中国科学: 数学 2018 年 第 48 卷 第 5 期: 643 ~ 660
SCIENTIA SINICA Mathematica
论 文

《中国科学》杂志社
SCIENCE CHINA PRESS
CrossMark
click for updates

Cox 比例风险模型中基于 SELO 惩罚函数的变量选择方法

曹永秀[1], 焦雨领[1], 石跃勇[2], 刘妍岩[3*]

1. 中南财经政法大学统计与数学学院, 武汉 430073;
2. 中国地质大学 (武汉) 经济管理学院, 武汉 430074;
3. 武汉大学数学与统计学院, 武汉 430072

E-mail: yxcao@zuel.edu.cn, yljiaostatistics@zuel.edu.cn, yueyongshi@cug.edu.cn, liuyy@whu.edu.cn

收稿日期: 2016-10-02; 接受日期: 2017-08-29; 网络出版日期: 2018-03-14; * 通信作者
国家自然科学基金 (批准号: 11501579 和 11701571)、湖北省自然科学基金 (批准号: 2016CFB486)、中央高校基本科研业务费专项资金 (批准号: CUGW150809)、中南财经政法大学研究生科研创新 (批准号: 2016J1304) 和中南财经政法大学引进人才启动金 (批准号: 31541711202) 资助项目

摘要 在生物医学研究中, 研究个体的失效时间往往存在删失, Cox 比例风险模型是经常被用来处理此类删失数据的模型. 对于带有删失的高维数据, 如何从众多协变量中挑选出少数的致病因素是研究者的兴趣所在. 本文针对高维删失数据利用 SELO 惩罚函数考虑了基于 Cox 比例风险模型框架下的变量选择及参数估计问题. 在允许协变量维数发散的条件下, 本文给出 SELO 惩罚估计量的相合性以及 oracle 性质. 计算方面若采用传统方法计算惩罚估计解, 当协变量维数较高时计算 Hesse 阵的逆矩阵需要花费大量的时间, 且 SELO 惩罚函数在原点的不光滑性也给计算 SELO 惩罚估计带来很大难度. 为此, 本文利用光滑化技术对 SELO 惩罚函数进行近似, 并利用 DFP 公式去代替 Hesse 阵的逆矩阵, 进而提出了 MSQN 算法. 模拟计算的结果表明, SELO 惩罚方法比已有常用的惩罚方法表现更好, 而且本文提出的新算法与常用的坐标下降算法相比表现更优. 在真实数据部分, 本文还分析了乳腺癌数据, 并利用留一交叉验证法来评估预测的好坏.

关键词 坐标下降算法 惩罚偏似然 拟 Newton 算法 变量选择

MSC (2010) 主题分类 62N02

1 引言

在许多实际问题中, 研究者往往感兴趣于从众多可能的协变量中挑选出对研究的客观现象有重要影响的因素, 这也就是变量选择问题. 基于惩罚函数的变量选择方法可以同时实现变量选择和参数估计的双重目的, 因此成为目前实现变量选择目的的常用方法. 基于 L_0 惩罚函数的惩罚方法是其中

英文引用格式: Cao Y X, Jiao Y L, Shi Y Y, et al. Penalized variable selection procedure for Cox proportional hazards model via seamless-L_0 penalty (in Chinese). Sci Sin Math, 2018, 48: 643–660, doi: 10.1360/SCM-2016-0609

© 2018《中国科学》杂志社

非常重要的一类, 如 C_p [1] 和 AIC (Akaike information criterion) [2] 等. L_0 惩罚方法直接对回归系数中非零元的个数做惩罚, 因此, 直观上利用 L_0 惩罚方法吻合变量选择的目的. 但由于该惩罚函数的不连续性, 直接利用 L_0 惩罚方法进行变量选择难以得到稳定的结果. 因此, 一些学者考虑利用一些特殊的连续函数去代替 L_0 惩罚函数来实现变量选择的目的, 如 LASSO (least absolute shrinkage and selection operator) [3]、SCAD (smoothly clipped absolute derivation) [4]、自适应 LASSO (adaptive ALASSO) [5]、MCP (minimax concave penalty) [6] 和 SELO (seamless L_0) [7] 惩罚函数等. 在这些惩罚方法中, SELO 惩罚函数可以很好地逼近 L_0 惩罚函数, 且该函数的连续性保证了 SELO 惩罚估计比 L_0 惩罚估计更稳定. 模拟计算的结果表明 SELO 惩罚方法比一些已有的惩罚方法表现更好, 尤其当信号较弱时这种优势更加明显 [7,8].

在生存分析中, 响应变量可能带有删失, 针对带有删失的高维生存数据, 变量选择显得尤为重要. 在过去的几十年里, 许多基于惩罚函数的变量选择方法都被用来处理分析高维生存数据 [9–16], 但 SELO 惩罚方法还没有被推广应用到高维生存数据上. 由于 SELO 惩罚方法在线性回归模型中的良好表现 [7], 我们将基于以下 4 个方面将基于 SELO 惩罚函数的惩罚方法推广到 Cox 模型 [17]: (1) 允许协变量维数随着样本量的增大而增大. (2) 模拟计算的结果表明 SELO 惩罚方法优于 LASSO、ALASSO、SCAD 和 MCP 惩罚方法. (3) 惩罚偏似然估计不存在显示解的事实给计算惩罚估计值带来很大难度. 针对线性回归模型求得惩罚最小二乘解的算法有很多, 如 Newton 法 [4]、坐标下降算法 (coordinate descent algorithm, CDA) [18] 和最小角回归 [19]. 但在 Cox 模型中, 惩罚估计的解路径不是逐段线性的. 因此, 一些学者提出利用 Taylor 展开的方法将偏似然函数近似成二次函数, 进而利用 CDA 来求解 [12,20]. 在这种近似方法中, 计算 Hesse 阵的逆矩阵不可避免, 这必然会花费大量的运算时间, 尤其当协变量维数很高时, 计算量将被大大增加, 而且此时也很难保证 Hesse 矩阵的非奇异性. 本文提出利用修正的光滑化拟 Newton (modified smoothed quasi-Newton, MSQN) 算法来求得惩罚偏似然估计. 特别地, 我们在 MSQN 算法中利用 Davidon、Fletcher 和 Powell 提出的 DFP (Daviidon-Fletcher-Powell) 公式 [21] 来求得 Hesse 阵的逆矩阵的近似, 从而减少运算量, 节约运算时间. 此外, 我们还引入向后线性搜索技术 [21] 来保证算法的收敛性. 在运用该算法进行求解时, 我们利用光滑化的方法来解决 SELO 惩罚函数在原点的不光滑问题. (4) 模拟计算结果表明我们提出的 MSQN 算法无论在计算速度上还是计算精度上都要优于 CDA 方法.

本文结构如下: 第 2 节主要介绍带有 SELO 惩罚函数的惩罚偏似然方法, 并给出惩罚估计量的大样本理论性质. 第 3 节详细描述我们提出的 MSQN 算法. 模拟计算以及真实数据分析在第 4 节给出. 第 5 节给出总结说明以及下一步可以进行的工作. 理论结果的证明放在附录 A 中.

2 Cox 比例风险模型中的变量选择

2.1 Cox 比例风险模型和 SELO 惩罚函数

令 T 和 C 分别表示个体的失效时间和删失时间, 从而个体的观测时间为 $X = \min\{T, C\}$. 令 $\delta = I\{T \leqslant C\}$ 表示删失指标, 其中 $I\{\cdot\}$ 是示性函数. $Z(t)$ 为可能与时间有关的协变量. 在给定协变量的条件下假定 T 和 C 条件独立. $\{(X_i, \delta_i, Z_i(t)), i = 1, \ldots, n\}$ 是从总体 $\{X, \delta, Z(t)\}$ 中抽取的容量为 n 的样本. 用 d_n 表示协变量 $Z(t)$ 的维数, 当 $n \to \infty$ 时, 允许 $d_n \to \infty$. 假定个体 i 的失效时间 T_i 的危险率函数服从如下的 Cox 比例风险模型:

$$\lambda(t \mid Z_i(s), s \leqslant t) = \lambda_0(t) \exp\{\boldsymbol{\beta}^{\mathrm{T}} Z_i(t)\}, \tag{2.1}$$

其中 $\lambda_0(t)$ 为未知的基准危险率函数, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{d_n})^{\mathrm{T}}$ 是 d_n 维未知的回归系数. 相应的 log 偏似然函数为

$$\ell_n(\boldsymbol{\beta})=\sum_{i=1}^{n}\delta_i\left\{\boldsymbol{\beta}^{\mathrm{T}}Z_i(X_i)-\log\left[\sum_{j=1}^{n}Y_j(X_i)\exp\{\boldsymbol{\beta}^{\mathrm{T}}Z_j(X_i)\}\right]\right\}.$$

考虑如下的惩罚偏似然函数:

$$Q_n(\boldsymbol{\beta})=\ell_n(\boldsymbol{\beta})-n\sum_{r=1}^{d_n}P(\beta_r;\lambda), \tag{2.2}$$

其中 $P(\beta_r;\lambda)$ 是回归系数 β_r 的惩罚函数, λ ($\lambda\geqslant 0$) 为调节 (tuning) 参数.

理论上 L_0 惩罚函数 $P_0(\boldsymbol{\beta};\lambda)=\lambda I\{|\boldsymbol{\beta}|\neq 0\}$ 具有很好的性质 [22]. 但 L_0 惩罚函数的不连续性导致了 L_0 惩罚估计在计算上的困难. 因此一些学者考虑利用一些连续型函数代替 L_0 惩罚函数来实现变量选择的目的, 例如,

(E1) LASSO 惩罚函数 [3]:

$$P_{\mathrm{LASSO}}(\beta;\lambda)=\lambda|\beta|.$$

(E2) Fan 和 Li 在文献 [4] 中提出的 SCAD 惩罚函数, 其导函数具有如下形式:

$$P'_{\mathrm{SCAD}}(|\beta|;\lambda)=\lambda\left\{I\{|\beta|\leqslant\lambda\}+\frac{(a\lambda-|\beta|)_+}{(a-1)\lambda}I\{|\beta|>\lambda\}\right\},\quad a\geqslant 2,$$

其中 $(t)_+=tI\{t\geqslant 0\}$, $a>0$ 为调节参数, 一般地可取 $a=3.7$.

(E3) Adaptive LASSO 惩罚函数 (简记为 ALASSO)

$$P_{\mathrm{ALASSO}}(\beta;\lambda,w)=\lambda w|\beta|,$$

其中 w 是依赖于数据的权重 [5].

(E4) Zhang [6] 提出了 MCP 惩罚函数

$$P_{\mathrm{MCP}}(\beta;\lambda,\upsilon)=\lambda\int_0^{|\beta|}\left\{1-\frac{x}{\lambda\upsilon}\right\}_+dx,$$

其中调节参数 υ ($\upsilon>1$) 用来控制惩罚函数的凸性.

(E5) SELO 惩罚函数

$$P_{\mathrm{SELO}}(\beta;\lambda)=\frac{\lambda}{\log(2)}\log\left(\frac{|\beta|}{|\beta|+\gamma}+1\right),$$

其中 $\gamma>0$ 是除 λ 之外的调节参数, 通常将 γ 取值为 0.01 [7].

本文采用 SELO 惩罚函数来实现 Cox 比例风险模型框架下的变量选择的目的, 相应地, 目标函数为

$$Q_{ns}(\boldsymbol{\beta})=\ell_n(\boldsymbol{\beta})-n\sum_{r=1}^{d_n}P_{\mathrm{SELO}}(\beta_r;\lambda,\gamma).$$

SELO 惩罚估计量 $\widehat{\boldsymbol{\beta}}$ 是目标函数 $Q_{ns}(\boldsymbol{\beta})$ 的极大值点.

2.2 渐近理论性质

与其他常见的惩罚估计量类似, 在适当的条件下, SELO 惩罚估计量 $\widehat{\boldsymbol{\beta}}$ 具有相合性和 oracle 性质. 本节只列出 $\widehat{\boldsymbol{\beta}}$ 的大样本理论结果, 理论性质所需条件和相关证明放在附录 A 中.

令 $\boldsymbol{\beta}_0 = (\beta_{10}, \ldots, \beta_{d_n 0})^{\mathrm{T}}$ 表示未知的真实回归系数, $\mathcal{A} = \{\beta_{j0} : \beta_{j0} \neq 0\}$ 代表真实活跃集, $\widehat{\mathcal{A}}$ 为估计的活跃集. s_n 表示集合 $\mathcal{A}$ 的 L_0 范数 $s_n = \|\mathcal{A}\|_0$. 对于向量 a, $\|a\|$ 表示 a 的 Euclid 范数.

定理 1 假定附录中条件 (A1)–(A7) 成立, 则存在 $Q_{ns}(\boldsymbol{\beta})$ 的局部极大值点 $\widehat{\boldsymbol{\beta}}$ 满足

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\left(\sqrt{\frac{d_n}{n}}\right).$$

定理 1 表明在适当的条件下, 存在 $\sqrt{n/d_n}$ 相合的 SELO 惩罚估计量. 不失一般性, 令 $\boldsymbol{\beta}_{10}$ 和 $\boldsymbol{\beta}_{20}$ 分别表示 $\boldsymbol{\beta}_0$ 的非零和零的部分. 相应地, 我们将 SELO 惩罚估计 $\widehat{\boldsymbol{\beta}}$ 写成

$$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}_2^{\mathrm{T}})^{\mathrm{T}}.$$

定理 2 假定附录中的条件 (A1)–(A7) 成立, 则有

(i) $\lim_{n\to\infty} \Pr(\widehat{\boldsymbol{\beta}}_2 = \mathbf{0}) = 1$;

(ii) 对于任意的满足 $\|c_n\| = 1$ 的 $s_n \times 1$ 维的常数向量 c_n, 有

$$\sqrt{n} c_n^{\mathrm{T}} \Gamma_{11}^{-1/2} A_{11} (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \to_d N(0, 1),$$

其中 A_{11} 和 Γ_{11} 分别是 $A(\boldsymbol{\beta}_0)$ 和 $\Gamma(\boldsymbol{\beta})$ 的前 s_n 行前 s_n 列元素组成的子阵. $A(\boldsymbol{\beta})$ 和 $\Gamma(\boldsymbol{\beta})$ 的定义见附录 A.

2.3 协方差估计

定义 $\frac{0}{0} = 0$, 为了克服 SELO 惩罚函数的不连续性, 当 $\beta_r \approx \beta_{r0}$ 时, 采用如下的二次逼近:

$$P_{\mathrm{SELO}}(\beta_r; \lambda, \gamma) \approx P_{\mathrm{SELO}}(\beta_{r0}; \lambda, \gamma) + \frac{1}{2|\beta_{r0}|} P'_{\mathrm{SELO}}(|\beta_{r0}|; \lambda, \gamma)(\beta_r^2 - \boldsymbol{\beta}_{r0}^2).$$

令 $\widehat{s}_n = \|\widehat{\mathcal{A}}\|_0$ 代表 SELO 惩罚估计中非零参数的个数, 相应地, 对 Hesse 矩阵做如下分解:

$$H = -\nabla^2 \ell_n(\widehat{\boldsymbol{\beta}}) = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

其中 H_{11} 是矩阵 H 的前 $\widehat{s}_n$ 行前 $\widehat{s}_n$ 列元素组成的子矩阵. 定义

$$\Sigma_n(\boldsymbol{\beta}; \lambda, \gamma) = \mathrm{diag}\left\{ \frac{P'_{\mathrm{SELO}}(|\beta_1|; \lambda, \gamma)}{|\beta_1|}, \ldots, \frac{P'_{\mathrm{SELO}}(|\beta_{d_n}|; \lambda, \gamma)}{|\beta_{d_n}|} \right\}.$$

类似地, 令 $\Sigma_n^{(1)}(\widehat{\boldsymbol{\beta}}; \lambda, \gamma)$ 表示 $\Sigma_n(\widehat{\boldsymbol{\beta}}; \lambda, \gamma)$ 的前 $\widehat{s}_n$ 行前 $\widehat{s}_n$ 列元素组成的子矩阵. 记

$$H_{22.1} = H_{22} - H_{21} H_{11}^{-1} H_{12},$$
$$\widetilde{H}_{11} = H_{11} + \Sigma_n^{(1)}(\widehat{\boldsymbol{\beta}}; \lambda, \gamma).$$

类似于 Zhang 和 Lu [12] 的估计方法, $\widehat{\boldsymbol{\beta}}_1$ 的协方差阵可用如下方法估计:

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}_1) = H_{11}^{-1} + (\widetilde{H}_{11}^{-1} - H_{11}^{-1})H_{12}H_{22.1}H_{21}(\widetilde{H}_{11}^{-1} - H_{11}^{-1}). \tag{2.3}$$

当调节参数 λ 很小时, Zhang 和 Lu [12] 指出 H_{11}^{-1} 可以很好地估计 $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}_1)$. 在我们的数值模拟实验中, 调节参数 λ 和 γ 的值很小, 因此, H_{11}^{-1} 可以作为 $\widehat{\boldsymbol{\beta}}_1$ 的协方差矩阵的一个估计, 即

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}_1) \approx H_{11}^{-1}. \tag{2.4}$$

3 算法设计

令 ϵ_0 代表一个固定的很小的正数, 用 $\sqrt{\beta_r^2 + \epsilon_0^2}$ 去近似 $|\beta_r|$. 显然, 当 $\epsilon_0 \to 0$ 时,

$$\sqrt{\beta_r^2 + \epsilon_0^2} \to |\beta_r|.$$

但 ϵ_0 越小导致计算过程越不稳定, 因此, 选取合适的 ϵ_0 的值非常重要. 我们的模拟计算表明, 当 ϵ_0 在集合 $\{0.001, 0.01\}$ 中取值时会得到稳定的结果. 从而我们将得到如下的伪目标函数:

$$D_n(\boldsymbol{\beta}) = \ell_n(\boldsymbol{\beta}) - n\sum_{r=1}^{d_n} P_{\mathrm{SELO}}(\sqrt{\beta_r^2 + \epsilon_0^2}; \lambda).$$

伪目标函数 $D_n(\boldsymbol{\beta})$ 关于 β 是光滑的, 因此, 我们可以利用 Newton 算法得到 $D_n(\boldsymbol{\beta})$ 的最优解. 为了避免 Hesse 阵的逆矩阵的计算, 我们用拟 Newton 算法对 $D_n(\boldsymbol{\beta})$ 进行求解. 拟 Newton 算法的第 $k+1$ 步迭代公式为

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \delta^{(k)} h^{(k)},$$

其中 $h^{(k)}$ 和 $\delta^{(k)}$ 分别代表第 k 步迭代的拟 Newton 方向和最优步长因子. 对于不同的 k, 最优步长因子可能不同. 在 Newton 算法中, $h^{(k)} = -[\nabla^2 D_n(\boldsymbol{\beta}^{(k)})]^{-1}\nabla D_n(\boldsymbol{\beta}^{(k)})$. 为了避免 $-[\nabla^2 D_n(\boldsymbol{\beta}^{(k)})]^{-1}$ 的计算在拟 Newton 算法中, 我们用满足如下条件的 $H^{(k)}$ 去代替 $-[\nabla^2 D_n(\boldsymbol{\beta}^{(k)})]^{-1}$:

(R1) $\{H^{(k)}\}_{k\geqslant 1}$ 计算方便;

(R2) $\{H^{(k)}\}_{k\geqslant 1}$ 正定;

(R3) $H^{(k)}$ 应保证拟 Newton 方向 $h^{(k)} = H^{(k)}\nabla D_n(\boldsymbol{\beta}^{(k)})$ 为目标函数的上升方向.

令

$$\begin{aligned} p^{(k)} &= \boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}, \\ q^{(k)} &= [\nabla D_n(\boldsymbol{\beta}^{(k+1)})] - [\nabla D_n(\boldsymbol{\beta}^{(k)})]. \end{aligned}$$

采用如下的由 Davidon、Fletcher 和 Powell 提出并发展的 DFP 公式更新 $H^{(k)}$ [21]:

$$H^{(k+1)} = H^{(k)} + \frac{p^{(k)}{p^{(k)}}^{\mathrm{T}}}{{p^{(k)}}^{\mathrm{T}}q^{(k)}} - \frac{H^{(k)}q^{(k)}{q^{(k)}}^{\mathrm{T}}H^{(k)}}{{q^{(k)}}^{\mathrm{T}}H^{(k)}q^{(k)}}. \tag{3.1}$$

一般地, 取 $H^{(1)} = I_{d_n}$, 其中 I_{d_n} 为 $d_n \times d_n$ 维单位矩阵. 易证当 ${p^{(k)}}^{\mathrm{T}}q^{(k)} > 0$ 时矩阵 $H^{(k+1)}$ 正定.

记 ϱ 为事先给定的阈值 (在模拟计算中, 令 $\varrho = 0.01$). 给定 λ 和 γ, 令 K 表示最大的迭代步数, 我们将如下算法命名为光滑化拟 Newton (smoothed quasi-Newton, SQN) 算法.

算法 1 SQN 算法

步骤 (1) 令 $k=1$, 记 $\boldsymbol{\beta}$ 的初始值为 $\boldsymbol{\beta}^{(1)}$.

步骤 (2) 若 $\|\nabla D_n(\boldsymbol{\beta}^{(k)})\| \leqslant \varrho$, 输出 $\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}^{(k)}$; 否则转入步骤 (3).

步骤 (3) (i) 更新拟 Newton 方向 $h^{(k)}=H^{(k)}\nabla D_n(\boldsymbol{\beta}^{(k)})$, 寻找满足如下条件的最优步长 $\delta^{(k)}$:

$$\delta^{(k)}=\arg\max_{\delta\geqslant 0} D_n(\boldsymbol{\beta}^{(k)}+\delta h^{(k)}). \tag{3.2}$$

令 $\boldsymbol{\beta}^{(k+1)}=\boldsymbol{\beta}^{(k)}+\delta^{(k)}h^{(k)}$.

(ii) 若 ${p^{(k)}}^{\mathrm{T}}q^{(k)}\geqslant 0$, 用 (3.1) 更新 $H^{(k)}$, 否则 $H^{(k+1)}=H^{(k)}$. 令 $k=k+1$, 返回步骤 (2).

步骤 (4) 重复步骤 (2)–(3) 直到 $k=K+1$ 或者算法收敛.

在算法 1 中, 涉及最优步长 $\delta^{(k)}$ 的求解. 一般来说求得最优步长的计算量很大, 为了进一步精简算法, 我们引入向后线性搜索 (backtracking line search, BLS) 技术来求得最优步长 $\delta^{(k)}$. BLS 技术是线性搜索中的常用手段, 关于 BLS 的详细介绍可以参见专著 [21].

令 $\eta\in(0,\frac{1}{2})$, ρ $(0<\rho<1)$ 为事先给定的常数 (模拟计算中, 取 $\eta=0.4$, $\rho=0.55$). 给定 $\boldsymbol{\beta}$ 和 h, 如下的 BLS 方法被用来在每一步迭代中寻找最优步长:

(B1) 令 $s=1$, $\delta_{(s)}=1$.

(B2) 若下面不等式 $D_n(\boldsymbol{\beta}+\delta_{(s)}h)\geqslant D_n(\boldsymbol{\beta})+\eta\delta_{(s)}[\nabla D_n(\boldsymbol{\beta})]^{\mathrm{T}}h$ 成立, 则令最优步长 $\delta=\delta_{(s)}$; 否则进入步骤 (B3).

(B3) 令 $\delta_{(s+1)}=\delta_{(s)}\rho$, $s=s+1$, 返回步骤 (B2).

所谓的修正的光滑化拟 Newton (modified SQN, MSQN) 算法是在算法 1 的基础上用 BLS 算法代替 (3.2) 来寻求最优步长. 由 $\delta^{(k)}$ 的定义, 我们可以证明 MSQN 算法的上升性质:

$$\begin{aligned} D_n(\boldsymbol{\beta}^{(k)}+\delta^{(k)}h^{(k)}) &\geqslant D_n(\boldsymbol{\beta}^{(k)})+\eta\delta^{(k)}[\nabla D_n(\boldsymbol{\beta}^{(k)})]^{\mathrm{T}}h^{(k)} \\ &= D_n(\boldsymbol{\beta}^{(k)})+\eta\delta^{(k)}[\nabla D_n(\boldsymbol{\beta}^{(k)})]^{\mathrm{T}}H^{(k)}[\nabla D_n(\boldsymbol{\beta}^{(k)})] \\ &\geqslant D_n(\boldsymbol{\beta}^{(k)}). \end{aligned}$$

$H^{(k)}$ 的正定性保证了二次型 $[\nabla D_n(\boldsymbol{\beta}^{(k)})]^{\mathrm{T}}H^{(k)}[\nabla D_n(\boldsymbol{\beta}^{(k)})]>0$, 从而最后一个不等式成立.

4 数值模拟和实证分析

4.1 数值模拟

我们通过模拟计算来检验 SELO 惩罚估计量 $\widehat{\boldsymbol{\beta}}$ 在有限样本下的表现. 个体的失效时间由 Cox 比例风险模型 $\lambda(t\mid\boldsymbol{Z})=\lambda_0(t)\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z})$, 其中 $\lambda_0(t)=1$. $\boldsymbol{\beta}_{10}^{\mathrm{T}}$ 和 $\boldsymbol{\beta}_{20}^{\mathrm{T}}$ 分别表示回归系数 $\boldsymbol{\beta}_0$ 的非零以及零的部分. 对于 $\boldsymbol{\beta}_0$ 的取值, 考虑如下两种情形:

模型一 (信号较弱的情形): $\boldsymbol{\beta}_0=(\boldsymbol{\beta}_{10}^{\mathrm{T}},\boldsymbol{\beta}_{20}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\boldsymbol{\beta}_{10}=(0.5,1,-0.6)^{\mathrm{T}}$.

模型二 (一般情形): 与 Bradic 等 [20] 的设置一样, 令 $\boldsymbol{\beta}_0$ 的前 4 个元素全部为 1, 其他元素全部为零.

d_n 维协变量 $\boldsymbol{Z}$ 服从均值为 $\mathbf{0}$ 协方差阵为 Σ 的正态分布, 其中 $\Sigma=(\sigma_{ij})$, $\sigma_{ij}=0.5^{|i-j|}$, $i,j=1,\ldots,d_n$. 删失时间服从 $(0,c)$ 上的均匀分布, 其中 c 的取值使得删失率为 25% 和 50%.

我们将 SELO 惩罚估计与基于 LASSO、SCAD、ALASSO 和 MCP 惩罚函数的惩罚估计进行比较. 令 SCAD 惩罚中的参数 $a=3.7$, MCP 惩罚函数中的调节参数 $\upsilon=2.7$. 根据文献 [7], 我们将 SELO

惩罚函数中的调节参数 γ 固定为 0.01. 所有惩罚估计中涉及的调节参数 λ 均由 BIC 准则确定.

对于上面提到的惩罚估计, 考虑如下指标: (M1) 估计的模型大小: $\|\widehat{\mathcal{A}}\|_0 = |\{r : \hat{\boldsymbol{\beta}}_r \neq 0\}|$. (M2) 是否选出正确模型的指标: $I\{\widehat{\mathcal{A}} = \mathcal{A}\}$. (M3) 模型误差 (简记为 ME): $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\Sigma(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. (M4) 多选指标: $N_+ = \|\widehat{\mathcal{A}}\backslash\mathcal{A}\|_0$. (M5) 少选指标: $N_- = \|\mathcal{A}\backslash\widehat{\mathcal{A}}\|_0$. 样本量 n 分别设为 200 或者 400, 协变量维数 d_n 为 10、20 和 40. 对于每种情形, 我们基于 1,000 次独立重复计算指标 (M1) 到 (M5) 的平均值.

表 1 和 2 给出了模型一的计算结果. 由表可知, 无论参数如何设置, 基于 SELO 惩罚函数的变量选择方法比其他方法可以以更高的频率选取出正确模型. 我们还可以发现模拟结果的好坏依赖于数据的删失率和协变量的维数. 例如, 当删失率为 25% 样本量 $n = 200$ 时, MCP 和 SELO 方法表现相当, 且当协变量维数 $d_n = 10$ 时, 这两种方法比其他三种惩罚方法表现更好. 当 $d_n = 20$ 或 $d_n = 40$ 时, 在 1,000 次模拟计算中, SELO 惩罚方法选出正确模型的频率可以高达 90% 以上, 明显优于其他 4 种方法. 当样本量 n 增加到 400 时, 本文考虑的 5 种变量选择方法表现相当. 当删失率达到 50%、协变量维数 $d_n = 20$ 或者 $d_n = 40$ 时, 所有的变量选择方法在是否选出正确模型的指标上表现均有失误, 但当我们将样本量增加到 $n = 400$ 时, SELO 惩罚方法选择出正确模型的频率可以达到 97.7%.

表 3 和 4 给出了模型二的变量选择结果. 在模型二中, 非零部分回归系数的真实值均高于模型一中的值. 5 种惩罚方法在少选指标 N_- 上均表现优良. 在所有情形下 ALASSO 在指标 Corr%、N_+、ME 和估计的模型大小上表现最差. 除了当 $d_n = 40$ 和 $n = 200$ 的情形外, SELO 惩罚方法在选择出正确模型的指标 Corr% 以及多选指标 N_+ 上表现最优, 且 SELO 惩罚方法在估计模型大小的指标上的误差是最小的. 针对样本量 $n = 400$ 的情形, SELO 在上述三个指标上表现最好. 当考虑指标模型误差 ME 时, SCAD、MCP 和 SELO 方法表现相当, 这三种方法均优于 LASSO 和 ALASSO. 且当样本量增加到 400 时, 通过比较表 1 与 3, 我们发现当信号较弱时, SELO 惩罚方法比其他 4 种方法表现更好.

表 1 模型大小 $\|\mathcal{A}\|_0 = 3$、$n = 200$ 下模型一 (信号较弱情形) 的模拟结果

		删失率 = 25%					删失率 = 50%				
d_n	方法	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$
10	LASSO	79.10	0.196	0.211	0.013	3.198	66.90	0.316	0.188	0.190	2.998
	ALASSO	89.70	0.047	0.099	0.017	3.082	70.20	0.093	0.347	0.044	3.303
	SCAD	88.30	0.048	0.038	0.089	2.949	46.90	0.198	0.079	0.668	2.411
	MCP	90.00	0.035	0.111	0.005	3.106	84.10	0.061	0.133	0.050	3.083
	SELO	97.30	0.033	0.015	0.013	3.002	89.80	0.057	0.056	0.054	3.002
20	LASSO	62.30	0.196	0.434	0.013	3.421	54.30	0.316	0.379	0.198	3.181
	ALASSO	64.90	0.097	0.470	0.017	3.453	32.80	0.209	1.314	0.029	4.285
	SCAD	78.80	0.061	0.179	0.087	3.092	40.90	0.215	0.294	0.617	2.677
	MCP	79.70	0.041	0.248	0.005	3.243	73.30	0.076	0.289	0.064	3.225
	SELO	95.20	0.036	0.042	0.014	3.028	82.90	0.066	0.147	0.053	3.094
40	LASSO	40.20	0.198	0.862	0.017	3.845	37.00	0.321	0.729	0.226	3.503
	ALASSO	13.40	0.348	2.463	0.012	5.451	1.40	0.763	5.405	0.031	8.374
	SCAD	72.50	0.073	0.343	0.073	3.270	36.60	0.239	0.602	0.576	3.026
	MCP	67.50	0.060	0.528	0.009	3.519	58.30	0.109	0.630	0.086	3.544
	SELO	90.00	0.046	0.109	0.013	3.096	69.30	0.093	0.366	0.060	3.306

Corr%: 选择正确模型的比例; N_+、N_- 和 ME 代表基于 1,000 次模拟计算的多选、少选指标和模型误差的平均值.

表 2　模型大小 $\|\mathcal{A}\|_0 = 3$、$n = 400$ 下模型一 (信号较弱情形) 的模拟结果

		删失率 = 25%					删失率 = 50%				
d_n	方法	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$
10	LASSO	94.90	0.186	0.051	0.000	3.051	87.80	0.305	0.040	0.083	2.957
	ALASSO	99.90	0.017	0.000	0.001	2.999	96.90	0.027	0.027	0.005	3.022
	SCAD	97.70	0.018	0.001	0.022	2.979	65.00	0.122	0.011	0.449	2.562
	MCP	96.30	0.015	0.038	0.000	3.038	96.60	0.022	0.035	0.000	3.035
	SELO	100.00	0.015	0.000	0.000	3.000	99.30	0.022	0.002	0.005	2.997
20	LASSO	93.50	0.184	0.064	0.001	3.063	86.00	0.304	0.059	0.085	2.974
	ALASSO	98.80	0.020	0.011	0.001	3.010	89.60	0.040	0.116	0.003	3.113
	SCAD	97.10	0.020	0.006	0.023	2.983	59.60	0.130	0.075	0.458	2.617
	MCP	91.20	0.016	0.096	0.000	3.096	90.80	0.025	0.105	0.001	3.104
	SELO	100.00	0.015	0.000	0.000	3.000	98.20	0.023	0.014	0.005	3.009
40	LASSO	91.00	0.183	0.090	0.000	3.090	83.60	0.303	0.082	0.091	2.991
	ALASSO	93.30	0.036	0.073	0.000	3.073	64.10	0.091	0.476	0.002	3.474
	SCAD	93.70	0.025	0.039	0.030	3.009	56.50	0.148	0.258	0.388	2.870
	MCP	81.10	0.021	0.235	0.000	3.235	82.80	0.031	0.236	0.002	3.234
	SELO	99.80	0.016	0.002	0.000	3.002	97.70	0.024	0.020	0.003	3.017

Corr%: 选择正确模型的比例; N_+、N_- 和 ME 代表基于 1,000 次模拟计算的多选、少选指标以及模型误差的平均值.

表 3　模型大小 $\|\mathcal{A}\|_0 = 4$、$n = 200$ 下模型二 (一般情形) 的模拟结果

		删失率 = 25%					删失率 = 50%				
d_n	方法	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$
10	LASSO	91.90	0.112	0.089	0.000	4.089	86.50	0.167	0.155	0.000	4.155
	ALASSO	90.20	0.100	0.114	0.000	4.114	70.90	0.181	0.386	0.000	4.386
	SCAD	91.60	0.074	0.090	0.000	4.090	89.30	0.114	0.120	0.000	4.120
	MCP	92.10	0.073	0.089	0.000	4.089	92.20	0.110	0.090	0.000	4.090
	SELO	98.10	0.070	0.022	0.000	4.022	93.40	0.107	0.074	0.000	4.074
20	LASSO	89.00	0.152	0.133	0.000	4.133	83.00	0.247	0.230	0.000	4.230
	ALASSO	63.60	0.190	0.499	0.000	4.499	27.70	0.415	1.516	0.000	5.516
	SCAD	82.70	0.092	0.226	0.000	4.226	80.90	0.139	0.282	0.000	4.282
	MCP	83.90	0.085	0.220	0.000	4.220	86.10	0.130	0.206	0.000	4.206
	SELO	94.60	0.080	0.066	0.000	4.066	87.50	0.124	0.160	0.000	4.160
40	LASSO	91.20	0.193	0.111	0.000	4.111	86.80	0.328	0.175	0.000	4.175
	ALASSO	13.00	0.679	2.613	0.000	6.613	0.70	1.614	6.443	0.000	10.443
	SCAD	70.40	0.123	0.469	0.000	4.469	63.40	0.203	0.678	0.002	4.676
	MCP	76.30	0.108	0.397	0.000	4.397	79.40	0.166	0.372	0.000	4.372
	SELO	88.50	0.108	0.140	0.000	4.140	72.80	0.177	0.382	0.000	4.382

Corr%: 选择正确模型的比例; ME: 模型误差的平均值; N_+ 和 N_- 分别代表多选和少选指标的均值; $\|\widehat{\mathcal{A}}\|_0$ 为估计的模型的平均大小.

针对模型一, 表 5 给出了删失率为 25% 时基于 1,000 次模拟计算的 SELO 惩罚估计的均值及相应的方差估计的平均值 (简记为 ESE) 和经验方差 (简记为 SE). 可知回归系数的估计值是渐近无偏的, ESE 和 SE 吻合得很好.

表 4 模型大小 $\|\mathcal{A}\|_0 = 4$、$n = 400$ 下模型二 (一般情形) 的模拟结果

		删失率 = 25%					删失率 = 50%				
d_n	方法	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$
10	LASSO	99.50	0.060	0.005	0.000	4.005	98.80	0.095	0.012	0.000	4.012
	ALASSO	99.80	0.039	0.002	0.000	4.002	97.60	0.061	0.025	0.000	4.025
	SCAD	97.80	0.033	0.025	0.000	4.025	96.50	0.047	0.039	0.000	4.039
	MCP	93.20	0.034	0.077	0.000	4.077	95.00	0.047	0.055	0.000	4.055
	SELO	99.80	0.032	0.002	0.000	4.002	99.30	0.044	0.007	0.000	4.007
20	LASSO	98.70	0.060	0.014	0.000	4.014	95.90	0.094	0.042	0.000	4.042
	ALASSO	99.00	0.051	0.010	0.000	4.010	86.40	0.098	0.155	0.000	4.155
	SCAD	94.10	0.036	0.068	0.000	4.068	89.50	0.056	0.132	0.000	4.132
	MCP	85.90	0.038	0.173	0.000	4.173	88.00	0.055	0.159	0.000	4.159
	SELO	99.70	0.033	0.003	0.000	4.003	97.60	0.048	0.026	0.000	4.026
40	LASSO	97.10	0.078	0.032	0.000	4.032	93.90	0.140	0.063	0.000	4.063
	ALASSO	91.70	0.104	0.093	0.000	4.093	55.80	0.248	0.659	0.000	4.659
	SCAD	89.30	0.045	0.131	0.000	4.131	79.20	0.078	0.319	0.000	4.319
	MCP	78.80	0.046	0.316	0.000	4.316	81.60	0.067	0.293	0.000	4.293
	SELO	98.80	0.040	0.012	0.000	4.012	96.10	0.057	0.044	0.000	4.044

Corr%: 选择正确模型的比例; ME: 模型误差的平均值; N_+ 和 N_- 分别代表多选和少选指标的均值; $\|\widehat{\mathcal{A}}\|_0$ 为估计的模型的平均大小.

表 5 模型一中基于 1,000 次独立模拟计算的 SELO 惩罚估计的均值以及相应的 SE 和 ESE

			$\beta_1 = 0.5$			$\beta_2 = 1$			$\beta_3 = -0.6$		
n	d_n	删失率	Mean	ESE	SE	Mean	ESE	SE	Mean	ESE	SE
200	10	25%	0.4970	0.0124	0.0134	1.0113	0.0173	0.0176	−0.5971	0.0129	0.0139
		50%	0.4891	0.0089	0.0093	0.9992	0.0143	0.0124	−0.5876	0.0100	0.0091
	20	25%	0.5006	0.0124	0.0139	1.0206	0.0173	0.0182	−0.6013	0.0129	0.0145
		50%	0.4947	0.0186	0.0252	1.0213	0.0232	0.0272	−0.5918	0.0195	0.0242
	40	25%	0.5090	0.0125	0.0149	1.0420	0.0174	0.0201	−0.6125	0.0129	0.0154
		50%	0.5010	0.0187	0.0284	1.0430	0.0233	0.0307	−0.6031	0.0195	0.0260
400	10	25%	0.4972	0.0061	0.0057	1.0017	0.0093	0.0084	−0.5933	0.0062	0.0061
		50%	0.4891	0.0089	0.0093	0.9992	0.0143	0.0124	−0.5876	0.0100	0.0091
	20	25%	0.5010	0.0061	0.0059	1.0076	0.0093	0.0084	−0.5970	0.0062	0.0061
		50%	0.4955	0.0089	0.0095	1.0068	0.0143	0.0126	−0.5934	0.0100	0.0092
	40	25%	0.5053	0.0061	0.0062	1.0173	0.0093	0.0090	−0.6017	0.0062	0.0065
		50%	0.5003	0.0089	0.0093	1.0176	0.0143	0.0130	−0.5987	0.0100	0.0095

Mean: 惩罚估计的均值; SE: 惩罚估计的经验方差; ESE: 估计方差的均值.

表 6　真实的模型大小为 3、删失率为 25% 下模型一的模拟结果

n	d_n	方法	time (s)	Corr%	ME	N_+	N_-	$\|\widehat{\mathcal{A}}\|_0$
100	20	MSQN	0.3266	54.00	0.1613	0.350	0.220	3.13
		CDA	0.8873	43.00	0.2416	0.580	0.240	3.34
	50	MSQN	0.4563	43.00	0.3509	0.790	0.350	3.44
		CDA	2.7710	28.00	0.3289	1.330	0.280	4.05
	80	MSQN	0.5642	36.00	0.5570	0.980	0.410	3.57
		CDA	5.1649	15.00	0.3957	2.100	0.290	4.81
	100	MSQN	0.6513	28.00	0.7773	1.310	0.470	3.84
		CDA	6.8469	11.00	0.4593	2.730	0.320	5.41
200	20	MSQN	0.6177	97.00	0.0360	0.030	0.010	3.02
		CDA	3.1347	66.00	0.1631	0.420	0.020	3.40
	50	MSQN	0.8344	89.00	0.0529	0.050	0.060	2.99
		CDA	8.2814	41.00	0.1835	1.010	0.030	3.98
	80	MSQN	1.0424	85.00	0.0787	0.130	0.070	3.06
		CDA	13.8166	26.00	0.2105	1.740	0.030	4.71
	100	MSQN	1.2200	80.00	0.0936	0.130	0.100	3.03
		CDA	18.1622	20.00	0.2297	2.150	0.030	5.12
	120	MSQN	1.3247	78.00	0.1157	0.180	0.120	3.06
		CDA	22.6929	19.00	0.2389	2.390	0.040	5.35
	150	MSQN	1.6464	73.00	0.1276	0.210	0.110	3.10
		CDA	29.9111	14.00	0.2628	3.030	0.040	5.35
	180	MSQN	1.9241	70.00	0.1515	0.280	0.120	3.16
		CDA	37.3164	11.00	0.2829	3.410	0.040	6.30
	200	MSQN	2.1147	69.00	0.1740	0.340	0.110	3.23
		CDA	42.6835	11.00	0.2997	3.740	0.040	6.70

time: 100 次模拟计算中求得 40 个 λ 所对应的最优解所花费的平均时间.

为了说明 MSQN 算法的有效性, 我们从运算时间以及指标 (M1) 到 (M5) 6 个方面来比较 MSQN 算法和基于二次逼近的 CDA, 其中模拟数据由模型一生成. 在每一次的模拟计算中, 我们均计算了对应于 40 个 λ 的惩罚解. 表 6 给出了基于 100 次独立模拟计算上述指标的平均值, 其中指标 time 为计算 40 个 λ 的对应惩罚解所需的的平均时间 (单位为秒). 由此表可知, CDA 算法得到惩罚估计所花费的时间远远大于 MSQN 算法所需时间, 在少选指标 N_- 上, CDA 算法优于 MSQN 算法, 但在其其他指标 corr%、N_+ 以及 $|\widehat{\mathcal{A}}|$ 上 MSQN 算法均优于 CDA 方法.

4.2　关于乳腺癌数据的实例分析

我们将本文的工作应用到关于乳腺癌数据 nki70 [23, 24] 的研究中. van't Veer 等 [24] 发现了 70 个可能与乳腺癌有关的基因. 我们的研究目的是从 5 个临床指标和 70 个基因指标中找出乳腺癌的可能致病因素. 我们将研究限制在那些淋巴指标呈阳性的 144 名患者上, 该组患者的追踪时间的中位数为 3.17 年, 取值范围为 0.35 年到 14.01 年. 删失率约为 66%. 可能致病的 5 个临床指标包括肿瘤细胞的大小 z_1 (若细胞大小 $\leqslant$ 2 cm, 则记为 1, 否则记为 2)、受影响的淋巴结数目 z_2 (若数目小于等于 3, 则

表 7 乳腺癌数据的分析结果

协变量	协变量所代表的指标名	LASSO		ALASSO		SCAD		MCP		SELO	
		$\widehat{\boldsymbol{\beta}}$	SE	$\widehat{\boldsymbol{\beta}}$	SE	$\widehat{\boldsymbol{\beta}}$	SE	$\widehat{\boldsymbol{\beta}}$	SE	$\widehat{\boldsymbol{\beta}}$	SE
z_5	Age	−0.0444	0.0053	−0.0589	0.0073	–	–	−0.2262	0.0117	–	–
z_9	NUSAP1	–	–	–	–	0.0164	0.0011	–	–	0.0101	0.0004
z_{12}	QSCN6L1	0.0523	0.0082	–	–	0.0196	0.0016	0.2922	0.0262	0.0123	0.0006
z_{15}	Contig32125-RC	–	–	–	–	–	–	0.1909	0.0145	–	–
z_{38}	ZNF533	−0.0714	0.0067	–	–	−0.0198	0.0015	−0.5122	0.0178	−0.0117	0.0006
z_{46}	DCK	–	–	0.0513	0.0093	–	–	–	–	–	–
z_{54}	ORC6L	–	–	–	–	–	–	–	–	–	–
z_{64}	PITRM1	–	–	–	–	–	–	–	–	–	–
z_{65}	IGFBP5.1	0.0548	0.0082	0.0666	0.0098	0.0170	0.0016	0.2211	0.0180	0.0097	0.0006
z_{69}	PRC1	0.1374	0.0143	0.1222	0.0171	0.0243	0.0014	0.5610	0.0439	0.0155	0.0006
z_{71}	CENPA	–	–	–	–	–	–	–	–	0.0100	0.0004
PE	Predition error	0.0061	–	0.0340	–	0.0005	–	0.8682	–	0.0002	–

PE: 预测误差的均值; $\widehat{\boldsymbol{\beta}}$: 惩罚估计; SE: 惩罚估计的估计方差;

记为 1, 否则记为 2)、雌激素受体状态 z_3 (阴性记为 1, 阳性为 2)、癌症等级 z_4 (1 代表无显著差异, 3 代表显著差异, 2 表示介于二者之间) 以及患者就诊时的年龄 z_5. 从而, 包含 5 个临床因素和 70 个基因表达 (用 z_6 到 z_{75} 表示) 协变量的 Cox 模型具有如下形式:

$$\lambda(t \mid Z) = \lambda_0(t) \exp\left\{\sum_{i=1}^{75} z_i\beta_i\right\},$$

其中 $\lambda_0(t)$ 为基准危险率函数.

首先对连续型变量做标准化处理使之具有零均值和单位方差. 我们考虑基于 LASSO、ALASSO、SCAD、MCP 和 SELO 惩罚函数的惩罚估计, 并用 MSQN 算法进行求解. 调节参数的确定方法同模拟计算. 我们考虑了 ϵ 在集合 $\{0.01, 0.005, 0.001\}$ 中取值的不同情形, 在所有情形下选择出的显著变量都是相同的, 且 ϵ_0 取值越小得到的惩罚估计越有效. 为此我们在表 7 中列出了当 $\epsilon_0 = 0.001$ 的结果, 此时停机准则为 $\|\nabla Q_n(\beta)\| \leqslant 0.001$. 我们利用留一交叉验证来评估惩罚方法的好坏, 为此考虑 Bradic 等 [20] 提出的预测误差:

$$[\exp\{-\beta^{\diamondsuit\mathrm{T}} Z\} - \exp\{-\hat{\beta}_{(-i)}^{\mathrm{T}} Z\}]^2, \quad i = 1, \ldots, n,$$

其中 $\beta^{\diamondsuit}$ 代表基于全部数据的惩罚估计, $\hat{\beta}_{(-i)}^{\mathrm{T}}$ 为去掉第 i 个样本后的惩罚解. 基于惩罚方法的预测误差的平均值及选择出的致病的临床指标和基因由此表给出. 所有的惩罚方法均选出 z_{65} (PITRM1) 和 z_{69} (PRC1) 为致病基因, 基于 LASSO、SCAD、MCP 和 SELO 惩罚函数的变量选择方法将变量 z_{12} (QSCN6L1) 和 z_{38} (ZNF533) 选出. 在所有方法中 SELO 惩罚方法具有最小的预测误差.

5 小结

本文在 Cox 比例风险模型的框架下研究了基于 SELO 惩罚偏似然方法的变量选择问题. 当样本量 $n \to \infty$ 时, 我们允许协变量维数 $d_n \to \infty$, 并在适当的条件下给出了 SELO 惩罚估计量的相合性以及 oracle 性质. 为了节约运算时间, 我们提出了 MSQN 算法, 该算法避免了 Hesse 阵逆矩阵的计算, 并得到了该算法的部分收敛性质. MSQN 算法亦适用于协变量维数高于样本量的相关计算问题. 模拟计算的结果表明, SELO 惩罚偏似然方法尤其适用于信号较弱的情形, 且本文提出的 MSQN 算法在计算速度和准确度方面优于最常用到的 CDA 算法. 虽然 MSQN 算法在选择重要变量方面表现优良, 但我们并不提倡用 Hesse 阵逆矩阵的近似去估计惩罚估计量的协方差, 因为这样做可能导致损失更多的信息.

关于 MSQN 算法的另一个问题是关于 ϵ_0 的选择. 显然, 当 $\epsilon_0 \to 0$ 时,

$$\sqrt{\beta + \epsilon_0^2} \to |\beta|,$$

但 ϵ_0 越趋于 0, 算法越不稳定. Shi 等 [25] 指出 $\epsilon_0 \in [0.001, 0.01]$ 是比较合适的. 为了研究 ϵ_0 对惩罚估计的影响, 我们做了更多的模拟计算, 模拟计算的结果表明, 当 ϵ_0 在集合 $\{0.1, 0.01, 0.001\}$ 中取值时会得到稳定的结果.

下面的几个问题还需要进一步地讨论研究: (1) 虽然本文提出的 MSQN 算法适用于 $d_n > n$ 的情形, 但缺少相关理论支持; (2) Cox 模型中关于个体危险率函数的假定和实际情形往往不符, 为此, 将本文提出的方法推广到其他常见模型 (如加速时效时间模型) 是我们将要做的工作; (3) 本文假定协变

量维数的阶是 $o(n^{1/4})$, 但在实际问题尤其是一些与基因有关的问题中, 该假定往往是不成立的, 因此, 在以后的工作中亦应考虑协变量维数是样本量的指数阶情形下的变量选择问题.

参考文献

1 Mallows C L. Some comments on C_p. Technometrics, 1973, 15: 661–675
2 Akaike H. A new look at the statistical model identification. IEEE Trans Automat Control, 1974, 19: 716–723
3 Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Ser B Stat Methodol, 1996, 58: 267–288
4 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statist Assoc, 2001, 96: 1348–1360
5 Zou H. The adaptive LASSO and its oracle properties. J Amer Statist Assoc, 2006, 101: 1418–1429
6 Zhang C H. Nearly unbiased variable selection under minimax concave penalty. Ann Statist, 2010, 38: 894–942
7 Dicker L, Huang B, Lin X. Variable selection and estimation with the seamless-L_0 penalty. Statist Sinica, 2013, 23: 929–962
8 Li Z, Wang S, Lin X. Variable selection and estimation in generalized linear models with the seamless L_0 penalty. Canad J Statist, 2012, 40: 745–769
9 Tibshirani R. The LASSO method for variable selection in the cox model. Stat Med, 1997, 16: 385–395
10 Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. Ann Statist, 2002, 30: 74–99
11 Cai J, Fan J, Li R, et al. Variable selection for multivariate failure time data. Biometrika, 2005, 92: 303–316
12 Zhang H H, Lu W. Adaptive LASSO for Cox's proportional hazards model. Biometrika, 2007, 94: 691–703
13 Zou H. A note on path-based variable selection in the penalized proportional hazards model. Biometrika, 2008, 95: 241–247
14 Du P, Ma S, Liang H. Penalized variable selection procedure for Cox models with semiparametric relative risk. Ann Statist, 2010, 38: 2092–2117
15 Long Q, Chung M, Moreno C S, et al. Risk prediction for prostate cancer recurrence through regularized estimation with simultaneous adjustment for nonlinear clinical effects. Ann Appl Stat, 2011, 5: 2003–2023
16 Ma S, Du P. Variable selection in partly linear regression model with diverging dimensions for right censored data. Statist Sinica, 2012, 22: 1003–1020
17 Andersen P K, Gill R D. Cox's regression model for counting processes: A large sample study. Ann Statist, 1982, 10: 1100–1120
18 Wu J F. Penalized regressions: The bridge versus the LASSO. J Comput Graph Statist, 1998, 7: 397–416
19 Efron B, Hastie T, Johnstone I, et al. Least angle regression. Ann Statist, 2004, 32: 407–499
20 Bradic J, Fan J, Jiang J. Regularization for Cox's proportional hazards model with NP-dimensionality. Ann Statist, 2011, 39: 3092–3120
21 Nocedal J, Wright S J. Numerical Optimization, 2nd ed. Springer Series in Operations Research and Financial Engineering. New York: Springer-Verlag, 2006, 30–139
22 Zhang C, Zhang T. A gegeral theorey of concave regularization for high-dimensional sparse estimation problems. Statist Sci, 2012, 27: 576–593
23 van de Vijver M J, He Y D, van't Veer L J, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med, 2002, 347: 1999–2009
24 van't Veer L J, Dai H, van de Vijver M J, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature, 2002, 415: 530–536
25 Shi Y Y, Cao Y X, Jiao Y L, et al. SICA for Cox's proportional hazards model with a diverging number of parameters. Acta Math Appl Sin Engl Ser, 2014, 30: 887–902

附录 A

令 $N_i(t)$ 代表计数过程 $N_i(t) = I\{T_i \leqslant t, \delta_i = 1\}$, $Y_i(t) = I(X_i \geqslant t)$ 代表在 t 时刻个体 i 是否处于

风险. 当 $l=0,1,2$ 时, 定义

$$S^{(l)}(\boldsymbol{\beta};t)=\sum_{i=1}^{n}Y_i(t)Z_i^{\otimes l}(t)\exp\{\boldsymbol{\beta}^{\mathrm{T}}Z_i(t)\},$$

对于向量 a, 有 $a^{\otimes 0}=1$, $a^{\otimes 1}=a$ 和 $a^{\otimes 2}=aa^{\mathrm{T}}$. 令 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别表示矩阵 A 的最小和最大特征根. $\|\boldsymbol{t}\|=(\boldsymbol{t}^{\mathrm{T}}\boldsymbol{t})^{1/2}$ 代表列向量 $\boldsymbol{t}$ 的 L_2 范数.

需要如下条件来保证 SELO 惩罚估计量的理论性质:

(A1) $\int_0^{\tau}\lambda_0(t)dt<\infty$.

(A2) 存在真值 $\boldsymbol{\beta}_0$ 的邻域 $\mathcal{B}$ 满足如下条件:

(i) 存在定义在 $\mathcal{B}\times[0,\tau]$ 上的矢量、向量和矩阵 $s^{(l)}(\boldsymbol{\beta};t)$ $(l=0,1,2)$ 满足

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}}\|S^{(l)}(\boldsymbol{\beta};t)-s^{(l)}(\boldsymbol{\beta};t)\|\to_p 0,$$

其中 $\to_p$ 代表依概率收敛.

(ii) $s^{(0)}(\boldsymbol{\beta};t)$ 在 $\mathcal{B}\times[0,\tau]$ 上远离 0. 对于 $l=0,1,2$, $\boldsymbol{\beta}\in\mathcal{B}$, $s^{(l)}(\boldsymbol{\beta};t)$ 关于 $t\in[0,\tau]$ 绝对连续.

(iii) 令 $M_i(t)=N_i(t)-\int_0^t Y(u)\exp(\boldsymbol{\beta}_0^{\mathrm{T}}Z_i(u))d\Lambda_0(u)$ 为局部平方可积鞅. 定义

$$D_i=\int_0^{\tau}\left[Z_i(t)-\frac{s^{(1)}(\boldsymbol{\beta}_0;t)}{s^{(0)}(\boldsymbol{\beta}_0;t)}\right]dM_i(t).$$

假定存在矩阵 $\Gamma(\boldsymbol{\beta}_0)$ 使得

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(D_i)-\Gamma(\boldsymbol{\beta}_0)\right\|\to 0.$$

进一步假定存在常数 C_1 和 C_2 使得

$$0<C_1<\lambda_{\min}(\Gamma(\boldsymbol{\beta}_0))\leqslant\lambda_{\max}(\Gamma(\boldsymbol{\beta}_0))<C_2<\infty.$$

(A3) 定义

$$v(\boldsymbol{\beta};t)=\frac{s^{(2)}(\boldsymbol{\beta};t)}{s^{(0)}(\boldsymbol{\beta};t)}-\left(\frac{s^{(1)}(\boldsymbol{\beta};t)}{s^{(0)}(\boldsymbol{\beta};t)}\right)^{\otimes 2},$$
$$A(\boldsymbol{\beta})=\int_0^{\tau}[v(\boldsymbol{\beta};t)s^{(0)}(\boldsymbol{\beta};t)]\lambda_0(t)dt.$$

假定存在常数 C_3 和 C_4 使得如下不等式成立:

$$0<C_3<\lambda_{\min}(A(\boldsymbol{\beta}_0))\leqslant\lambda_{\max}(A(\boldsymbol{\beta}_0))<C_4<\infty.$$

(A4) 令 D_{ik} 表示 D_i 的第 k 个元素, 假定

$$\sup_{1\leqslant i\leqslant n}\mathrm{E}(D_{ik}^2D_{il}^2)<C_5<\infty,\quad 1\leqslant k,l\leqslant d_n.$$

(A5) $d_n^4/n\to 0$.

(A6) 调节参数 λ_n 和 γ_n 满足 $\gamma_n=O(\sqrt{1/(d_nn)})$, $\lambda_n\gamma_n(n/d_n)^{3/2}\to\infty$.

(A7) 令 $\rho_n=\min_{j\in\mathcal{A}}|\beta_j|$, 则 $\lambda_n/\rho_n^2\to 0$.

当协变量维数 d_n 发散时, 条件 (A1)–(A5) 与 Cai 等 [11] 提出的条件一致, 这些条件用来保证偏似然函数的局部渐近二次性质, 从而保证了惩罚估计量的渐近正态性.

定理 1 的证明 令 $\alpha_n=\sqrt{d_n/n}$, 我们只需证明对于任意的 $\epsilon>0$, 存在足够大的常数 C 使得

$$\mathrm{P}\Big\{\sup_{\|\boldsymbol{u}\|=c} Q_{ns}(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})<Q_{ns}(\boldsymbol{\beta}_0)\Big\}\geqslant 1-\epsilon.$$

从而存在 $Q_{ns}(\boldsymbol{\beta})$ 的局部极大值点 $\widehat{\boldsymbol{\beta}}$ 满足 $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|=O_p(\alpha_n)$.

由 $Q_{ns}(\boldsymbol{\beta})$ 的定义知,

$$\begin{aligned}
&Q_{ns}(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-Q_{ns}(\boldsymbol{\beta}_0)\\
&\quad=\{\ell_n(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-\ell_n(\boldsymbol{\beta}_0)\}-n\sum_{r=1}^{d_n}\{P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0}+\alpha_n u_r;\lambda_n,\gamma_n)-P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0};\lambda_n,\gamma_n)\}\\
&\quad\leqslant\{\ell_n(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-\ell_n(\boldsymbol{\beta}_0)\}-n\sum_{r\in K(\boldsymbol{u})}\{P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0}+\alpha_n u_r;\lambda_n,\gamma_n)-P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0};\lambda_n,\gamma_n)\}\\
&\quad\triangleq I_1+I_2,
\end{aligned}$$

其中

$$K(\boldsymbol{u})=\{r: P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0}+\alpha_n u_r;\lambda_n,\gamma_n)-P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0};\lambda_n,\gamma_n)<0\},$$

u_r 代表 $\boldsymbol{u}$ 的第 r 个元素. 由 Taylor 展开式得

$$\begin{aligned}
I_1&=\ell_n(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-\ell_n(\boldsymbol{\beta}_0)\\
&=\alpha_n\boldsymbol{u}^{\mathrm{T}}\ell_n'(\boldsymbol{\beta}_0)+\frac{1}{2}\alpha_n^2\boldsymbol{u}^{\mathrm{T}}\ell_n''(\boldsymbol{\beta}^*)\boldsymbol{u}\\
&\triangleq I_{11}+I_{12},
\end{aligned}$$

其中 $\boldsymbol{\beta}^*$ 位于 $\boldsymbol{\beta}_0$ 与 $\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u}$ 之间. 由文献 [11] 中的结论, 当 $d_n^4/n\to 0$ 时, $\|\ell_n'(\boldsymbol{\beta}_0)\|=O_p(\sqrt{nd_n})$,

$$\|n^{-1}\ell_n''(\boldsymbol{\beta})+A(\boldsymbol{\beta})\|=o_p(d_n^{-1}),\tag{A.1}$$

在 $\boldsymbol{\beta}\in\mathcal{B}$ 上一致成立. $A(\boldsymbol{\beta}_0)$ 为有限的正定阵, 因此, 当 C 足够大时, $I_{11}=O_p(n\alpha_n^2)\|\boldsymbol{u}\|$ 在 $\|\boldsymbol{u}\|=C$ 上被 $I_{12}=O_p(n\alpha_n^2)\|\boldsymbol{u}\|^2$ 一致控制.

由 SELO 惩罚函数在 $[0,\infty)$ 上的凸性可知, 对于任意的 $r\in K(\boldsymbol{u})$ 和足够大的 n, 有

$$P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0}+\alpha_n u_r;\lambda_n,\gamma_n)-P_{\mathrm{SELO}}(\boldsymbol{\beta}_{r0};\lambda_n,\gamma_n)>-P_{\mathrm{SELO}}'(\boldsymbol{\beta}_{r0}+\alpha_n u_r;\lambda_n,\gamma_n)\alpha_n|u_r|.$$

从而, 由 P_{SELO} 的表达式、Cauchy-Schwartz 不等式、条件 (A6) 和 (A7), 可得

$$\begin{aligned}
I_2&\leqslant n\sum_{r\in K(\boldsymbol{u})}\frac{\lambda_n\gamma_n}{2\log(2)\rho_n^2}\alpha_n|u_r|\\
&\leqslant n\alpha_n\sqrt{d_n}\|\boldsymbol{u}\|\frac{\lambda_n\gamma_n}{2\log(2)\rho_n^2}\\
&=O(n\alpha_n\lambda_n\gamma_n\rho_n^{-2}\sqrt{d_n})\|\boldsymbol{u}\|\\
&=o(n\alpha_n^2)\|\boldsymbol{u}\|.
\end{aligned}$$

定理得证. □

为了证明 SELO 惩罚估计量的 oracle 性质, 我们首先证明惩罚估计量是稀疏的, 即 $\widehat{\boldsymbol{\beta}}_2=\mathbf{0}$.

引理 1 假定定理 2 的条件成立, 则对于任意满足 $\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10}\|=O_p(\sqrt{d_n/n})$ 的 $\boldsymbol{\beta}_1$ 以及任意正数 C 以趋向于 1 的概率, 有

$$Q_{ns}\{(\boldsymbol{\beta}_1{}^{\mathrm{T}},\mathbf{0}^{\mathrm{T}})\}=\max_{\|\boldsymbol{\beta}_2\|\leqslant C\sqrt{d_n/n}}Q_{ns}\{(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}})\}.$$

证明 记 $\epsilon_n=C\sqrt{d_n/n}=C\alpha_n$. 只需证明对于任意满足条件 $\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10}\|=O_p(\alpha_n)$ 和 $\|\boldsymbol{\beta}_2\|\leqslant C\alpha_n$ 的 $\boldsymbol{\beta}=(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}})$, 当 $r=s_n+1,\ldots,d_n$ 时, 如下不等式以趋向于 1 的概率成立:

$$\begin{aligned}&\frac{\partial Q_{ns}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}<0,\quad \text{当 } 0<\boldsymbol{\beta}_r<\epsilon_n \text{ 时},\\&\frac{\partial Q_{ns}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}>0,\quad \text{当 } -\epsilon_n<\boldsymbol{\beta}_r<0 \text{ 时}.\end{aligned}$$

令 $\mathrm{sgn}(\cdot)$ 为符号函数, 则有

$$\begin{aligned}\frac{\partial Q_{ns}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}&=\frac{\partial\ell_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}-nP'_{\mathrm{SELO}}(|\boldsymbol{\beta}_r|;\lambda_n,\gamma_n)\mathrm{sgn}(\boldsymbol{\beta}_r)\\&=\frac{\partial\ell_n(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}_r}+\sum_{j=1}^{d_n}\frac{\partial^2\ell_n(\boldsymbol{\beta}^*)}{\partial\boldsymbol{\beta}_r\partial\boldsymbol{\beta}_j}(\boldsymbol{\beta}_j-\boldsymbol{\beta}_{j0})-nP'_{\mathrm{SELO}}(|\boldsymbol{\beta}_r|;\lambda_n,\gamma_n)\mathrm{sgn}(\boldsymbol{\beta}_r)\\&\triangleq J_1+J_2+J_3,\end{aligned}$$

其中 $\boldsymbol{\beta}^*$ 位于 $\boldsymbol{\beta}$ 与 $\boldsymbol{\beta}_0$ 之间. 由文献 [17] 中的结论知,

$$J_1=O_p(\sqrt{n}).$$

$A_r(\boldsymbol{\beta}_0)$ 为矩阵 $A(\boldsymbol{\beta}_0)$ 的第 r 行, $A_{rl}(\boldsymbol{\beta}_0)$ 表示矩阵 $A(\boldsymbol{\beta}_0)$ 的第 (r,l) 元素. 由文献 [11] 以及条件 (A2) 可知,

$$\begin{aligned}|J_2|&=|-nA_r(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\{1+o_p(1)\}|\\&\leqslant nO_p(\alpha_n)\Big\{\sum_{l=1}^{d_n}A_{rl}(\boldsymbol{\beta}_0)^2\Big\}^{1/2}.\end{aligned}$$

由条件 (A3) 知, 矩阵 $A(\boldsymbol{\beta}_0)$ 的最大特征根有界, 从而,

$$\begin{aligned}\Big\{\sum_{l=1}^{d_n}A_{rl}(\boldsymbol{\beta}_0)^2\Big\}&=(A(\boldsymbol{\beta}_0)A(\boldsymbol{\beta}_0)^{\mathrm{T}})_{r,r}=e_r^{\mathrm{T}}(A(\boldsymbol{\beta}_0)A(\boldsymbol{\beta}_0)^{\mathrm{T}})e_r\\&\leqslant\lambda_{\max}(A(\boldsymbol{\beta}_0)A(\boldsymbol{\beta}_0)^{\mathrm{T}})\\&<\infty,\end{aligned}$$

其中 e_r 是第 r 个元素为 1、其他所有元素均为 0 的 d_n 维向量. $(A(\boldsymbol{\beta}_0)A(\boldsymbol{\beta}_0)^{\mathrm{T}})_{r,r}$ 是矩阵 $A(\boldsymbol{\beta}_0)A(\boldsymbol{\beta}_0)^{\mathrm{T}}$ 的第 (r,r) 元. 从而, $J_2=O_p(\sqrt{nd_n})$, 则有 $J_1+J_2=O_p(\sqrt{nd_n})$ 以及

$$\begin{aligned}\frac{\partial Q_{ns}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}&=O_p(\sqrt{nd_n})-nP'_{\mathrm{SELO}}(|\boldsymbol{\beta}_r|;\lambda_n,\gamma_n)\mathrm{sgn}(\boldsymbol{\beta}_r)\\&=n\alpha_n\Big\{O_p(1)-\frac{P'_{\mathrm{SELO}}(|\boldsymbol{\beta}_r|;\lambda_n,\gamma_n)}{\alpha_n\mathrm{sgn}(\boldsymbol{\beta}_r)}\Big\}.\end{aligned}$$

当 $\|\boldsymbol{\beta}_2\| \leqslant C\alpha_n$ 时, 由 SELO 的凸性可知, 对于任意的 $r = s_n + 1, \ldots, d_n$, 均有

$$\begin{aligned}\frac{P'_{\mathrm{SELO}}(|\boldsymbol{\beta}_r|; \lambda_n, \gamma_n)}{\alpha_n} &> \frac{\lambda_n\gamma_n/\alpha_n}{\log(2)(2C\alpha_n + \gamma_n)(C\alpha_n + \gamma_n)} \\ &= O\bigg(\frac{\lambda_n\gamma_n/\alpha_n}{\alpha_n^2}\bigg) \\ &\to \infty.\end{aligned}$$

从而 $\partial Q_{ns}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_r$ 的符号由 $\boldsymbol{\beta}_r$ 的符号完全确定. 引理得证. □

定理 2 的证明 第一部分的证明可由上面的引理直接得到. 下证定理 2 的第二部分.

由 $\widehat{\boldsymbol{\beta}}$ 的定义, 当 $r = 1, \ldots, s_n$ 时,

$$\frac{\partial Q_{ns}(\widehat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}_r} = 0, \quad r = 1, \ldots, s_n,$$

即

$$\ell'_{n1}(\widehat{\boldsymbol{\beta}}) - nP'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}; \lambda_n, \gamma_n) = 0, \tag{A.2}$$

其中

$$P'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}; \lambda_n, \gamma_n) = (P'_{\mathrm{SELO}}(|\hat{\boldsymbol{\beta}}_1|; \lambda_n, \gamma_n)\mathrm{sgn}(\hat{\boldsymbol{\beta}}_1), \ldots, P'_{\mathrm{SELO}}(|\hat{\boldsymbol{\beta}}_{s_n}|; \lambda_n, \gamma_n)\mathrm{sgn}(\hat{\boldsymbol{\beta}}_{s_n}))^{\mathrm{T}},$$

$\ell'_{n1}(\widehat{\boldsymbol{\beta}})$ 是 $\ell'_n(\widehat{\boldsymbol{\beta}})$ 的前 s_n 个元素组成的列向量.

令 $\ell''_{n11}(\boldsymbol{\beta}^*)$ 代表 $\ell''_n(\boldsymbol{\beta}^*)$ 的前 s_n 行前 s_n 列元素组成的子阵. 由 (A.2), 对 $\ell_n(\boldsymbol{\beta})$ 在 $\boldsymbol{\beta}_0$ 处 Taylor 展开, 则有

$$\ell'_{n1}(\boldsymbol{\beta}_0) - nP'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}_1; \lambda_n, \gamma_n) = -\ell''_{n11}(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}),$$

其中 $\boldsymbol{\beta}^*$ 位于 $\boldsymbol{\beta}_0$ 与 $\widehat{\boldsymbol{\beta}}$ 之间. 由文献 [11]、条件 (A5) 和定理 1, 可得

$$n^{-1}\ell'_{n1}(\boldsymbol{\beta}_0) - P'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}_1; \lambda_n, \gamma_n) = A_{11}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) + o_p(n^{-1/2}).$$

由 $P'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}_1; \lambda_n, \gamma_n)$ 的定义、条件 (A7) 和定理 1, 有

$$\begin{aligned}\|\sqrt{n}P'_{\mathrm{SELO}1}(\widehat{\boldsymbol{\beta}}_1; \lambda_n, \gamma_n)\| &\leqslant \sqrt{ns_n}\max_{1\leqslant j\leqslant s_n}(P'_{\mathrm{SELO}}(|\hat{\boldsymbol{\beta}}_j|; \lambda_n, \gamma_n)) \\ &= \sqrt{ns_n}O_p(\rho_n^{-2}\lambda_n\gamma_n) = o_p(1).\end{aligned}$$

从而, 对于任意的 s_n 维常数向量 c_n, 有

$$\sqrt{n}c_n^{\mathrm{T}}\Gamma_{11}^{-1/2}A_{11}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) = n^{-1/2}c_n^{\mathrm{T}}\Gamma_{11}^{-1/2}\ell'_{n1}(\boldsymbol{\beta}_0) + o_p(1).$$

当条件 (A1)–(A5) 满足时, Cai 等 [11] 证明了

$$n^{-1/2}c_n^{\mathrm{T}}\Gamma_{11}^{-1/2}\ell'_{n1}(\boldsymbol{\beta}_0) \to_d N(0, 1),$$

其中符号 $\to_d$ 表示依分布收敛. 从而有 Slutsky 定理,

$$\sqrt{n}c_n\Gamma_{11}^{-1/2}A_{11}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \to_d N(0, 1).$$

定理 2 证毕. □

Penalized variable selection procedure for Cox proportional hazards model via seamless-L_0 penalty

Yongxiu Cao, Yuling Jiao, Yueyong Shi & Yanyan Liu

Abstract In biomedical studies, it is of substantial interest to select important risk factors using high-dimensional data such as gene expression data for outcomes which are subject to censoring. The Cox proportional hazards model is widely used for the analysis of censored outcome data. This paper concerns variable selection and estimation problem for the analysis of high-dimensional censored outcome data in Cox model using the seamless-L_0 penalized partial likelihood method. Under the assumption of diverging number of covariates, we establish the consistency and oracle property of the corresponding penalized estimators. The calculation is challenge because of the non-smoothness of the seamless-L_0 penalty and the expensive calculation of the inverse of Hessian matrix when the number of coefficients is large. We approximate the seamless-L_0 penalty by a smoothing technique and the inverse of Hessian matrix by the DFP formula which is widely used in optimization theory. We show through simulation studies that our proposed approach has better performance compared with several existing methods and the proposed algorithm has some advantages over the generally used coordinate descent algorithm. We analyzed a breast cancer data and evaluated the prediction performance of several methods based on a leave-one-out cross validation method.

Keywords **coordinate descent algorithm, penalized partial likelihood, quasi-Newton algorithm, variable selection**

MSC(2010) **62N02**

doi: 10.1360/SCM-2016-0609