

REGRESSION SELECTION VIA THE ADAPTIVE LASSO FOR CURRENT STATUS DATA UNDER THE ADDITIVE HAZARDS MODEL

ZHANG Yi-jin¹, WANG Cheng-yong²

(1. School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China)

(2. School of Mathematics and Statistics, Hubei University of Arts and Science,
Xiangyang 441053, China)

Abstract: Variable selection is commonly employed when the true underlying model has a sparse representation. Identifying significant predictors will enhance the prediction performance of the fitted model. To solve this problem, among others, Zhang and Lu [1] developed a variable selection method under the framework of the proportional hazards model when one observes right-censored data. In this paper, We consider the variable selection problem for the additive hazards model when one faces current status data. Motivated by Zhang and Lu [1], we develop an adaptive Lasso method for this problem. Some theoretical properties, including consistency and oracle properties are established under some weak regularity conditions. An extensive simulation is performed to show that the method performs competitively. This method is also applied to a real data set from a tumorigenicity study.

Keywords: Additive hazards model; current status data; adaptive Lasso; ADMM algorithm

2010 MR Subject Classification: 62N02; 62F10; 62F12

Document code: A **Article ID:** 0255-7797(2021)03-0189-16

1 Introduction

Current status data or Case I interval-censored failure time data occur frequently in survival analysis when an exact event time of interest is not available, and only whether or not the event has occurred up to a certain random monitoring time. That event's current status' is known. This kind of data are often encountered in epidemiological studies, carcinogenicity experiments, econometrics and reliability studies among others. Regression analysis of failure time data is one of the main objectives in survival analysis. In regression analysis, an important and challenging task is to identify the risk factors and their risk contributions. Often, not all the collected covariates may contribute to the predication of outcomes and we need to remove some unimportant covariates.

There are many variable selection techniques in linear regression models. Some of them have been extended to the survival analysis, for example, Bayesian variable selection methods

* **Received date:** 2020-10-12

Accepted date: 2020-12-11

Foundation item: Supported by National Natural Science Foundation of China(71371066).

Biography: Zhang Yijin (1996-), male, born at Xiaogan, Hubei, postgraduate, major in statistics.

Corresponding author: Wang Chengyong.

for censored survival data were proposed by Faraggi and Simon [2]. However, the sampling properties of this selection methods are largely unknown(see Fan and Li [3]). The least absolute shrinkage and selection operator (Lasso), proposed by Tibshirani [4], is a member of variable selection family based on a penalized likelihood approach with the L_1 -penalty. It can delete insignificant variables by estimating their coefficients as 0. Tibshirani [5] proposed using the Lasso for estimation and variable selection under the Cox model. However, the Lasso estimator does not possess the oracle properties(see [3]). Many other variable selection methods have been developed following Tibshirani [4]. For example, the smoothly clipped absolute derivation (SCAD) by Fan and Li [6] and the adaptive Lasso (aLasso) by Zou [7]. Both of them have nice properties.

So far many literatures have developed variable selection methods for right-censored data (see for example, [3], [5], [8]). In particular, some penalized methods have been established under the Cox's proportional hazards model. For example, Tibshirani [5] proposed using the Lasso for the variable selection under the Cox model and right-censored data. Fan and Li [3] generalized the SCAD to the Cox model with right-censored data. The aLasso method also has been extended to the context of proportional hazards model when one observes right-censored data by Zhang and Lu [1]. Huang et al. [9] studied the Lasso estimator in sparse, high-dimensional Cox model. Zhao et al [10] studied the simultaneous estimation and variable selection for interval-censored data under the Cox model.

The additive hazards model as an alternative model, which describes a different aspect of the association between the failure time and covariates than the proportional hazards model, is another commonly used regression model in survival analysis. A lot of theoretical results of the estimated regression parameters under additive hazards model have been well established (see for example, [11–13]). It is well-known that many efforts have been focused on the methods of variable selection for Cox model with right-censored observation data. However, as mentioned by Zhao et al [10], there exists little literature on variable selection for interval-censored data. There are relatively less studies developed for the additive hazards model with interval-censored data. This paper considers the variable selection method for case I interval-censored data under the additive hazards model.

The remainder of the paper is organized as follows. In Section 2, we will introduce some notations and assumptions that will be used in this paper. In Section 3, we develop an adaptive lasso method, and give its statistical properties. Section 4 gives some details about the ADMM algorithm that will be applied to solve the adaptive lasso. Section 5 provides some numerical results from an extensive simulation study to assess the performance of the proposed method, and Section 6 applies the proposed method to a real data set from a tumorigenicity study.

2 Notations and Models

Consider a random sample of n independent subjects. For $i = 1, \dots, n$, let T_i and C_i denote the failure time of interest and censoring time of the i -th subject, and $Z_i(t) =$

$(Z_{i1}(t), \dots, Z_{ip}(t))'$ be the vector of possibly time-dependent covariates. Furthermore, since only current status data are available for failure time T_i 's, the observed data are given by $\{C_i, \delta_i = I(T_i \geq C_i), Z_i(t), i = 1, \dots, n\}$. In the next section, we present methods for the cases in which the monitoring time C is independent or dependent of T and Z .

2.1 Independent Censoring

In this subsection, we suppose that C is independent of T and Z . To model the covariate effect, we assume that the hazard function of T at time t , given the history of a p -dimensional covariate process $Z(\cdot)$ up to t , has the form

$$\lambda(t|Z) = \lambda_0(t) + \beta_0'Z(t), \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and β_0 is a p -vector of unknown regression parameters.

For $i = 1, \dots, n$, define $N_i(t) = \delta_i I(C_i \leq t)$, and $Y_i(t) = I(C_i \geq t)$. It can be shown that the counting process $N_i(t)$ has the Cox type intensity process as follows

$$dH_i(t) = e^{-\beta_0'Z_i^*(t)}dH_0(t), \tag{2.2}$$

where $dH_0(t) = e^{-\Lambda_0(t)}d\Lambda_c(t)$, $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, $\Lambda_c(t) = \int_0^t \lambda_c(s)ds$, $Z_i^*(t) = \int_0^t Z_i(s)ds$, and $\lambda_c(t)$ is the hazard function of C . Therefore,

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)e^{-\beta_0'Z_i^*(s)}dH_0(s), \quad i = 1, \dots, n$$

are martingales with respect to the σ -filtration $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s) : s \leq t, i = 1, \dots, n\}$. Thus, we can make inferences about β_0 by applying the partial likelihood principle to model (2.2). For this, we first define the partial likelihood function as follows

$$L_1(\beta) = \prod_{i=1}^n \left(\frac{e^{-\beta'Z_i^*(C_i)}}{\sum_{j=1}^n Y_j(C_i)e^{-\beta'Z_j^*(C_i)}} \right)^{\delta_i}.$$

Taking logarithm of it yields that

$$l_1(\beta) = \log L_1(\beta) = \sum_{i=1}^n \int_0^\tau -\beta'Z_i^*(t)dN_i(t) - \int_0^\tau \log \left(\sum_{j=1}^n Y_j(t)e^{-\beta'Z_j^*(t)} \right) d\bar{N}(t),$$

where $\bar{N}(t) = \sum_{i=1}^n N_i(t)$, and τ is the longest follow-up time. For $k = 0, 1, 2$, we also define $S^{(k)}(t, \beta) = \sum_{j=1}^n (Z_j^*(t))^{\otimes k} Y_j(t) e^{-\beta'Z_j^*(t)}$, where $Z^{\otimes 0} = 1, Z^{\otimes 1} = Z, Z^{\otimes 2} = ZZ'$. By differentiation and rearrangement of terms, the gradient of $l_n(\beta)$ is

$$U_1(\beta) = \frac{\partial l_1(\beta)}{\partial \beta} = \sum_{i=1}^n \int_0^\tau \left(-Z_i^*(t) + \frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)} \right) dN_i(t),$$

and the Hessian matrix is

$$\mathcal{H}_1(\beta) = \frac{\partial^2 l_1(\beta)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n \int_0^\tau \left(Z_i^*(t) - \frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)} \right)^{\otimes 2} Y_i(t) e^{-\beta' Z_i^*(t)} \frac{d\bar{N}(t)}{S^{(0)}(t, \beta)}.$$

It can be seen that the Hessian matrix of $l_1(\beta)$ is negative definite, so $l_1(\beta)$ is concave in β , that is, $l_1(\beta)$ has a unique maximizer $\tilde{\beta}$. The estimate $\tilde{\beta}$ of β_0 can be obtained by maximizing the function $l_1(\beta)$, or solving the equation $U_1(\beta) = 0$.

2.2 Dependent Censoring

When the censoring time C is not independent of the covariate vector Z , we describe the relationship between C and Z by the following hazards model,

$$d\Lambda_c(t|Z) = e^{\gamma_0' Z(t)} d\Lambda_{c0}(t) := \lambda_{c0} e^{\gamma_0' Z(t)} dt, \quad (2.3)$$

where $\Lambda_{c0}(t)$ is an unknown cumulative baseline hazard function, and γ_0 is a p -vector of unknown regression parameters. We assumed that C is conditionally independent of T given the covariate vector Z .

By the arguments leading to (2.2), it can be shown that, under model (2.1) and (2.3), the compensated counting processes

$$\tilde{M}_i(t) = N_i(t) - \int_0^t Y_i(s) e^{-\beta_0' Z_i^*(s) + \gamma_0' Z_i(s)} dH_0(s), \quad i = 1, \dots, n \quad (2.4)$$

are martingales with respect to the σ -filtration \mathcal{F}_t . The notations $N_i(t)$ and $H_0(t)$ are the same as those defined in subsection 2.1. We can also apply the partial likelihood principle to model (2.4) to make inferences for the unknown parameters β_0 and γ_0 . That is, we can consider the following partial likelihood function

$$L_2(\beta, \gamma) = \prod_{i=1}^n \left(\frac{e^{-\beta' Z_i^*(C_i) + \gamma' Z_i(C_i)}}{\sum_{j=1}^n Y_j(C_i) e^{-\beta' Z_j^*(C_i) + \gamma' Z_j(C_i)}} \right)^{\delta_i}.$$

However, the function $L_2(\beta, \gamma)$ above utilizes only the information of C_i 's with non zero δ_i 's, and we mainly focus on β , it would be more efficient to estimate γ_0 by applying the partial likelihood theory directly to the model (2.3). Hence, for the estimate of γ_0 , we first consider the following partial likelihood function

$$L_3(\gamma) = \prod_{i=1}^n \left(\frac{e^{\gamma' Z_i(C_i)}}{\sum_{j=1}^n Y_j(C_i) e^{\gamma' Z_j(C_i)}} \right).$$

The maximum partial likelihood estimator $\hat{\gamma}$ of γ_0 can be obtained by maximizing the function $L_3(\gamma)$. Of course, $\hat{\gamma}$ can also be obtained by solving the score equation $U_\gamma(\gamma) = 0$, where

$$U_\gamma(\gamma) = \sum_{i=1}^n \left(Z_i - \frac{\sum_{j=1}^n Y_j(C_i) e^{\gamma' Z_j(C_i)} Z_j(C_i)}{\sum_{j=1}^n Y_j(C_i) e^{\gamma' Z_j(C_i)}} \right). \quad (2.5)$$

Given $\hat{\gamma}$, we estimate β_0 by the following function

$$L_2(\beta, \hat{\gamma}) = \prod_{i=1}^n \left(\frac{\exp(-\beta' Z_i^*(C_i) + \hat{\gamma}' Z_i(C_i))}{\sum_{j=i}^n Y_j(C_i) \exp(-\beta' Z_j^*(C_i) + \hat{\gamma}' Z_j(C_i))} \right)^{\delta_i}.$$

The estimate $\hat{\beta}$ of β_0 can be obtained by maximizing the function $L_2(\beta, \hat{\gamma})$ or $l_2(\beta)$, where $l_2(\beta)$ is defined as

$$l_2(\beta) = \log L_2(\beta, \hat{\gamma}) = \sum_{i=1}^n \int_0^\tau (-\beta' Z_i^*(t) + \hat{\gamma}' Z_i(t)) dN_i(t) - \int_0^\tau \log \left(\sum_{j=1}^n Y_j(t) e^{-\beta' Z_j^*(t) + \hat{\gamma}' Z_j(t)} \right) d\bar{N}(t).$$

For $k = 0, 1, 2$, define $\tilde{S}^{(k)}(t, \beta, \gamma) = \sum_{j=1}^n (Z_j^*(t))^{\otimes k} Y_j(t) e^{-\beta' Z_j^*(t) + \gamma' Z_j(t)}$. Similar to the process above, we can get the following score function

$$U_2(\beta) = \frac{\partial l_2(\beta)}{\partial \beta} = \sum_{i=1}^n \int_0^\tau \left(-Z_i^*(t) + \frac{\tilde{S}^{(1)}(t, \beta, \hat{\gamma})}{\tilde{S}^{(0)}(t, \beta, \hat{\gamma})} \right) dN_i(t).$$

The estimate $\hat{\beta}$ also can be obtained by solving the equation $U_2(\beta) = 0$.

In the following, we will discuss the development of a penalized or regularized procedure for covariate selection based on the functions $l_1(\beta)$ and $l_2(\beta)$.

3 Adaptive Lasso Estimation Procedure

We assume that one observes right-censored data, to select and estimate important variables under the proportional hazards model, Zhang and Lu [1] proposed to minimize the penalized log partial likelihood function,

$$-\frac{1}{n} l_n^*(\beta) + \lambda \sum_{j=1}^p |\beta_j| / |\tilde{\beta}_j|$$

where $l_n^*(\beta)$ denotes the log partial likelihood based on the right-censored data and the proportional hazards model, the positive weights $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)'$ is the maximizer of the log partial likelihood, λ is a nonnegative penalization tuning parameter.

Consider the current status data under model (2.1), note that the intensity process of the counting process $N_i(t)$ also satisfies Cox type. This suggests that we can select variables by employing a similar method of Zhang and Lu [1]. We propose the adaptive Lasso estimator $\hat{\beta}_n$ as follows,

$$\hat{\beta}_n = \arg \min_{\beta} \left\{ -\frac{1}{n} l_1(\beta) + \lambda_n \sum_{j=1}^p |\beta_j| \omega_j \right\}, \tag{3.1}$$

or

$$\hat{\beta}_n = \arg \min_{\beta} \left\{ -\frac{1}{n} l_2(\beta) + \lambda_n \sum_{j=1}^p |\beta_j| \omega_j \right\}. \tag{3.2}$$

The values of ω_j 's can be chosen by different ways. In this paper, we specify $\omega_j = 1/|\tilde{\beta}_j|$, where $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)'$ is the maximizer of the log partial likelihood $l_i(\beta)$, $i = 1, 2$.

To study the oracle properties of the estimators, we first consider the penalized log partial likelihood function

$$Q_i(\beta) = l_i(\beta) - n\lambda_n \sum_{j=1}^p |\beta_j|/|\tilde{\beta}_j|, \quad i = 1, 2. \tag{3.3}$$

Let $\beta_0 = (\beta'_{10}, \beta'_{20})'$ denote the true parameter vector, where β_{10} consists of all q nonzero components and β_{20} consists of the remaining zero components. Similarly, we use $\hat{\beta}_n = (\hat{\beta}'_{1n}, \hat{\beta}'_{2n})'$ to denote the maximizer of (3.1) or (3.2). In the case of independent censoring, we can get the Fisher information matrix $\Omega(\beta_0)$, which is the limit of $n^{-1}(-\mathcal{H}_1(\beta))$. As usual, we assume that $\Omega(\beta_0)$ is nonsingular. In the case of dependent censoring, let

$$\hat{\Omega}_\beta(\beta; \gamma) = -n^{-1} \frac{\partial U_2(\beta; \gamma)}{\partial \beta'}, \quad \hat{\Omega}_{\beta\gamma}(\beta; \gamma) = n^{-1} \frac{\partial U_2(\beta; \gamma)}{\partial \gamma'}, \quad \hat{D}_\gamma(\gamma) = -n^{-1} \frac{\partial U_\gamma(\gamma)}{\partial \gamma'}$$

and let Ω_β , $\Omega_{\beta\gamma}$ and D_γ denote their limits at $\beta = \beta_0$ and $\gamma = \gamma_0$.

Using some similar arguments as those of Lin et al [11], we can prove that the random vectors $n^{-\frac{1}{2}}U_2(\beta_0; \hat{\gamma})$ and $n^{\frac{1}{2}}(\tilde{\beta} - \beta_0)$ converge in distribution to zero-mean normal random vectors with covariance matrices $M(\beta_0) = \Omega_\beta - \Omega_{\beta\gamma}D_\gamma^{-1}\Omega'_{\beta\gamma}$ and $V(\beta_0) = \Omega_\beta^{-1} - \Omega_\beta^{-1}\Omega_{\beta\gamma}D_\gamma^{-1}\Omega'_{\beta\gamma}\Omega_\beta^{-1}$, respectively.

Let $\Omega_1(\beta_{10}) = \Omega_{11}(\beta_{10}, 0)$, where $\Omega_{11}(\beta_{10}, 0)$ is the leading $q \times q$ submatrix of $\Omega(\beta_0)$ with $\beta_{20} = 0$ and $V_1(\beta_{10}) = V_{11}(\beta_{10}, 0)$, where $V_{11}(\beta_{10}, 0)$ is the leading $q \times q$ submatrix of $V(\beta_0)$ with $\beta_{20} = 0$. The following theorem shows that $\hat{\beta}_n$ is root- n consistent if $\lambda_n \rightarrow 0$ at an appropriate rate.

Theorem 3.1 Assume that $(Z_1, T_1, C_1), \dots, (Z_n, T_n, C_n)$ are independently and identically distributed, and that C_i is independent of T_i or conditionally independent of T_i given Z_i . If $\sqrt{n}\lambda_n = O_p(1)$, then the adaptive Lasso estimator satisfies $\|\hat{\beta}_n - \beta_0\| = O_p(n^{-1/2})$.

Proof As mentioned earlier, in the case of independent censoring, the log partial likelihood is

$$l_1(\beta) = \sum_{i=1}^n \int_0^\tau -\beta' Z_i^*(t) dN_i(t) - \int_0^\tau \log \left(\sum_{i=1}^n Y_i(t) e^{-\beta' Z_i^*(t)} \right) d\bar{N}(t). \tag{3.4}$$

By Theorem 4.1 and Lemma 3.1 of Andersen and Gill [14], it follows that for each β in a neighbourhood of β_0 ,

$$\frac{l_1(\beta) - l_1(\beta_0)}{n} = \int_0^\tau \left[(\beta - \beta_0)' s^{(1)}(\beta_0, t) - \log \left(\frac{s^{(0)}(\beta, t)}{s^{(0)}(\beta_0, t)} \right) s^{(0)}(\beta_0, t) \right] \lambda_0(t) dt + O_p \left(\frac{\|\beta - \beta_0\|}{\sqrt{n}} \right).$$

It is sufficient to show that for any given $\varepsilon > 0$, there exists a large constant K such that

$$P \left\{ \sup_{\|u\|=K} Q_1(\beta_0 + n^{-1/2}u) < Q_1(\beta_0) \right\} \geq 1 - \varepsilon. \tag{3.5}$$

where $u = (u_1, \dots, u_p)'$. This implies with probability at least $1 - \varepsilon$ that there exists a local maximum in the ball $B_n(K) = \{\beta_0 + n^{-1/2}u, \|u\| \leq K\}$, $K > 0$. Hence, there exists a local maximizer such that $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$.

In the case of independent censoring, because $U_1(\beta_0)/\sqrt{n} \rightarrow N\{0, \Omega(\beta_0)\}$ in distribution and $-\mathcal{H}_1(\beta_0)/n \rightarrow \Omega(\beta_0)$ in probability, we can get $U_1(\beta_0)/\sqrt{n} = O_p(1)$ and $-\mathcal{H}_1(\beta_0)/n = \Omega(\beta_0) + o_p(1)$. For any $\beta \in \partial B_n(K)$, where $\partial B_n(K)$ denotes $B_n(K)$'s boundary, by the second-order Taylor expansion of the log partial likelihood, we have

$$\begin{aligned} \frac{1}{n} (l_1(\beta_0 + n^{-1/2}u) - l_1(\beta_0)) &= \frac{1}{n} U_1'(\beta_0) n^{-1/2}u - \frac{1}{2n} u' \{-\mathcal{H}_1(\beta_0)/n\}u + \frac{1}{n} u' o_p(1)u \\ &= -\frac{1}{2n} u' \{\Omega(\beta_0) + o_p(1)\}u + \frac{1}{n} O_p(1) \sum_{j=1}^p |u_j|. \end{aligned}$$

Then we have

$$\begin{aligned} &\frac{1}{n} (Q_1(\beta_0 + n^{-1/2}u) - Q_1(\beta_0)) \\ &= \frac{1}{n} \{l_1(\beta_0 + n^{-1/2}u) - l_1(\beta_0)\} - \lambda_n \sum_{j=1}^p \left(\frac{|\beta_{j0} + n^{-1/2}u_j|}{|\tilde{\beta}_j|} - \frac{|\beta_{j0}|}{|\tilde{\beta}_j|} \right) \\ &\leq \frac{1}{n} \{l_1(\beta_0 + n^{-1/2}u) - l_1(\beta_0)\} - \lambda_n \sum_{j=1}^q (|\beta_{j0} + n^{-1/2}u_j| - |\beta_{j0}|) / |\tilde{\beta}_j| \\ &\leq \frac{1}{n} \{l_1(\beta_0 + n^{-1/2}u) - l_1(\beta_0)\} + n^{-1/2} \lambda_n \sum_{j=1}^q |u_j| / |\tilde{\beta}_j| \\ &= -\frac{1}{2n} u' \{\Omega(\beta_0) + o_p(1)\}u + \frac{1}{n} O_p(1) \sum_{j=1}^p |u_j| + \frac{1}{\sqrt{n}} \lambda_n \sum_{j=1}^q |u_j| / |\tilde{\beta}_j|. \end{aligned} \tag{3.6}$$

In the case of dependent censoring, we can write

$$\frac{1}{n} (l_2(\beta_0 + n^{-1/2}u) - l_2(\beta_0)) = -\frac{1}{2n} u' (\Omega_\beta + o_p(1)) u + \frac{1}{n} O_p(1) \sum_{j=1}^p |u_j|.$$

Then we have

$$\frac{Q_2(\beta_0 + u/\sqrt{n}) - Q_2(\beta_0)}{n} \leq -\frac{1}{2n} u' [\Omega_\beta + o_p(1)]u + \frac{1}{n} O_p(1) \sum_{j=1}^p |u_j| + \frac{1}{\sqrt{n}} \lambda_n \sum_{j=1}^q \frac{|u_j|}{|\tilde{\beta}_j|}. \tag{3.7}$$

Since the maximum partial likelihood estimator $\tilde{\beta}$ satisfies $\|\tilde{\beta} - \beta_0\| = O_p(n^{-1/2})$, by the Taylor expansion, we have, for $1 \leq j \leq q$,

$$\frac{1}{|\tilde{\beta}_j|} = \frac{1}{|\beta_{j0}|} - \frac{\text{sign}(\beta_{j0})}{\beta_{j0}^2} (\tilde{\beta}_j - \beta_{j0}) + o_p(|\tilde{\beta}_j - \beta_{j0}|) = \frac{1}{|\beta_{j0}|} + \frac{O_p(1)}{\sqrt{n}}.$$

In addition, since $\sqrt{n}\lambda_n = O_p(1)$, we have

$$\frac{1}{\sqrt{n}} \lambda_n \sum_{j=1}^q |u_j| / |\tilde{\beta}_j| = \frac{1}{\sqrt{n}} \sum_{j=1}^q \left(\frac{|u_j|}{\beta_{j0}} + \frac{|u_j|}{\sqrt{n}} O_p(1) \right) \leq K n^{-1/2} \lambda_n O_p(1) = K n^{-1} O_p(1).$$

Therefore in (3.6) or (3.7), if we choose a sufficiently large K , the first term is of the order K^2n^{-1} . The second and third terms are of the order Kn^{-1} , which are dominated by the first term. Therefore (3.5) holds and it completes the proof.

If the λ_n is chosen properly, the adaptive Lasso estimator has the oracle property. There are the properties we will show next.

Theorem 3.2 Assume that $\sqrt{n}\lambda_n \rightarrow \lambda_0$ and $n\lambda_n \rightarrow \infty$. Then, under the conditions of Theorem 3.1, with probability tending to 1, the root- n consistent adaptive Lasso estimator $\hat{\beta}_n$ must satisfy the following conditions:

- (1) (Sparsity) $\hat{\beta}_{2n} = 0$;
- (2) (Asymptotic normality) $\sqrt{n}(\hat{\beta}_{1n} - \beta_{10})$ converges in distribution to the normal distribution of $N(0, \Omega_1^{-1}(\beta_{10}))$ for the independent censoring case, or $N(0, V_1(\beta_{10}))$ for the dependent censoring case.

Proof (1) Here we show that $\hat{\beta}_{2n} = 0$. It is sufficient to show that, for any sequence β_1 satisfying $\|\beta_1 - \beta_{10}\| = O_p(n^{-1/2})$ and for any constant K ,

$$Q_i(\beta_1, 0) = \max_{\|\beta_2\| \leq Kn^{-1/2}} Q_i(\beta_1, \beta_2), \quad i = 1, 2.$$

We will show that, with probability tending to 1, for any β_1 satisfying $\|\beta_1 - \beta_{10}\| = O_p(n^{-1/2})$, $\partial Q_i(\beta)/\partial \beta_j$ and β_j have different signs for $\beta_j \in (-Kn^{-1/2}, Kn^{-1/2})$ with $j = q + 1, \dots, p$. For each β in a neighbourhood of β_0 , by Taylor expansion,

$$l_i(\beta) = l_i(\beta_0) + nf_i(\beta) + O_p(\sqrt{n}\|\beta - \beta_0\|), \quad i = 1, 2,$$

where $f_1(\beta) = -\frac{1}{2}(\beta - \beta_0)'(\Omega(\beta_0) + o(1))(\beta - \beta_0)$ or $f_2(\beta) = -\frac{1}{2}(\beta - \beta_0)'(\Omega_\beta + o(1))(\beta - \beta_0)$. For $j = q + 1, \dots, p$, we have

$$\frac{\partial Q_i(\beta)}{\partial \beta_j} = \frac{\partial l_i(\beta)}{\partial \beta_j} - n\lambda_n \frac{\text{sign}(\beta_j)}{|\tilde{\beta}_j|} = O_p(n^{1/2}) - (n\lambda_n)n^{1/2} \frac{\text{sign}(\beta_j)}{|n^{1/2}\tilde{\beta}_j|}.$$

Note that $n^{1/2}(\tilde{\beta}_j - 0) = O_p(1)$, so that we have

$$\frac{\partial Q_i(\beta)}{\partial \beta_j} = n^{1/2} \left(O_p(1) - n\lambda_n \frac{\text{sign}(\beta_j)}{|O_p(1)|} \right). \tag{3.8}$$

Since $n\lambda_n \rightarrow \infty$, the sign of $\frac{\partial Q_i(\beta)}{\partial \beta_j}$ in (3.8) is completely determined by the sign of β_j when n is large, and they always have different signs.

(2) We need to show the asymptotic normality of $\hat{\beta}_{1n}$. From the proof of Theorem 3.1, it is easy to show that there exists a root- n consistent maximizer $\hat{\beta}_{1n}$ of $Q_i(\beta_1, 0)$, i.e.

$$\left. \frac{\partial Q_i(\beta)}{\partial \beta_1} \right|_{\beta_1 = (\hat{\beta}'_{1n}, 0)'} = 0.$$

In the case of independent censoring, let $U_{11}(\beta)$ be the first q elements of $U_1(\beta)$ and let $\hat{I}_{11}(\beta)$ be the first $q \times q$ submatrix of $-\mathcal{H}_1(\beta)$. Then

$$\begin{aligned} 0 &= \left. \frac{\partial Q_1(\beta)}{\partial \beta_1} \right|_{\beta_1=(\hat{\beta}'_{1n}, 0)'} = \left. \frac{\partial l_1(\beta)}{\partial \beta_1} \right|_{\beta_1=(\hat{\beta}'_{1n}, 0)'} - n\lambda_n \left(\frac{\text{sign}(\hat{\beta}_1)}{\tilde{\beta}_1}, \dots, \frac{\text{sign}(\hat{\beta}_q)}{\tilde{\beta}_q} \right)' \\ &= U_{11}(\beta_0) - \hat{I}_{11}(\beta^*)(\hat{\beta}_{1n} - \beta_{10}) - n\lambda_n \left(\frac{\text{sign}(\beta_{10})}{\tilde{\beta}_1}, \dots, \frac{\text{sign}(\beta_{q0})}{\tilde{\beta}_q} \right)', \end{aligned}$$

where β^* is between $\hat{\beta}_n$ and β_0 . The last equation is implied by $\text{sign}(\hat{\beta}_{jn}) = \text{sign}(\beta_{j0})$ when n is large. Using Theorem 3.2 of Andersen and Gill [14], we can prove that $U_{11}(\beta_0)/\sqrt{n} \rightarrow N\{0, \Omega_1(\beta_0)\}$ in distribution and $\hat{I}_{11}(\beta^*)/n \rightarrow \Omega_1(\beta_{10})$ in probability as $n \rightarrow \infty$. Furthermore, if $n \rightarrow \infty$ and $\sqrt{n}\lambda_n \rightarrow \lambda_0$, a nonnegative constant, we have

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) = \Omega_1^{-1}(\beta_{10}) \left(\frac{1}{\sqrt{n}} U_{11}(\beta_0) - \lambda_0 b_1 \right) + o_p(1)$$

with $b_1 = \left(\frac{\text{sign}(\beta_{10})}{|\beta_{10}|}, \dots, \frac{\text{sign}(\beta_{q0})}{|\beta_{q0}|} \right)'$, since $\tilde{\beta}_j \rightarrow \beta_{j0} \neq 0$ for $1 \leq j \leq q$. Then by Slutsky's Theorem, $\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) \rightarrow N(-\lambda_0 \Omega_1^{-1}(\beta_{10}) b_1, \Omega_1^{-1}(\beta_{10}))$ in distribution as $n \rightarrow \infty$. In particular, if $n \rightarrow \infty$ and $\sqrt{n}\lambda_n \rightarrow 0$, we have

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) \xrightarrow{d} N(0, \Omega_1^{-1}(\beta_{10})),$$

where \xrightarrow{d} means converging in distribution.

In the case of dependent censoring, let $U_{21}(\beta; \gamma)$ be the first q elements of $U_2(\beta; \gamma)$ and let $\hat{I}_{11}(\beta; \gamma)$ be the first $q \times q$ submatrix of $\hat{\Omega}_\beta(\beta; \gamma)$. Then

$$\begin{aligned} 0 &= \left. \frac{\partial Q_2(\beta)}{\partial \beta_1} \right|_{\beta_1=(\hat{\beta}'_{1n}, 0)'} = \left. \frac{\partial l_2(\beta)}{\partial \beta_1} \right|_{\beta_1=(\hat{\beta}'_{1n}, 0)'} - n\lambda_n \left(\frac{\text{sign}(\hat{\beta}_1)}{\tilde{\beta}_1}, \dots, \frac{\text{sign}(\hat{\beta}_q)}{\tilde{\beta}_q} \right)' \\ &= U_{21}(\beta; \hat{\gamma}) - \hat{I}_{11}(\beta^*; \hat{\gamma})(\hat{\beta}_{1n} - \beta_{10}) - n\lambda_n \left(\frac{\text{sign}(\beta_{10})}{\tilde{\beta}_1}, \dots, \frac{\text{sign}(\beta_{q0})}{\tilde{\beta}_q} \right)', \end{aligned}$$

where β^* is between $\hat{\beta}_n$ and β_0 . The last equation is implied by $\text{sign}(\hat{\beta}_{jn}) = \text{sign}(\beta_{j0})$ when n is large. Let $M_1(\beta_{10}) = M_{11}(\beta_{10}, 0)$, where $M_{11}(\beta_{10}, 0)$ is the leading $q \times q$ submatrix of $M(\beta_0)$ with $\beta_{20} = 0$ and $\Omega_{\beta 1}(\beta_{10}) = \Omega_{\beta_{11}}(\beta_{10}, 0)$, where $\Omega_{\beta_{11}}(\beta_{10}, 0)$ is the leading $q \times q$ submatrix of Ω_β with $\beta_{20} = 0$. Since $U_{21}(\beta_0; \hat{\gamma})/\sqrt{n} \rightarrow N(0, M_1(\beta_0))$ in distribution and $\hat{I}_{11}(\beta^*)/n \rightarrow \Omega_{\beta 1}(\beta_{10})$ in probability as $n \rightarrow \infty$. Furthermore, if $\sqrt{n}\lambda_n \rightarrow \lambda_0$, a nonnegative constant, we have

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) = \Omega_{\beta 1}^{-1}(\beta_{10}) \left(\frac{1}{\sqrt{n}} U_{21}(\beta_0; \hat{\gamma}) - \lambda_0 b_1 \right) + o_p(1)$$

with $b_1 = \left(\frac{\text{sign}(\beta_{10})}{|\beta_{10}|}, \dots, \frac{\text{sign}(\beta_{q0})}{|\beta_{q0}|} \right)^T$, since $\tilde{\beta}_j \rightarrow \beta_{j0} \neq 0$ for $1 \leq j \leq q$. Then by Slutsky's Theorem, $\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) \rightarrow N(-\lambda_0 \Omega_{\beta 1}^{-1}(\beta_{10}) b_1, V_1(\beta_{10}))$ in distribution as $n \rightarrow \infty$. In

particular, if $\sqrt{n}\lambda_n \rightarrow 0$, we have

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_{10}) \rightarrow N(0, V_1(\beta_{10}))$$

in distribution as $n \rightarrow \infty$.

Remark It is worth noting that as n goes to infinity, the adaptive Lasso can perform as well as the correct submodel was known. Since the proofs only require the root- n consistency of $\tilde{\beta}$, any root- n consistent estimator of β_0 can be used as the adaptive weight ρ without changing the asymptotic properties.

4 Computational Algorithm

The optimization problem (3.1) or (3.2) is strictly convex and therefore can be solved by many convex optimization algorithm. Here we present an algorithm based on the Alternating Direction Method of Multipliers (ADMM)^[15]. The ADMM algorithm solves problem in the form

$$\begin{aligned} & \text{minimize} && f(x) + g(z) \\ & \text{subject to} && Ax + Bz = c \end{aligned}$$

with variables $x \in R^n$ and $z \in R^m$, where $A \in R^{p \times n}$, $B \in R^{p \times m}$, and $c \in R^p$. The augmented Lagrangian is

$$L_\rho(x, z, y) = f(x) + g(z) + y'(Ax + Bz - c) + (\rho/2)\|Ax + Bz - c\|_2^2.$$

ADMM consists of the iterations

$$\begin{aligned} x^{k+1} &= \arg \min_x L_\rho(x, z^k, y^k), \\ z^{k+1} &= \arg \min_z L_\rho(x^{k+1}, z, y^k), \\ y^{k+1} &= y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \end{aligned}$$

with $\rho > 0$.

In ADMM form, the problem (3.1) or (3.2) can be written as

$$\begin{aligned} & \min && f(\beta) + g(z) \\ & \text{s.t.} && \beta - z = 0, \end{aligned}$$

where $f(\beta)$ is equal to $-l_1(\beta)/n$ or $-l_2(\beta)/n$, and $g(z) = \lambda \sum_{j=1}^p |z_j| \omega_j$. The updates performed by the algorithm during each iteration are

$$\begin{aligned} \beta^{k+1} &= \arg \min_\beta (f(\beta) + \rho(u^k)'(\beta - z^k) + (\rho/2)\|\beta - z^k\|_2^2), \\ z_i^{k+1} &= S_{\frac{\lambda \omega_i}{\rho}}(\beta_i^{k+1} + u_i^k), \quad i = 1, \dots, p, \\ u^{k+1} &= u^k + \beta^{k+1} - z^{k+1}, \end{aligned}$$

where S is the soft thresholding operator satisfying

$$S_{\kappa}(a) = \begin{cases} a - \kappa, & a > \kappa, \\ 0, & |a| \leq \kappa, \\ a + \kappa, & a < -\kappa. \end{cases}$$

The β -update can be done by solving the equation $-\frac{U_i(\beta)}{n} + u^k \rho + \rho(\beta - z^k) = 0$, $i = 1, 2$. To solve the equation, there are many standard methods, such as the Newton-Raphson method.

This algorithm gives very small values to the coefficients which should be estimated as zero and it converges quickly based on our empirical experience.

5 A Simulation Study

In this section, we examine the performance of the adaptive Lasso method under the additive hazards model and as a comparison, Lasso, smoothly clipped absolute deviation (SCAD), maximum partial likelihood estimators (MPLE) are also considered. For given p , the covariates Z are assumed to follow the multivariate normal distribution with mean zero, variance one, and the correlation between Z_j and Z_k being $\rho^{|j-k|}$ with $\rho = 0.5$, $j, k = 1, \dots, p$. We set $\beta_{0j} = 1$ for the first and last two components of the covariates and $\beta_{0j} = 0$ for other components. The results given below are based on sample size $n = 300$ and 500 replications.

To measure prediction accuracy, we define the mean weighted squared error (MWSE) to be $(\hat{\beta} - \beta_0)' E(ZZ') (\hat{\beta} - \beta_0)$. Besides MWSE, we also use the averaged number of nonzero estimates of parameters whose true values are not zero (TP): $TP = \sum_{i=1}^p I(\beta_{0i} \neq 0) I(\hat{\beta}_i \neq 0)$, and the averaged number of nonzero estimates of parameters whose true values are zero (FP): $FP = \sum_{i=1}^p I(\beta_{0i} = 0) I(\hat{\beta}_i \neq 0)$. It is easy to see that TP and FP provide the estimates of the true and false positive probabilities, respectively. For the selection of the tuning parameters in the proposed method, we use the Bayesian information criterion based on $BIC(\lambda) = -2l_i(\hat{\beta}) + q_n \times \log(n)$, for $i = 1$ or 2 with q_n denoting the number of the nonzero β estimates. Then one choose the values of λ that minimize $BIC(\lambda)$.

Table 1 displays the results on the covariate selection with $p = 10$ or 20 in the case of independent censoring. In this case, the failure times T_i are generated from model (2.1) with $\lambda_0 = 0.5$ or 1. For the observation times C_i , we generated it from the uniform distribution over $(0, 3.5)$ and the exponential distribution with parameter $\lambda = 0.5$ or 0.7. One can see from Table 1 that the aLasso approach gives the smallest FP compared with other methods which means the aLasso chooses unimportant variables much less often than the other methods. At the same time, it kept a fairly high TP and low MWSE. The SCAD method gave the largest TP in most cases among the method considered here.

Table 2 displays the results on the covariate selection with $p = 10$ or 20 in the case of dependent censoring. In this case, we consider different combinations of λ_0, λ_c and γ_0 . Here, we set all components of γ_0 to be the same, for example, in Table 2, $\gamma_0 = 0.1$ means $\gamma'_0 = (0.1, 0.1, \dots, 0.1, 0.1)$ in model (2.3). Keeping γ_0 unchanged, we list four combinations

of λ_0 and λ_c in each part, which corresponds to $\lambda_0 = 0.5$ or 1 , $\lambda_c = 0.5$ or 0.7 . As in the case of independent censoring, the aLasso approach gave the smallest FP in all dependent cases.

Also, it can be seen from Tables 1–2 that, as the number of covariates increases, the aLasso tends to give the smallest MWSE and largest TP among the methods considered. Overall, the adaptive Lasso performs well in terms of both variable selection and prediction accuracy.

Table 1. Results in the case of independent censoring.

| C | λ_0 | Method | $p = 10$ | | | $p = 20$ | | |
|-------------------|-------------|--------|----------|--------|--------|----------|--------|--------|
| | | | MWSE | TP | FP | MWSE | TP | FP |
| $U(0, 3.5)$ | 0.5 | aLasso | 0.3443 | 3.9960 | 0.3160 | 0.4409 | 3.9960 | 0.7940 |
| | | Lasso | 0.5292 | 3.9940 | 1.0020 | 0.9642 | 3.9780 | 1.8360 |
| | | SCAD | 0.2264 | 3.9900 | 0.6800 | 0.3920 | 3.9640 | 1.2280 |
| | | MPLE | 0.4964 | 4 | 6 | 1.1390 | 4 | 16 |
| | 1 | aLasso | 0.3612 | 3.9900 | 0.3420 | 0.4856 | 3.9860 | 0.8540 |
| | | Lasso | 0.5784 | 3.9880 | 1.0400 | 1.1075 | 3.9740 | 1.7060 |
| | | SCAD | 0.2610 | 3.9540 | 0.7120 | 0.3982 | 3.9660 | 1.4940 |
| | | MPLE | 0.5784 | 4 | 6 | 1.3472 | 4 | 16 |
| $\text{Exp}(0.5)$ | 0.5 | aLasso | 0.5889 | 3.9940 | 0.7060 | 0.8958 | 3.9680 | 1.1780 |
| | | Lasso | 1.0489 | 3.9620 | 1.6140 | 1.2094 | 3.9340 | 1.9320 |
| | | SCAD | 0.4166 | 3.9980 | 1.7320 | 0.9359 | 3.9460 | 1.9280 |
| | | MPLE | 0.5016 | 4 | 6 | 1.2456 | 4 | 16 |
| | 1 | aLasso | 0.6726 | 3.9760 | 0.6400 | 0.9977 | 3.9500 | 1.2000 |
| | | Lasso | 1.1606 | 3.9340 | 1.4240 | 1.3246 | 3.9280 | 1.9020 |
| | | SCAD | 0.4685 | 3.9840 | 1.4240 | 0.9425 | 3.9520 | 1.9620 |
| | | MLE | 0.5849 | 4 | 6 | 1.4114 | 4 | 16 |
| $\text{Exp}(0.7)$ | 0.5 | aLasso | 0.6137 | 3.9760 | 0.7320 | 0.9191 | 3.9380 | 1.0860 |
| | | Lasso | 1.0797 | 3.9500 | 1.4600 | 1.2501 | 3.8940 | 1.8780 |
| | | SCAD | 0.4181 | 3.9840 | 1.4380 | 0.9047 | 3.9240 | 1.9300 |
| | | MPLE | 0.5212 | 4 | 6 | 1.2681 | 4 | 16 |
| | 1 | aLasso | 0.6456 | 3.9660 | 0.6320 | 0.9614 | 3.9260 | 1.0840 |
| | | Lasso | 1.1521 | 3.9240 | 1.4080 | 1.3458 | 3.8920 | 1.8120 |
| | | SCAD | 0.4661 | 3.9800 | 1.1540 | 0.9514 | 3.9160 | 1.9120 |
| | | MPLE | 0.5435 | 4 | 6 | 1.3800 | 4 | 16 |

Table 2. Results in the case of dependent censoring.

| (λ_0, λ_c) | γ_0 | Method | $p = 10$ | | | $p = 20$ | | |
|--------------------------|------------|--------|----------|--------|--------|----------|--------|--------|
| | | | MWSE | TP | FP | MWSE | TP | FP |
| (0.5, 0.5) | 0.1 | aLasso | 0.5257 | 3.9960 | 0.4860 | 0.8838 | 3.9660 | 1 |
| | | Lasso | 0.9696 | 3.9720 | 1.3640 | 1.1813 | 3.9420 | 1.5380 |
| | | SCAD | 0.3535 | 3.9960 | 1.5260 | 0.8042 | 3.9580 | 1.5020 |
| | | MPLE | 0.4444 | 4 | 6 | 1.0386 | 4 | 16 |
| (0.5, 0.7) | | aLasso | 0.5366 | 3.9820 | 0.4560 | 0.9294 | 3.9420 | 0.8360 |
| | | Lasso | 0.9656 | 3.9740 | 1.2400 | 1.3217 | 3.8740 | 1.4360 |
| | | SCAD | 0.3610 | 3.9920 | 1.1580 | 0.9002 | 3.9120 | 1.4460 |
| | | MPLE | 0.4326 | 4 | 6 | 1.0446 | 4 | 16 |
| (1, 0.5) | | aLasso | 0.5542 | 3.9900 | 0.5740 | 0.9412 | 3.9480 | 0.8500 |
| | | Lasso | 0.9936 | 3.9800 | 1.3420 | 1.2948 | 3.9220 | 1.4580 |
| | | SCAD | 0.3314 | 3.9960 | 1.3560 | 0.8923 | 3.9380 | 1.5180 |
| | | MPLE | 0.4490 | 4 | 6 | 1.2111 | 4 | 16 |
| (1, 0.7) | | aLasso | 0.5439 | 3.9740 | 0.5400 | 0.9224 | 3.9560 | 0.8480 |
| | | Lasso | 1.0439 | 3.9480 | 1.1720 | 1.3822 | 3.8820 | 1.3900 |
| | | SCAD | 0.3190 | 3.9820 | 1.0300 | 0.8903 | 3.9380 | 1.4640 |
| | | MPLE | 0.4490 | 4 | 6 | 1.1385 | 4 | 16 |
| (0.5, 0.5) | 0.2 | aLasso | 0.5628 | 3.9860 | 0.5600 | 1.2515 | 3.8780 | 1.0840 |
| | | Lasso | 1.0970 | 3.9620 | 1.4560 | 1.7642 | 3.7140 | 1.6320 |
| | | SCAD | 0.3567 | 3.9920 | 1.4860 | 1.2253 | 3.8020 | 1.7140 |
| | | MPLE | 0.4404 | 4 | 6 | 1.3247 | 4 | 16 |
| (0.5, 0.7) | | aLasso | 0.6401 | 3.9700 | 0.5400 | 1.3662 | 3.8440 | 0.9880 |
| | | Lasso | 1.2872 | 3.9020 | 1.2940 | 2.1243 | 3.4860 | 1.4780 |
| | | SCAD | 0.3679 | 3.9900 | 1.1460 | 1.3346 | 3.7240 | 1.7060 |
| | | MPLE | 0.4730 | 4 | 6 | 1.5113 | 4 | 16 |
| (1, 0.5) | | aLasso | 0.6465 | 3.9680 | 0.5740 | 1.3275 | 3.8340 | 1.1180 |
| | | Lasso | 1.2081 | 3.9180 | 1.3240 | 1.8881 | 3.6540 | 1.5180 |
| | | SCAD | 0.3799 | 3.9820 | 1.3160 | 1.3251 | 3.7780 | 1.6320 |
| | | MPLE | 0.4831 | 4 | 6 | 1.5177 | 4 | 16 |
| (1, 0.7) | | aLasso | 0.6896 | 3.9540 | 0.5140 | 1.4032 | 3.7900 | 1.1240 |
| | | Lasso | 1.3210 | 3.9020 | 1.2460 | 2.1546 | 3.4760 | 1.3720 |
| | | SCAD | 0.3527 | 3.9740 | 0.9760 | 1.3999 | 3.7060 | 1.6340 |
| | | MPLE | 0.4929 | 4 | 6 | 1.6897 | 4 | 16 |

Table 2 Continued.

| (λ_0, λ_c) | γ_0 | Method | $p = 10$ | | | $p = 20$ | | |
|--------------------------|------------|--------|----------|--------|--------|----------|--------|--------|
| | | | MWSE | TP | FP | MWSE | TP | FP |
| (0.5, 0.5) | 0.3 | aLasso | 0.7607 | 3.9480 | 0.6220 | 1.8972 | 3.6700 | 1.3160 |
| | | Lasso | 1.5474 | 3.8240 | 1.4420 | 2.8727 | 3.0960 | 1.4220 |
| | | SCAD | 0.4513 | 3.9680 | 1.3200 | 2.0540 | 3.3800 | 1.7120 |
| | | MPLE | 0.4819 | 4 | 6 | 2.1382 | 4 | 16 |
| (0.5, 0.7) | | aLasso | 0.9324 | 3.8900 | 0.5620 | 2.1734 | 3.5140 | 1.3160 |
| | | Lasso | 1.9573 | 3.5880 | 1.3580 | 3.3883 | 2.6800 | 1.2260 |
| | | SCAD | 0.5280 | 3.9500 | 1.0560 | 2.4260 | 3.1020 | 1.5700 |
| | | MPLE | 0.5491 | 4 | 6 | 2.2098 | 4 | 16 |
| (1, 0.5) | | aLasso | 0.7896 | 3.9500 | 0.5740 | 1.9295 | 3.6220 | 1.4860 |
| | | Lasso | 1.6051 | 3.8000 | 1.2680 | 2.9691 | 2.9920 | 1.3700 |
| | | SCAD | 0.4170 | 3.9660 | 1.1680 | 2.0186 | 3.3840 | 1.7700 |
| | | MPLE | 0.5317 | 4 | 6 | 2.3245 | 4 | 16 |
| (1, 0.7) | | aLasso | 0.9099 | 3.9080 | 0.5620 | 2.2117 | 3.4420 | 1.3220 |
| | | Lasso | 1.9178 | 3.6220 | 1.2320 | 3.2215 | 2.7740 | 1.2900 |
| | | SCAD | 0.4330 | 3.9460 | 1.0620 | 2.4330 | 3.1260 | 1.6200 |
| | | MPLE | 0.5692 | 4 | 6 | 2.7490 | 4 | 16 |

6 An Application

In this section, we apply the proposed regression selection procedure to a set of data on mice hepatocellular adenoma. This data set arises from a 2-year tumorigenicity study conducted by National Toxicology Program. In the study, groups of mice were exposed to chloroprene at different concentrations by inhalation. Each mouse was examined once for various tumors when it died. Some mice died naturally during the study, and the others who survived at the end of study were sacrificed for examinations. At each examination time, tumors were observed if have developed, but the exact tumor onset times were unknown, therefore, only current status data can be obtained.

Here we considered the liver tumor data, and the covariates on which the information was collected include the initial weight of the mouse, the body weight change, the weight change rate, the gender of the mouse, the dose. For the analysis below, we will focus on 200 mice that either belong to the control group or belong to the PPM80 group.

To apply the aLasso regression procedure, let IW denote the initial weight of the mouse, BWC denote the body weight change and BWCR denote the weight change rate. We de-

fine Gender=1 if the mouse was male and 0 otherwise, PPM80=0 if the mouse was in the control group and 1 otherwise. For the analysis, we performed the standardization on the three continuous covariates IW, BWC and BWCR. The analysis results given by the aLasso procedure are presented in Table 3. As in the simulation study and for comparison, we also include the analysis results obtained by applying the other penalized procedures discussed here. ALasso, Lasso and SCAD all suggest that the Gender and the initial weight of the mouse had no relationship with or significant influence on the existence of hepatocellular adenoma.

Table 3 Analysis results of mice hepatocellular adenoma data.

| Method | IW | BWC | BWCR | GENDER | PPM80 |
|--------|---------|--------|---------|---------|--------|
| aLasso | 0 | 0.1538 | 0.1197 | 0 | 0.5938 |
| Lasso | 0 | 0.0452 | 0.1459 | 0 | 0.2732 |
| SCAD | 0 | 0.0561 | 0.1094 | 0 | 0.6469 |
| MPLE | -0.2544 | 1.1425 | -1.0347 | -0.1879 | 0.7037 |

7 Concluding Remarks

This paper has discussed the variable selection problem for the additive hazards model based on current status data. In order to select important variables, a penalized log partial likelihood method is developed and the oracle properties are provided. The simulated results suggest that the proposed method performs well for dropping the unimportant variables and retaining the important variables.

As mentioned above, the proposed method can be seen as a generalization of the method given in Zhang and Lu [1], for the case that the model is proportional hazards model and the data is right-censored data. Therefore it could be generalized in several directions. For one, note that in the preceding sections, we assume that C is independent of Z and T , it is straightforward to generalize the proposed method to the case where the censoring time C is not independent of Z or other type data.

The second direction is that we can change the weights ρ_j with other estimators since the proofs only require the root- n consistency of $\tilde{\beta}$.

References

- [1] Zhang, H.H., Lu, W. Adaptive-lasso for cox's proportional hazard model [J]. *Biometrika*, 2007, 94: 691–703.
- [2] Faraggi, D., Simon, R. Bayesian variable selection method for censored survival data [J]. *Biometrics*, 1998, 54: 1475–1485.
- [3] Fan, J., Li, R. Variable selection for Cox's proportional hazards model and frailty model [J]. *Ann. Statist.*, 2002, 30: 74–99.

- [4] Tibshirani, R. Regression shrinkage and selection via the lasso [J]. *Journal of the Royal Statistical Society Series B*, 1996, 58: 267–288.
- [5] Tibshirani, R. The Lasso method for variable selection in the Cox model [J]. *Stat. Med.*, 1997, 16: 385–396.
- [6] Fan, J., Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *J. Amer. Statist. Assoc.*, 2001, 96: 1348–1360.
- [7] Zou, H. The adaptive lasso and its oracle properties [J]. *J. Amer. Statist. Assoc.*, 2006, 101: 1418–1429.
- [8] Huang, J., Ma, S. Variable selection in accelerated failure time model via the bridge method [J]. *Lifetime Data Analysis*, 2010, 16: 176–195.
- [9] Huang, J., Sun, T., Ying, Z., Yu, Y. and Zhang, C.H. Oracle inequalities for the lasso in the Cox model [J]. *Ann. Statist.*, 2013, 41: 1142–1165.
- [10] Zhao, H., Wu, Q., Li, G., Sun, J. Simultaneous estimation and variable selection for interval-censored data with broken adaptive ridge regression [J]. *J. Amer. Statist. Assoc.*, 2020, 115: 204–216.
- [11] Lin, D., Oakes, D., Ying, Z. Additive hazards regression with current status data [J]. *Biometrika*, 1998, 85: 289–298.
- [12] Wang, L., Sun, J., Tong, X. Regression analysis of case II interval censored failure time data with the additive hazards model [J]. *Statistica Sinica*, 2010, 20: 1709–1723.
- [13] Feng, Y., Ma, L., Sun, J. Additive hazards regression with auxiliary covariates for case I interval-censored data [J]. *Scand. J. Statist.*, 2015, 42: 118–136.
- [14] Andersen, P.K., Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study [J]. *Ann. Statist.*, 1982, 10: 1100–1120.
- [15] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends in Machine Learning*, 2010, 3(1): 1–122.

基于当前状态数据的加法风险模型的自适应LASSO回归选择

张悻瑾¹, 王成勇²

(1. 武汉大学数学与统计学院, 湖北 武汉 430072)

(2. 湖北文理学院数学与统计学院, 湖北 襄阳 441053)

摘要: 当真实的潜在模型具有稀疏表示时通常需要使用变量选择方法, 确定模型中的重要预测因子可提高被拟合模型的预测性能, 许多文献研究了这类问题, 其中张和吕[1]针对右删失数据开发了一种基于比例风险模型的变量选择方法. 本文研究了基于当前状态数据的加法风险模型的变量选择问题. 在文献[1]的启发下, 我们提出一种自适应Lasso方法来解决这个问题, 并在弱正则性条件下, 建立了估计量的相合性和oracle性质等理论结果. 大量的模拟数据分析证明了该方法的有效性. 我们用该方法分析了一组来自肿瘤研究的真实数据.

关键词: 加法风险模型; 当前状态数据; 自适应Lasso; ADMM算法

MR(2010)主题分类号: 62N02; 62F10; 62F12 **中图分类号:** O212.1