Statistics and Probability Letters 135 (2018) 20–25

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Screening group variables in the proportional hazards model

Kwang Woo Ahn *, Natasha Sahr, Soyoung Kim

Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53132, USA

ARTICLE INFO

Article history:
Received 5 June 2017
Received in revised form 10 August 2017
Accepted 29 November 2017
Available online 13 December 2017

MSC:
00-01
99-00

Keywords:
Proportional hazards model
Sure screening property
Joint screening
High dimensional data

ABSTRACT

We propose a method to screen group variables under the high dimensional group variable setting for the proportional hazards model. We study the sure screening property of the proposed method for independent and clustered survival data. The simulation study shows that the proposed method performs better for group variable screening than some existing procedures.

1. Introduction

There is rich literature on group and within-group variable selection with possibly overlapping groups for the proportional hazards (PH) model. Huang et al. (2014) proposed the group bridge for bi-level selection when the number of covariates p is less than the number of observations n. An alternative approach by Wu and Wang (2013) considered double penalties: lasso and group lasso penalties for $p > n$. However, it lacks theoretical justification. Wang et al. (2009) proposed a hierarchically penalized Cox regression. They used an adaptive penalty to achieve the bi-level selection consistency when $p < n$. For $p > n$, they proposed to use the ridge regression to construct the adaptive penalty without theoretical justification. Although these penalized regression methods may work when p is relatively larger than n, they may be computationally expensive and unstable when p is much larger than n ($p \gg n$).

Screening techniques for the PH model with $p \gg n$ have recently been given much attention (Zhao and Li, 2012; Yang et al., 2016). However, these existing methods are all limited to individual variable screening for independent survival data to the best of the authors' knowledge. Therefore, motivated by Yang et al. (2016), we propose a sure group joint screening (SGJS) procedure to screen group variables for the PH models when $p \gg n$. We show the SGJS enjoys the sure screening property for group variable screening under independent and clustered survival data. We also propose two ways to handle overlapping group variables for screening. Simulation study is also conducted.

2. Sure group joint variable screening

2.1. Notations and assumptions

We first define the notations used throughout the rest of the paper. We assume there are m clusters and the ith cluster has L individuals. The clusters may have different sizes by defining censoring times as zero when observed times are

* Corresponding author.
E-mail addresses: kwooahn@mcw.edu (K.W. Ahn), nsahr@mcw.edu (N. Sahr), skim@mcw.edu (S. Kim).

https://doi.org/10.1016/j.spl.2017.11.014
0167-7152/

missing (Spiekerman and Lin, 1998). We denote $n = mL$ as the total sample size. It is assumed that there are p number of covariates. Let T_{ij}, C_{ij}, and $\mathbf{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})^T$ be the event time, censoring time, and covariate vector of individual j in cluster i, respectively, for $i = 1, \ldots, m$ and $j = 1, \ldots, L$. The covariate vector $\mathbf{Z}_{ij}$ may depend on time t. We omit dependency on t in $\mathbf{Z}_{ij}(t)$ when the context is clear. Without loss of generality, we assume that $\mathbf{Z}_{ij}$'s are standardized. The parameter vector of interest is denoted as $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$. Define $\boldsymbol{\beta}_0 = (\beta_{10}, \ldots, \beta_{p0})^T$ as the true parameter vector. Let $\mathbf{T}_i = \{T_{ij}, j = 1, \ldots, L\}$, $\mathbf{C}_i = \{C_{ij}, j = 1, \ldots, L\}$, and $\mathbf{Z}_i = \{\mathbf{Z}_{ij}, j = 1, \ldots, L\}$. Suppose that $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are independent and identically distributed. We assume that the T_{ij}'s are independent of the C_{ij}'s given $\mathbf{Z}_{ij}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, L$. Let $X_{ij} = T_{ij} \wedge C_{ij}$ be the observed time and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is an indicator function. Let $N_{ij}(t) = I(X_{ij} \leq t, \Delta_{ij} = 1)$ and $Y_{ij}(t) = I(X_{ij} \geq t)$. The study period is assumed to be $[0, \tau]$. Given covariate vector $\mathbf{Z}$, the hazard function $\lambda(t \mid \mathbf{Z})$ is defined as $\lambda(t \mid \mathbf{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\mathbf{Z})$, where $\lambda_0(t)$ is an unspecified baseline hazard function at time t. Then, Spiekerman and Lin (1998) proposed the marginal log-partial likelihood function for clustered survival data as follows:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{m}\sum_{j=1}^{L}\Delta_{ij}\Big[\boldsymbol{\beta}^T\mathbf{Z}_{ij} - \log\Big\{\sum_{a=1}^{m}\sum_{b=1}^{L}Y_{ab}(X_{ij})\exp(\boldsymbol{\beta}^T\mathbf{Z}_{ab})\Big\}\Big]. \tag{1}$$

Next, we define some notations on group variables and their memberships. Assume that we have K groups of variables. Let $A_1, \ldots, A_K$ be subsets of $\{1, \ldots, p\}$ representing group memberships of variables. Define $\boldsymbol{\beta}_A = (\beta_j, j \in A)^T$ and $\boldsymbol{\beta}_{A,0} = (\beta_{j0}; j \in A)^T$ for a set A. Denote $|A|$ as the cardinality of a set A. Without loss of generality, for disjoint groups we assume that β_i's are on the order of groups such that $\boldsymbol{\beta} = (\boldsymbol{\beta}_{A_1}^T, \ldots, \boldsymbol{\beta}_{A_K}^T)^T$, where $\boldsymbol{\beta}_{A_1} = (\beta_1, \ldots, \beta_{|A_1|})^T$ and $\boldsymbol{\beta}_{A_k} = (\beta_{|A_{k-1}|+1}, \ldots, \beta_{|A_{k-1}|+|A_k|})^T$ for $k = 2, \ldots, K$.

2.2. Sure group joint variable screening for disjoint groups

In this section, we study the sure group joint variable screening when A_k's are disjoint. To screen zero group variables for the PH model, we propose the following penalized partial likelihood:

$$\arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}) \quad \text{subject to number of non-zero groups } \leq q, \tag{2}$$

where q is pre-specified. We call this as the sure group joint screening (SGJS). Because $p > n$, it is not feasible to directly solve (2). Thus, motivated by the Taylor expansion of $\ell(\boldsymbol{\alpha})$ at $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\alpha}$, Yang et al. (2016) considered the following function to approximate the log-partial likelihood function:

$$g(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) + (\boldsymbol{\alpha} - \boldsymbol{\beta})^T\ell'(\boldsymbol{\beta}) - \frac{u}{2}(\boldsymbol{\alpha} - \boldsymbol{\beta})^T\mathbf{W}(\boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta}), \tag{3}$$

where u is a pre-specified constant and $\mathbf{W}(\boldsymbol{\beta})$ is a diagonal matrix consisting of the diagonal elements of $-\ell''(\boldsymbol{\beta})$. Yang et al. (2016) used $g(\boldsymbol{\alpha} \mid \boldsymbol{\beta})$ in the optimization procedure to maximize $\ell(\boldsymbol{\beta})$ with L_0 penalty.

The function $g(\boldsymbol{\alpha} \mid \boldsymbol{\beta})$ in (3) may also be used to solve (2). However, for group structured variables, variables within the same group often tend to be more correlated than variables between different groups. Motivated by this, we consider a block diagonal matrix $\mathbf{W}^*(\boldsymbol{\beta})$ instead of the diagonal matrix $\mathbf{W}(\boldsymbol{\beta})$ in (3), where each block in $\mathbf{W}^*(\boldsymbol{\beta})$ consists of the sub-square matrix of $-\ell''(\boldsymbol{\beta})$ corresponding to each group. To elaborate $\mathbf{W}^*(\boldsymbol{\beta})$, define $|A_0| = 0$. For $k = 1, \ldots, K$, we define $\mathbf{W}_k^*(\boldsymbol{\beta}_{A_k})$ as the sub-square matrix of $-\ell''(\boldsymbol{\beta})$ corresponding to $\boldsymbol{\beta}_{A_k}$ as follows:

$$\mathbf{W}_k^*(\boldsymbol{\beta}_{A_k}) = -\begin{pmatrix} \ell''(\boldsymbol{\beta})_{|A_{k-1}|+1,|A_{k-1}|+1} & \cdots & \ell''(\boldsymbol{\beta})_{|A_{k-1}|+1,|A_k|} \\ \vdots & \ddots & \vdots \\ \ell''(\boldsymbol{\beta})_{|A_k|,|A_{k-1}|+1} & \cdots & \ell''(\boldsymbol{\beta})_{|A_k|,|A_k|} \end{pmatrix},$$

where $\ell''(\boldsymbol{\beta})_{a,b}$ is the (a, b)th entry of $\ell''(\boldsymbol{\beta})$. Define $\mathbf{W}^*(\boldsymbol{\beta})$ as follows:

$$\mathbf{W}^*(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{W}_1^*(\boldsymbol{\beta}_{A_1}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2^*(\boldsymbol{\beta}_{A_2}) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_K^*(\boldsymbol{\beta}_{A_K}) \end{pmatrix},$$

where all non-block-diagonal elements are zeros. For group variable screening, we propose the following approximation at $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\alpha}$:

$$g^*(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) + (\boldsymbol{\alpha} - \boldsymbol{\beta})^T\ell'(\boldsymbol{\beta}) - \frac{u}{2}(\boldsymbol{\alpha} - \boldsymbol{\beta})^T\mathbf{W}^*(\boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta}). \tag{4}$$

For group variable screening, we propose to maximize $g^*(\boldsymbol{\alpha} \mid \boldsymbol{\beta})$ as follows:

$$\max_{\boldsymbol{\alpha}} g^*(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) \quad \text{subject to number of non-zero groups } \leq q. \tag{5}$$