–803.

He X, Wang L, Hong HG. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. The Annals of Statistics. 2013; 41:342–369.

Hoeffding W. Probability inequality for sums of bounded random variables. Journal of the American Statistical Association. 1963; 58:13–30.

Iglewicz, B.; Hoaglin, DC. How to Detect and Handle Outliers. Milwaukee, WI: American Society for Quality Control; 1993.

Johnson BA. Variable selection in semiparametric linear regression with censored data. Journal of the Royal Statistical Society, Series B. 2008; 70:351–370.

Johnson BA, Lin DY, Zeng D. Penalized estimating functions and variable selection in semi- parametric regression models. Journal of the American Statistical Association. 2008; 103:672– 680. [PubMed: 20376193]

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2nd ed.. New York: Wiley; 2011.

Kendall, MG. Rank Correlation Methods. 3rd ed.. London: Griffin & Co; 1962.

Kosorok MR, Lee BL, Fine JP. Robust inference for univariate proportional hazards frailty regression models. The Annals of Statistics. 2004; 32:1448–1491.

Kuo C-L, Zaykin DV. Novel rank-based approaches for discovery and replication in genome-wide association studies. Genetics. 2011; 189:329–340. [PubMed: 21705758]

Meinshausen N, Rice J. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. The Annals of Statistics. 2006; 34:373–393.

Meinshausen N, Yu B. Lasso-type recovery of sparse representations for high-dimensional data. The Annals of Statistics. 2009; 37:246–270.

Peng L, Fine J. Competing risks quantile regression. Journal of the American Statistical Association. 2009; 104:1440–1453.

Scharfstein DO, Tsiatis AA, Gilbert PB. Semiparametric efficient estimation in the generalized odds- rate class of regression models for right-censored time-to-event data. Lifetime Data Analysis. 1998; 4:355–391. [PubMed: 9880995]

Sellke TM, Sellke SH. Chebyshev inequalities for unimodal distributions. The American Statistician. 1997; 51:34–39.

Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association. 1968; 63:1379–1389.

Skol D, Scott L, Abecasis G, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature Genetics. 2006; 38:209–213. [PubMed: 16415888]

Tibshirani RJ. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B. 1996; 58:267–288.

Tibshirani RJ. The lasso method for variable selection in the Cox model. Statistics in Medicine. 1997; 16:385–395. [PubMed: 9044528]

Tibshirani RJ. Univariate shrinkage in the cox model for high dimensional data. Statistical Applications in Genetics and Molecular Biology. 2009; 8:1–18.

van de Geer S. High-dimensional generalized linear models and the lasso. The Annals of Statistics. 2008; 36:614–645.

van Houwelingen HC, Bruinsma T, Hart AAM, Van 't Veer LJ, Wessels LFA. Cross-validated cox regression on microarray gene expression data. Statistics in Medicine. 2006; 25:3201–3216. [PubMed: 16143967]

van 't Veer L, Dai H, van de Vijver MJ, He Y, Hart A, Mao M, Peterse H, van der Kooy K, Marton MJ, Witteveen AT, Schreiber G, Kerkhoven R, Roberts C, Linsley P, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002; 415:530–536. [PubMed: 11823860]

Zhang HH, Lu W. Adaptive lasso for Cox's proportional hazards model. Biometrika. 2007; 94:691–703.

Zhao DS, Li Y. Principled sure independence screening for Cox models with ultra-high-dimensional covariates. Journal of Multivariae Analysis. 2012; 105:397–411.
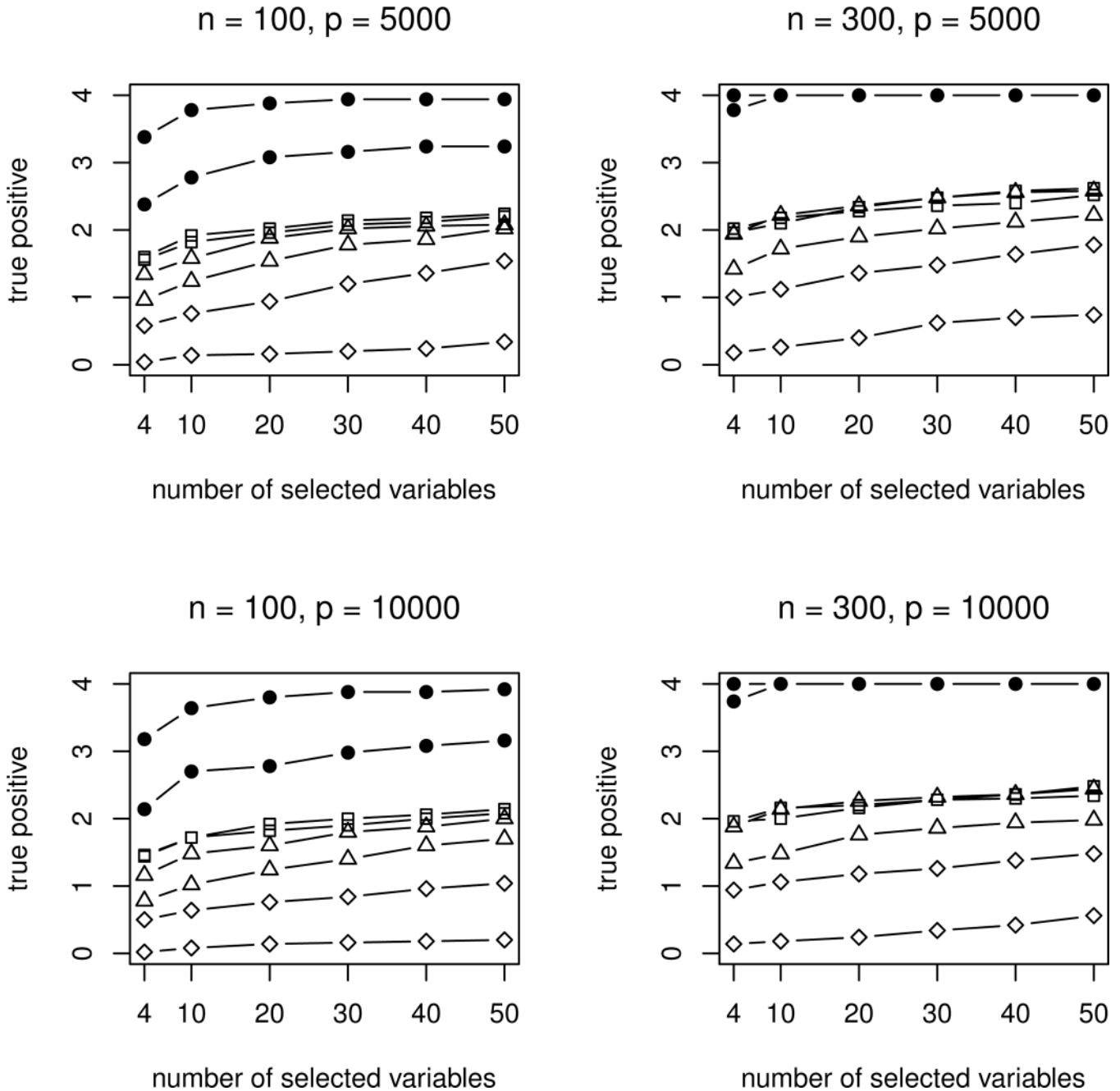
Fig. 1.
Average numbers of active variables contained in the top 4, 10, 20, 30, 40, and 50 selected variables for the proportional hazards model under Scenario 1. Solid circles denote the results obtained using the proposed method; triangles denote feature aberration at survival times screening, squares denote partial likelihood ratio screening; and diamonds denote correlation screening. For each method, the upper line is for case of 15% censoring, and the lower line is for case of 40% censoring.
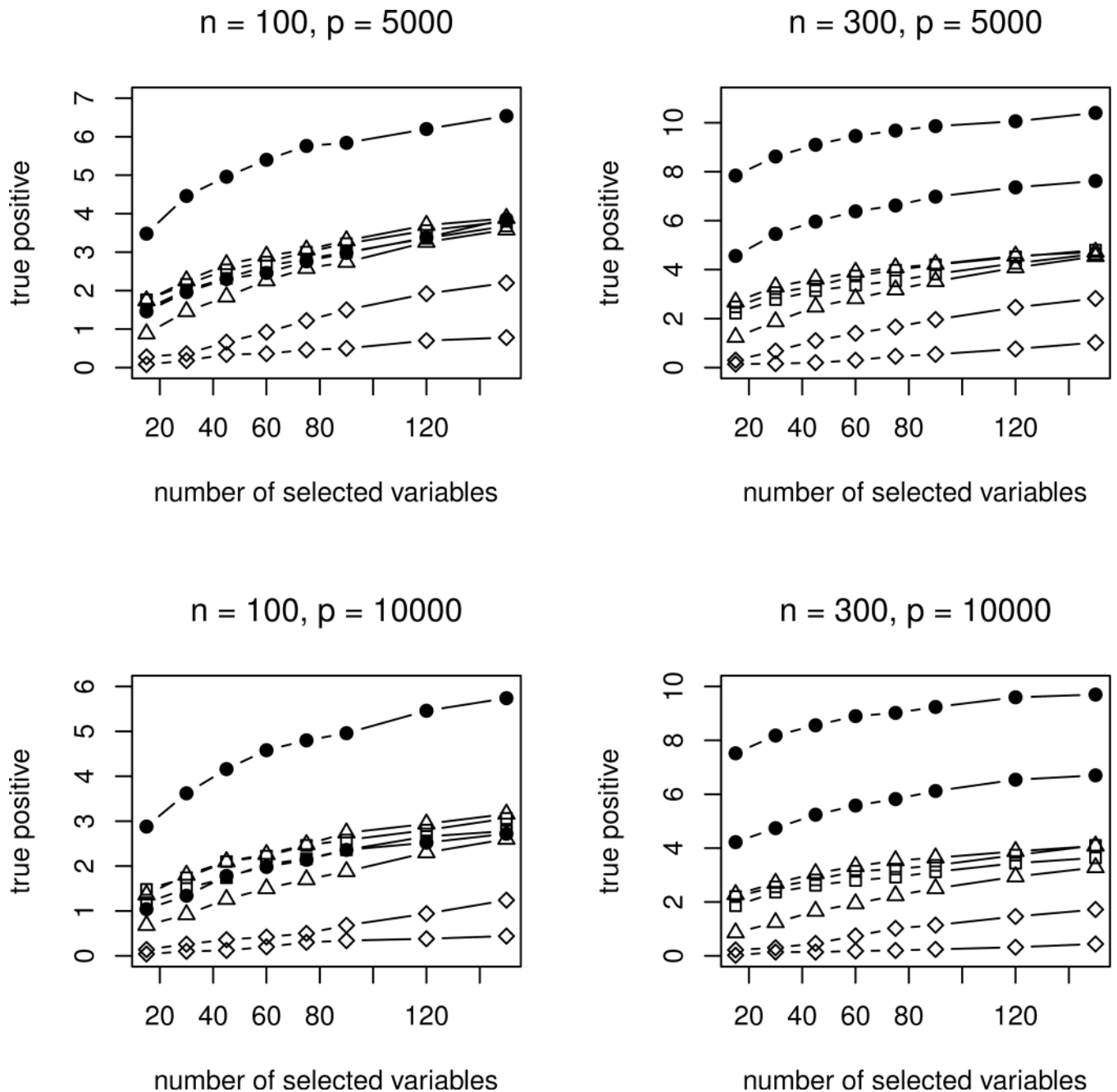
Fig. 2.
The average numbers of active variables contained in the top 15, 30, 45, 60, 75, 90, 120, and 150 selected variables for the proportional hazards model under Scenario 2. Solid circles denote the results obtained using the proposed method; triangles denote feature aberration at survival times screening, squares denote partial likelihood ratio screening; and diamonds denote correlation screening. For each method, the upper line is for case of 15% censoring, and the lower line is for case of 40% censoring.

Table 1

Average numbers of important selected predictors.

Model	Scenario	n = 100				n = 300			
		N.sel (SD)	N.sel (SD)	N.sel (SD)	N.sel (SD)	N.sel (SD)	N.sel (SD)	N.sel (SD)	N.sel (SD)
		CP = 15%	CP = 40%			CP = 15%	CP = 40%		
PH	1	13.8 (4.8)	28.1 (8.9)			7.4 (3.0)	28.7 (10.2)		
	2	31.3 (16.4)	116.7 (80.5)			9.6 (6.5)	101.6 (79.2)		
PO	1	11.6 (4.7)	26.1 (7.8)			7.5 (3.3)	23.1 (10.6)		
	2	45.0 (15.6)	125.4 (87.3)			25.0 (8.0)	129.8 (100.4)		

[†] CP denotes censoring proportion; N.sel denotes the average number of selected important predictors over 100 simulations with the number in parenthesis being the standard deviation over 100 simulations; PH denotes the proportional hazards model; PO denotes the proportional odds model.

Table 2

Top 20 selected genes for breast cancer data.

order	CS		PLRS		CRIS	
	W	WO	W	WO	W	WO
1	NM_007117	NM_005310	NM_018410	NM_001942	Contig38288_RC	Contig38288_RC
2	AL137517	NM_001871	NM_006607	NM_000509	Contig57584_RC	Contig57584_RC
3	U22029	AB037821	D38553	NM_004496	NM_001168	NM_001168
4	NM_000764	NM_007117	D14678	NM_016267	Contig48270_RC	NM_003258
5	NM_012067	NM_004291	NM_004701	Contig45588_RC	NM_003258	Contig48270_RC
6	NM_001819	NM_000764	Contig57584_RC	NM_006121	NM_006623	NM_006623
7	M33318	NM_001819	NM_003981	Contig44191_RC	NM_001333	NM_006607
8	NM_001871	U22029	D43950	NM_000423	NM_006607	NM_006739
9	NM_004291	M33318	NM_016359	Contig51687_RC	NM_006096	Contig31288_RC
10	NM_005310	NM_012242	NM_001168	Contig27882_RC	Contig31288_RC	NM_006096
11	NM_001741	NM_001327	U96131	NM_002590	NM_004701	NM_004701
12	NM_005982	Contig50979_RC	NM_001333	NM_001844	M96577	M96577
13	NM_016569	Contig21930_RC	Contig38288_RC	NM_000427	NM_004217	NM_001333
14	NM_000439	NM_002809	Contig31288_RC	NM_005310	NM_016359	NM_004217
15	Contig50979_RC	NM_000735	NM_001809	NM_003360	U96131	NM_016359
16	AF070536	NM_001741	NM_004217	NM_014211	NM_004456	NM_007019
17	AF131851	NM_020411	M96577	NM_006536	U74612	U96131
18	NM_000363	AJ275978	NM_001605	Contig53968_RC	NM_007019	NM_003981
19	Contig49589_RC	NM_004988	NM_005733	Contig43195_RC	NM_003981	NM_006579
20	Contig32563_RC	Contig42751_RC	NM_007019	NM_001453	Contig41413_RC	Contig41413_RC

†CS denotes correlation screening; PLRS denotes partial likelihood ratio screening; CRIS denotes proposed censored rank independence screening; W denotes the original data; and WO denotes the data with outliers removed.