DOI: 10.14188/j.1671-8836.2018.06.008

高维广义线性模型的惩罚拟似然 SCAD 估计

闫 莉,陈 夏†
(陕西师范大学 数学与信息科学学院,陕西 西安 710119)

摘 要: 基于光滑切片绝对偏差(smoothly clipped absolute deviation,SCAD)惩罚的拟似然方法,研究高维广义线性模型的变量选择和参数估计问题.所提方法只需要对响应变量期望函数和方差函数的正确设定.在适当的正则条件下,证明了拟似然 SCAD 估计具有相合性和 Oracle 性质.最后通过数据模拟和实例分析,验证了所提方法的有限样本性质.

关 键 词: 广义线性模型;高维数据;变量选择;拟似然方法;光滑切片绝对偏差
中图分类号: O 212.1 **文献标识码:** A **文章编号:** 1671-8836(2018)06-0533-07

Penalized Quasi-Likelihood SCAD Estimator in High-Dimensional Generalized Linear Models

YAN Li, CHEN Xia†
(School of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710119, Shaanxi, China)

Abstract: This paper considers the problem of variable selection and estimation in high-dimensional generalized linear models based on the quasi-likelihood method with SCAD(smoothly clipped absolute deviation) penalty. This method requires only the correct specification of the expectation and variance function of the response variable. Under reasonable conditions, the consistency and the Oracle property of the quasi-likelihood SCAD estimator is proved. Some simulations and a real data analysis are given to illustrate the and good performance of the proposed method.

Key words: generalized linear models; high-dimensional data; variable selection; quasi-likelihood method; SCAD (smoothly clipped absolute deviation)

0 引 言

广义线性模型(generalized linear models, GLMs)的理论是对经典线性模型理论的重要推广,自 Nelder 与 Wedderburn[1] 正式引入 GLMs 以来,已经成功应用到了很多领域.起初,研究者们只局限于响应变量服从指数分布族的情形.1974 年,Wedderburn[2] 提出了 GLMs 拟似然函数的概念,该方法在建模时可以只要求响应变量期望函数与方差函数有正确设定,不要求响应变量的分布为指数型分布族.后续研究表明,只对期望函数有正确设定但方差函数未知的情况下,此方法依然有效[3,4].这种研究方法称为拟似然方法,得到的参数估计称为拟似然估计.

在经济和生物医学等领域,常常会遇到多变量大样本的数据,因此变量选择在高维数据分析中发挥着重要作用.在高维数据变量选择问题的研究中,应用较为广泛的是基于惩罚函数的估计方法.它可以同时实现变量选择和参数估计.常见的惩罚函数有 SCAD(smoothly clipped absolute deviation)[5]、Bridge[6]、Lasso[7] 以及 Lasso 的衍生版本,如 Adaptive Lasso[8]、Elastic Net[9]、Adaptive Elastic Net[10] 及针对群组数据的 Group Lasso[11] 等.对于经典线性模型,基于惩罚函数的变量选择方法已经取得了众多研究成果.例如,Huang 和 Xie[12] 研究了高维线性模

收稿日期: 2018-04-05 †通信联系人 E-mail: xchen80@snnu.edu.cn
基金项目: 国家自然科学基金(11801346);教育部人文社会科学研究青年基金(18YJC910014);陕西省自然科学基础研究计划(2018JM1024);中央高校基本科研业务费专项资金(GK201503012, GK201503015)
作者简介: 闫 莉,女,讲师,主要从事复杂数据统计分析方面的研究. E-mail: lyan@snnu.edu.cn

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

型下 SCAD 估计的渐近性质. Knight 和 Fu [13] 研究了参数个数有限时,Bridge 估计的渐近性质. Huang 等 [14] 又将此结果拓展到参数无穷维的情形. 在此基础上,Wang 等 [15] 研究了广义线性模型中参数发散时,Bridge 估计的渐近性质. Xie 和 Huang [16] 研究了 SCAD 惩罚下,高维部分线性模型参数估计及其渐近性质. Fan 和 Li [5] 以及 Fan 和 Peng [17] 证明了参数发散时,SCAD 估计仍具有 Oracle 性质. Zou [8] 证明了 Lasso 方法不具有 Oracle 性质,并提出了 Adaptive Lasso 方法. 更多关于高维变量选择的研究参见文献[18].

在文献中尚未见到利用惩罚的拟似然 SCAD 方法,研究高维广义线性模型的参数估计和变量选择问题. 拟似然方法不需要知道响应变量的分布,只需均值函数和方差函数的正确设定,这大大扩大了其应用范围. 基于此,本文考虑在高维广义线性模型中,利用惩罚拟似然 SCAD 方法进行变量选择和参数估计. 在适当的条件下,证明了拟似然 SCAD 估计的相合性和 Oracle 性质. 最后通过数据模拟和实例分析,说明了所提方法的有限样本性质.

1 GLMs 的惩罚拟似然 SCAD 估计

考虑如下广义线性模型

$$y_i = g(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_n) + \varepsilon_i, \quad i = 1,2,\cdots,n$$

其中 y_i 是响应变量,$\boldsymbol{X}_i = (X_{i1},\cdots,X_{ip_n})^{\mathrm{T}}$ 是 $p_n \times 1$ 维协变量,$\boldsymbol{\beta}_n$ 是 $p_n \times 1$ 维的未知参数向量,$g(\cdot)$ 的逆是联系函数. $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是相互独立的随机误差序列项,满足 $E(\varepsilon_i) = 0$,$\mathrm{Var}(\varepsilon_i) = \sigma(\mu_i)$,这里 $\mu_i = g(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$.

1974 年,Wedderburn 提出如下拟似然函数 [2]

$$Q_n^*(\boldsymbol{\beta}_n) = \sum_{i=1}^{n}\int_{y_i}^{\mu_i(\boldsymbol{\beta}_n)}\frac{y_i - t}{\sigma(t)}\mathrm{d}t$$

其中 $\mu_i(\boldsymbol{\beta}_n) = g(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$. 参数 $\boldsymbol{\beta}_n$ 的极大拟似然估计定义为方程

$$U(\boldsymbol{\beta}_n) = \frac{\partial Q_n^*(\boldsymbol{\beta}_n)}{\partial\boldsymbol{\beta}_n} = \sum_{i=1}^{n}(y_i - \mu_i(\boldsymbol{\beta}_n))\sigma^{-1}(\mu_i(\boldsymbol{\beta}_n))g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_n)\boldsymbol{X}_i = 0 \quad (1)$$

的根. 这里 $Q_n^*(\boldsymbol{\beta}_n)$ 和指数型情况下所定义的对数似然函数的作用是相同的,而方程(1)的作用也与指数情形下的对数似然方程一致.

惩罚拟似然目标函数定义为

$$Q_n(\boldsymbol{\beta}_n) = Q_n^*(\boldsymbol{\beta}_n) - n\sum_{j=1}^{p_n}p_{\lambda_n}(|\boldsymbol{\beta}_{nj}|) \quad (2)$$

其中 $p_{\lambda_n}(\cdot)$ 是惩罚函数,λ_n 是调整参数,SCAD 惩罚函数 $p_{\lambda_n}(\cdot)$ 的一阶导函数为

$$p'_{\lambda_n}(|x|) = \lambda_n\{I(x \leqslant \lambda_n) + \frac{(a\lambda_n - x)_+}{(a-1)\lambda_n}I(x > \lambda_n)\}$$

其中 $\lambda_n > 0$,$a > 2$. 这里取 Fan 和 Li 所建议的固定调整参数 $a = 3.7$ [5]. 最大化惩罚拟似然目标函数 $Q_n(\boldsymbol{\beta}_n)$,得到参数 $\boldsymbol{\beta}_n$ 的拟似然 SCAD 估计 $\hat{\boldsymbol{\beta}}_n$.

记参数真值为 $\boldsymbol{\beta}_{n0}$,简记为 $\boldsymbol{\beta}_0$. 令 $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}^{\mathrm{T}},\boldsymbol{\beta}_{20}^{\mathrm{T}})^{\mathrm{T}}$,其中 $\boldsymbol{\beta}_{10}$ 是 k_n 维非零向量,$\boldsymbol{\beta}_{20}$ 是 m_n 维零向量,这里 k_n 为非零系数的个数,m_n 为零系数的个数. 同样地,令 $\boldsymbol{X} = (\boldsymbol{w}^{\mathrm{T}},\boldsymbol{z}^{\mathrm{T}})^{\mathrm{T}}$,相应的样本为 $\boldsymbol{X}_i = (\boldsymbol{w}_i^{\mathrm{T}},\boldsymbol{z}_i^{\mathrm{T}})^{\mathrm{T}}$,其中 $\boldsymbol{w}_i = (X_{i1},\cdots,X_{ik_n})^{\mathrm{T}}$,$\boldsymbol{z}_i = (X_{i(k_n+1)},\cdots,X_{ip_n})^{\mathrm{T}}$.

2 主要结果

首先给出本文中的一些记号.

$$U(\boldsymbol{\beta}) = \frac{\partial Q_n^*(\boldsymbol{\beta})}{\partial\beta} = \sum_{i=1}^{n}(y_i - \mu_i(\boldsymbol{\beta}))\sigma^{-1}(\mu_i(\boldsymbol{\beta}))\cdot g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{X}_i \triangleq \sum_{i=1}^{n}D(\boldsymbol{\beta},\boldsymbol{X}_i)\boldsymbol{X}_i$$

$$I_n^*(\boldsymbol{\beta}) = -\frac{\partial^2 Q_n^*(\beta)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} = \sum_{i=1}^{n}\{[g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})]^2\sigma^{-1}(\mu_i) + [g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})]^2\sigma^{-2}(\mu_i)\sigma'(\mu_i) - g''(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})\sigma^{-1}(\mu_i)]\cdot(y_i - \mu_i(\boldsymbol{\beta}))\}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}} \triangleq \sum_{i=1}^{n}d_1(\boldsymbol{\beta},\boldsymbol{X}_i)\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}$$

$$\frac{\partial^3 Q_n^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{nj}\partial\boldsymbol{\beta}_{nk}\partial\boldsymbol{\beta}_{nl}} = \sum_{i=1}^{n}\{3g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})g''(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})\sigma^{-1}(\mu_i) - 2[g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})]^3\sigma^{-2}(\mu_i)\sigma'(\mu_i) + [3g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})g''(\boldsymbol{X}_\beta^{\mathrm{T}}\boldsymbol{\beta})\cdot\sigma^{-2}(\mu_i)\sigma'(\mu_i) - 2[g'(\boldsymbol{\beta})]^3\sigma^{-3}(\mu_i)[\sigma'(\mu_i)]^3 + [g'(\boldsymbol{\beta})]^3\sigma^{-2}(\mu_i)\sigma\sigma''(\mu_i) - g'''(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})\sigma^{-1}(\mu_i)]\cdot(y_i - \mu_i(\boldsymbol{\beta}))\}\boldsymbol{X}_{ij}\boldsymbol{X}_{ik}\boldsymbol{X}_{il} \triangleq \sum_{i=1}^{n}d_2(\boldsymbol{\beta},\boldsymbol{X}_i)\boldsymbol{X}_{ij}\boldsymbol{X}_{ik}\boldsymbol{X}_{il}$$

$$\boldsymbol{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}[g'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})]^2\sigma^{-1}(\mu_i)\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}$$

$$\boldsymbol{\Sigma}_{1n} = \frac{1}{n}\sum_{i=1}^{n}[g'(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{\beta})]^2\sigma^{-1}(\mu_i)\boldsymbol{w}_i\boldsymbol{w}_i^{\mathrm{T}}$$

令 λ_{1n} 和 λ_{2n} 分别是 $\boldsymbol{\Sigma}_n$ 的最小和最大特征值,τ_{1n} 和 τ_{2n} 分别是 $\boldsymbol{\Sigma}_{1n}$ 的最小和最大特征值. 下面列出本文所需要的条件:

条件 1 g 定义于 $[A,B] = [\underline{\mu} - \varepsilon,\bar{\mu} + \varepsilon]$,对于某个 $\varepsilon > 0$,其中 $\underline{\mu} = \inf\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}: i \geqslant 1,\boldsymbol{\beta} \in \Theta_n\}$,$\bar{\mu} = \sup\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}: i \geqslant 1,\boldsymbol{\beta} \in \Theta_n\}$,$g$ 的三阶导数存在且在 $[A,B]$ 上有界,g 的导数 $[A,B]$ 上处处不为 0,σ 定义于 $[g(A),g(B)]$,在其上有有界的二阶导数,且 σ 在 $[g(A),g(B)]$ 上有非 0 下界.

条件 2 $\{\boldsymbol{X}_i, i \geqslant 1\}$ 有界.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

条件 3　$p_n\to\infty$, $p_n^3/n\to 0$, $(nk_n)1/2\lambda_n\to 0$.

条件 4　存在常数 b_0, b_1, 且 $0<b_0<b_1$, 使得 $b_0\leqslant\min\{|\boldsymbol{\beta}_{10j}|, 1\leqslant j\leqslant k_n\}\leqslant\max\{|\boldsymbol{\beta}_{10j}|, 1\leqslant j\leqslant k_n\}\leqslant b_1$, 其中 $\boldsymbol{\beta}_{10j}$ 是 $\boldsymbol{\beta}_{10}$ 的第 j 个元素.

条件 5　存在常数 $\lambda>0$ 和 $0<\tau_1<\tau_2<\infty$ 使得 $\lambda\leqslant\lambda_{1n}, \tau_1\leqslant\tau_{1n}\leqslant\tau_{2n}\leqslant\tau_2$.

下面考虑惩罚拟似然估计的渐近性质. 在本文中 C 表示一个正常数, 它可以在不同的位置表示不同的数.

定理 1(相合性)　在条件 1 ~ 5 下, 如果 $n\lambda_n^2=O(1)$, 则(2)式存在一个局部最大值点 $\hat{\boldsymbol{\beta}}_n$, 使得

$$\|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\|=O_p(\sqrt{p_n/n})$$

证　令 $\alpha_n=\sqrt{p_n/n}$, 下面要证明对任给的 $\varepsilon>0$, 有一个充分大的常数 C, 对足够大的 n, 有

$$\Pr\{\sup_{\|\boldsymbol{u}\|=C}Q_n(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})<Q_n(\boldsymbol{\beta}_0)\}\geqslant 1-\varepsilon$$

即以概率趋向于 1 在球$\{\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u}, \|\boldsymbol{u}\|\leqslant C\}$内存在一个局部最大值点 $\hat{\boldsymbol{\beta}}_n$, 使得 $\|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\|=O_p(\sqrt{p_n/n})$, 其中 $\boldsymbol{u}$ 是一个 $p_n\times 1$ 的向量.

易知

$$Q_n(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-Q_n(\boldsymbol{\beta}_0)=\{Q_n^*(\boldsymbol{\beta}_0+\alpha_n\boldsymbol{u})-Q_n^*(\boldsymbol{\beta}_0)\}+n\sum_{j=1}^{k_n}p_{\lambda_n}(|\boldsymbol{\beta}_{0j}|)\triangleq I_1+I_2 \quad (3)$$

首先考虑第一项 I_1. 利用泰勒展开, 得到

$$I_1=\alpha_n U(\boldsymbol{\beta}_0)\boldsymbol{u}-\frac{1}{2}\alpha_n^2\boldsymbol{u}^{\mathrm{T}}I_n^*(\boldsymbol{\beta}_0)\boldsymbol{u}-\frac{1}{6}\alpha_n^3\nabla^{\mathrm{T}}(\boldsymbol{u}^{\mathrm{T}}I_n^*(\beta_n^*)\boldsymbol{u})\boldsymbol{u}\triangleq I_{11}+I_{12}+I_{13}$$

其中 $\boldsymbol{\beta}_n^*$ 是介于 $\boldsymbol{\beta}_0+a_n\boldsymbol{u}$ 和 $\boldsymbol{\beta}_0$ 之间的向量. 对于 I_{11}, 在条件 1 和条件 2 下, 有 $U(\boldsymbol{\beta}_0)=O_p(\sqrt{np_n})$. 根据 Cauchy-Schwarz 不等式, 有

$$I_{11}\leqslant\alpha_n\|U(\boldsymbol{\beta}_0)\|\|\boldsymbol{u}\|=\alpha_n\sqrt{np_n}\|\boldsymbol{u}\|=O_P(n\alpha_n^2)\|\boldsymbol{u}\|$$

对于 I_{12}, 有

$$I_{12}=-\frac{1}{2}n\alpha_n^2\boldsymbol{u}^{\mathrm{T}}\left(\frac{1}{n}I_n^*(\boldsymbol{\beta}_0)-\boldsymbol{\Sigma}_n\right)-\frac{1}{2}n\alpha_n^2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_n\boldsymbol{u}=I_{121}+I_{122}$$

在条件 1 下, 可以得到 $E[U_1(\boldsymbol{\beta}_0, X_1)]^2\leqslant C$, 再由切比雪夫不等式, 得到

$$\|\frac{1}{n}I_n^*(\boldsymbol{\beta}_0)-\boldsymbol{\Sigma}_n\|=o_p(\frac{1}{\sqrt{p_n}}) \quad (4)$$

因此, 由 Cauchy-Schwarz 不等式, 知

$$|I_{121}|\leqslant\frac{1}{2}n\alpha_n^2\|\boldsymbol{u}\|^2\|\frac{1}{n}I_n^*(\boldsymbol{\beta}_0)-\boldsymbol{\Sigma}_n\|=\frac{1}{2}n\alpha_n^2\|\boldsymbol{u}\|^2o_p(\frac{1}{\sqrt{p_n}})=n\alpha_n^2\|\boldsymbol{u}\|^2o_p(1)$$

对于 I_{13}, 由条件 1 ~ 3, 可以得到

$$I_{13}=O_P(\alpha_n^3np_n\|\boldsymbol{u}\|^2)=o_p(n\alpha_n^2)\|\boldsymbol{u}\|^2$$

现在考虑 I_2, 如果 $n\lambda_n^2=O(1)$, 由条件 3 和条件 4, 有

$$I_2=n\sum_{j=1}^{p_n}(p_{\lambda_n}|\boldsymbol{\beta}_{0j}|)\leqslant\frac{nk_n(a+1)\lambda_n^2}{2}=O(n\alpha_n^2)$$

因此对于一个充分大的常数 C, (3)式的符号完全由 I_{122} 决定, 而这一项是负的. 定理证毕.

定理 2(Oracle 性质)　令在定理 1 中的局部最大值点为 $\hat{\boldsymbol{\beta}}_n=(\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{2n}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\hat{\boldsymbol{\beta}}_{1n}$ 和 $\hat{\boldsymbol{\beta}}_{2n}$ 分别是 $\boldsymbol{\beta}_{1n}$ 和 $\boldsymbol{\beta}_{2n}$ 的估计. 假设 $\sqrt{p_n/n}\to 0$ 且条件 1 ~ 5 都成立, 我们有

1)　**稀疏性:** $\Pr(\hat{\boldsymbol{\beta}}_{2n}=0)\to 1$, $n\to\infty$.

2)　**渐近正态性:** 令 $S_n^2\triangleq\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}\boldsymbol{\alpha}_n$, 其中 $\boldsymbol{\alpha}_n$ 为任意的 $k_n\times 1$ 维向量且满足 $\|\boldsymbol{\alpha}_n\|_2\leqslant 1$. 则

$$n^{1/2}\boldsymbol{S}_n^{-1}\boldsymbol{\alpha}_n^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10})\xrightarrow{D}N(0,1)$$

证　首先证明定理的第一个部分. 令 $\varepsilon_n=C\sqrt{p_n/n}$, 只需证明当 $n\to\infty$ 时以趋于 1 的概率对于任意给的 $\boldsymbol{\beta}_{1n}$ 满足 $\|\boldsymbol{\beta}_{1n}-\boldsymbol{\beta}_{10}\|=O_p(\sqrt{p_n/n})$, 有

$$\frac{\partial Q_n(\boldsymbol{\beta}_n)}{\partial\beta_{nj}}<0, \quad 0<\beta_{nj}<\varepsilon_n$$

$$\frac{\partial Q_n(\boldsymbol{\beta}_n)}{\partial\beta_{nj}}>0, \quad -\varepsilon_n<\beta_{nj}<0$$

其中 $j=k_n+1, \cdots, p_n$. 由于 $\|\boldsymbol{\beta}_{2n}\|\leqslant C\sqrt{p_n/n}$, 由泰勒展开, 有

$$\frac{\partial Q_n(\boldsymbol{\beta}_n)}{\partial\beta_{nj}}=U_j(\boldsymbol{\beta}_n)-n\lambda_n\mathrm{sgn}(\beta_{nj})$$

$$=U_j(\boldsymbol{\beta}_0)+\sum_{j=1}^{p_n}\frac{\partial U_j(\boldsymbol{\beta}_0)}{\partial\beta_{nk}}(\beta_{nk}-\beta_{0k})+\frac{1}{2}\sum_{k,l=1}^{p_n}\frac{\partial^2U_j(\beta_0^*)}{\partial\beta_{nk}\partial\beta_{nl}}(\beta_{nk}-\beta_{0k})(\beta_{nl}\beta_{0l})-n\lambda_n\mathrm{sgn}(\beta_{nj})$$

$$\triangleq J_{n1}+J_{n2}+J_{n3}+J_{n4}$$

其中 $\boldsymbol{\beta}_0^*$ 是介于 $\boldsymbol{\beta}_n$ 和 $\boldsymbol{\beta}_0$ 之间的向量, 对于第一项 J_{n1}, 由条件 1, 我们得到

$$J_{n1}=O_p(\sqrt{n})=o_p(\sqrt{np_n}) \quad (5)$$

对于 J_{n2}, 有

$$J_{n2}=n\sum_{k=1}^{p_n}\left[\frac{1}{n}\frac{\partial U_j(\boldsymbol{\beta}_0)}{\partial\beta_{nk}}+\Sigma_{njk}\right](\beta_{nk}-\beta_{0k})-n\sum_{k=1}^{p_n}\Sigma_{njk}(\beta_{nk}-\beta_{0k})\triangleq J_{n21}+J_{n22}$$

根据 Cauchy-Schwarz 不等式和(4)式, 得到

$$|J_{n21}|\leqslant n\left\{\sum_{k=1}^{n}\left[\frac{1}{n}I_n^*(\boldsymbol{\beta}_0)-\Sigma_{njk}\right]^2\right\}\frac{1}{2}\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|=O_p(n/\sqrt{p_n})O_p(\sqrt{\frac{p_n}{n}})=O_p(\sqrt{np_n})$$

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

对于 J_{n22}，有

$$|J_{n22}| \leqslant n\left\{\sum_{k=1}^{p_n}\Sigma_{njk}^2\right\}^{1/2}\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\| \leqslant n\lambda_{\max}(\boldsymbol{\Sigma}_n)$$

$$O_p\left(\sqrt{\frac{p_n}{n}}\right) = O_p(\sqrt{np_n})$$

因此，得到

$$|J_{n2}| = O_p(\sqrt{np_n}) \tag{6}$$

对于 J_{n3}，有

$$J_{n3} = \frac{1}{2}\left(\sum_{l,k=1}^{p_n}\frac{\partial^2 U_j(\boldsymbol{\beta}_0^*)}{\partial\beta_{nk}\partial\beta_{nl}} - E\frac{\partial^2 U_j(\boldsymbol{\beta}_0^*)}{\partial\beta_{nk}\partial\beta_{nl}}\right)(\beta_{nk}-\beta_{0k})$$

$$(\beta_{nl}-\beta_{0l}) + \frac{1}{2}\sum_{l,k=1}^{p_n}E\frac{\partial^2 U_j(\beta_0^*)}{\partial\beta_{nk}\partial\beta_{nl}}(\beta_{nk}-\beta_{0k})(\beta_{nl}-\beta_{0l})$$

$$\triangleq J_{n31}+J_{n32}$$

由条件 1，得到 $E[\partial^2 U_j(\boldsymbol{\beta}_0^*)/\partial\beta_{nk}\partial\beta_{nl}]^2 = O(n^2)$.

利用 Cauchy-Schwarz 不等式，知

$$|J_{n32}| \leqslant \frac{1}{2}\left(\sum_{l,k=1}^{p_n}E\left[\frac{\partial^2 U_j(\boldsymbol{\beta}_0^*)}{\partial\beta_{nk}\partial\beta_{nl}}\right]^2\right)^{\frac{1}{2}}\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|^2$$

$$\leqslant \frac{1}{2}np_n\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|^2 = O_p(p_n^2) = o_p(\sqrt{np_n})$$

且

$$J_{n31}^2 \leqslant \frac{1}{4}\sum_{l,k=1}^{p_n}\left(\frac{\partial^2 U_j(\boldsymbol{\beta}_0^*)}{\partial\beta_{nk}\partial\beta_{nl}} - E\frac{\partial^2 U_j(\boldsymbol{\beta}_0^*)}{\partial\beta_{nk}\partial\beta_{nl}}\right)^2\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|^4$$

在条件 1 到条件 3 下，有

$$J_{n31} = O_p\left\{\left(n^2p_n^2\frac{p_n^2}{n^2}\right)^{\frac{1}{2}}\right\} = o_p(\sqrt{np_n})$$

由 J_{n31} 和 J_{n32}，得

$$|J_{n3}| = O_p(p_n^2) = o_p(\sqrt{np_n}) \tag{7}$$

由(5)，(6)和(7)式，有

$$\frac{\partial Q_n(\boldsymbol{\beta}_n)}{\partial\beta_{nj}} = O_P(\sqrt{np_n}) - n\lambda_n\operatorname{sgn}(\beta_{nj}) =$$

$$n\lambda_n[O_p(\sqrt{p_n/n}/\lambda_n) - \operatorname{sgn}(\beta_{nj})]$$

由 $\sqrt{p_n/n}/\lambda_n\to 0$，易得 β_{nj} 的符号决定了 $\partial Q_n(\beta_n)/\partial\beta_{nj}$.

现在来证明定理的第二部分. 由定理 1 知，存在一个局部最大值点 $\hat{\boldsymbol{\beta}}_n$ 使得

$$\frac{\partial Q_n(\hat{\boldsymbol{\beta}}_{1n},0)}{\partial\boldsymbol{\beta}_{1n}} = \sum_{i=1}^{n}(y_i-\mu_i(\hat{\boldsymbol{\beta}}_{1n},0))\sigma^{-1}(\mu_i(\hat{\boldsymbol{\beta}}_{1n},0))$$

$$g'(\boldsymbol{w}_{1n}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{1n})\boldsymbol{w}_i = 0$$

由泰勒展开，有

$$\frac{\partial Q_n(\hat{\boldsymbol{\beta}}_{1n},0)}{\partial\boldsymbol{\beta}_{1n}} = U(\hat{\boldsymbol{\beta}}_{1n}) = U(\boldsymbol{\beta}_{10}) - I_n^*(\boldsymbol{\beta}_{10})$$

$$(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10}) + R_n(\boldsymbol{\beta}_{10}^*) = 0$$

其中 $\boldsymbol{\beta}_{10}^*$ 是介于 $\hat{\boldsymbol{\beta}}_{1n}$ 和 $\boldsymbol{\beta}_{1n}$ 之间的向量. 所以

$$\frac{1}{n}I_n^*(\boldsymbol{\beta}_{10})(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10}) = \frac{1}{n}U(\boldsymbol{\beta}_{10}) + \frac{1}{n}R_n^*(\boldsymbol{\beta}_{10}^*)$$

根据 Cauchy-Schwarz 不等式和(4)式，易知

$$\left|\left[\frac{1}{n}I_n^*(\boldsymbol{\beta}_{10}) - \boldsymbol{\Sigma}_{1n}\right](\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10})\right| \leqslant$$

$$\left\|\frac{1}{n}I_n^*(\boldsymbol{\beta}_{10}) - \boldsymbol{\Sigma}_{1n}\right\|\|(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10})\|$$

$$= o_p\left(\frac{1}{\sqrt{p_n}}\right)O_p\left(\sqrt{\frac{p_n}{n}}\right) = o_p\left(\frac{1}{\sqrt{n}}\right)$$

再利用上述证明 J_{n3} 的方法，可以得到

$$\left|\frac{1}{n}R_n^*(\boldsymbol{\beta}_{10}^*)\right| = o_p\left(\frac{1}{\sqrt{n}}\right)$$

因此

$$\boldsymbol{\Sigma}_{1n}(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10}) = \frac{1}{n}U(\boldsymbol{\beta}_{10}) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{8}$$

对于任意的 $k_n\times 1$ 的向量 $\boldsymbol{\alpha}_n$，由于 $y_i=\mu_i(\boldsymbol{\beta}_{10},0)+\varepsilon_i$，(8)式可改写为

$$n^{1/2}\boldsymbol{\alpha}_n^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10}) = \frac{1}{\sqrt{n}}\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}U(\boldsymbol{\beta}_{10}) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma^{-1}(\mu_i(\hat{\boldsymbol{\beta}}_{1n},0))g'(\boldsymbol{w}_{1n}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{1n})\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{w}_i\varepsilon_i + o_p(1)$$

令 $a_i=\sigma^{-1}(\mu_i(\hat{\boldsymbol{\beta}}_{1n},0))g'(\boldsymbol{w}_{1n}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{1n})\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{w}_i$，则有

$$S_n^2 \triangleq \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i\varepsilon_i\right) = \frac{1}{n}\sum_{i=1}^{n}[g'(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{\beta}_{10})]^2\sigma^{-1}$$

$$(\mu_i(\boldsymbol{\beta}_{10},0))\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{w}_i\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{\alpha}_n$$

令 $v_i=n^{-1/2}S_n^{-1}a_i$，$w_i=v_i\varepsilon_i$，现在验证 Lindeberg-Feller 中心极限定理的条件，也就是对任给的 $\varepsilon>0$，

$$\sum_{i=1}^{n}E(w_i^2I_{\{|w_i|\geqslant\varepsilon\}})\to 0$$

只需要证明 $\max\limits_{1\leqslant i\leqslant n}E\varepsilon_i^2I_{\{|v_i\varepsilon_i|\geqslant\varepsilon\}}\to 0$，或者等价于 $\max\limits_{1\leqslant i\leqslant n}|v_i|\to 0$. 由条件 1 到条件 3，可以得到

$$\max_{1\leqslant i\leqslant n}|v_i| =$$

$$\frac{1}{\sqrt{n}}S_n^{-1}\max_{1\leqslant i\leqslant n}|\sigma^{-1}(\mu_i(\boldsymbol{\beta}_{10},0))g'(w_i^{\mathrm{T}}\boldsymbol{\beta}_{10})\boldsymbol{\alpha}_n^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{w}_i| \leqslant$$

$$\frac{C}{\sqrt{n}}\max_{1\leqslant i\leqslant n}(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{\Sigma}_{1n}^{-1}\boldsymbol{w}_i)^{\frac{1}{2}} \leqslant \frac{C}{\sqrt{n}}(\tau_{1n})^{-\frac{1}{2}}\max_{1\leqslant i\leqslant n}(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{w}_i)^{\frac{1}{2}}\to 0$$

因此

$$n^{\frac{1}{2}}S_n^{-1}\boldsymbol{\alpha}_n^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{1n}-\boldsymbol{\beta}_{10}) \xrightarrow{D} N(0,1)$$

定理证毕.

3 数值模拟和实例分析

3.1 计 算

本节通过局部二次算法(LQA)来计算惩罚拟似然估计[5]. 调整参数 λ_n 的选择方式，采用 Wang 等[19]所建议的贝叶斯信息准则(BIC).

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

假设初值 $\boldsymbol{\beta}_n^{(0)}$，若 $\beta_{nj}^{(0)}$ 非常接近于 0，则令 $\hat{\beta}_{nj}^{(0)}=0$. 惩罚函数 $p_{\lambda_n}(|\beta_{nj}|)$ 可由下式进行逼近

$$p_{\lambda_n}(|\beta_{nj}|)\approx p_{\lambda_n}(|\beta_{nj}^{(0)}|)+\frac{1}{2}\{p'_{\lambda_n}(|\beta_{nj}^{(0)}|)/|\beta_{nj}^{(0)}|\}(\beta_{nj}^2-\beta_{nj}^{(0)2})$$

令 $\hat{\boldsymbol{\beta}}_n^{(k)}$ 是最优化算法第 k 步的迭代值，由牛顿迭代法得到下述迭代公式

$$\hat{\boldsymbol{\beta}}_n^{(k+1)}=\hat{\boldsymbol{\beta}}_n^{(k)}-[\nabla^2 Q_n^*(\hat{\boldsymbol{\beta}}_n^{(k)})-n\sum\nolimits_{\lambda_n}(\hat{\boldsymbol{\beta}}_n^{(k)})]^{-1}\cdot[U(\hat{\boldsymbol{\beta}}_n^{(k)})-n\sum\nolimits_{\lambda_n}(\hat{\boldsymbol{\beta}}_n^{(k)})\hat{\boldsymbol{\beta}}_n^{(k)}]$$

其中 $\nabla^2 Q_n^*(\hat{\boldsymbol{\beta}}_n^{(k)})=\partial^2 Q_n^*(\hat{\boldsymbol{\beta}}_n^{(k)})/\partial\boldsymbol{\beta}_n\partial\boldsymbol{\beta}_n^{\mathrm{T}}$，$\sum_{\lambda_n}(\hat{\boldsymbol{\beta}}_n^{(k)})$ 是一个对角矩阵，其第 (j,j) 个元素是 $p'_{\lambda_n}(\hat{\beta}_{nj}^{(k)})/|\hat{\beta}_{nj}^{(k)}|$. 当 $\max\limits_{1\leqslant j\leqslant p_n}|\hat{\beta}_{nj}^{(k+1)}-\hat{\beta}_{nj}^{(k)}|\leqslant\tau$ 时迭代终止，本文中我们设定 $\tau=1.0\times10^{-4}$.

3.2　数值模拟

本节通过数值模拟验证所提估计的有限样本性质. 为了进行比较，在模拟中分别给出 SCAD、Lasso、Bridge 和 Oracle 4 种估计，这里 Oracle 估计是指事先已知道模型参数中哪些是零系数. 用 (C,I) 来评估变量选择的结果，其中 C 表示参数零系数被正确选择为零的平均个数，I 表示参数非零系数错误估计为零的平均个数. 同时，利用模型误差 $\mathrm{ME}(\hat{\boldsymbol{\beta}}_n)=E[\mu(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)-\mu(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n)]$ 来比较不同方法的表现.

考虑下面的广义线性模型:

$$y_i=g(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})+\sigma V^{1/2}(g(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}))\varepsilon_i$$

其中 $g(x)=2\cos x$，$V(x)=\exp(x)$，$\sigma=0.8$，$\boldsymbol{\beta}$ 是一个 p_n 维的向量，且 $\beta_j=j$，$1\leqslant j\leqslant4$，$\beta_j=0$，$5\leqslant j\leqslant p_n$. $\boldsymbol{X}_i$ 来自于 p_n 维的多元正态分布，均值为零，协方差矩阵的第 (i,j) 个元素为 $r^{|i-j|}$，其中 $r=0.2$ 或 0.8，$\varepsilon_i\sim U(-\sqrt{3},\sqrt{3})$. 考虑 3 种样本情况 $n=200$，$p_n=17$; $n=400$，$p_n=21$ 和 $n=600$，$p_n=24$.

模拟的结果列在表 1 和表 2 中. 由表 1 和表 2 可以得到: 1) 4 种方法下非零系数的估计都接近真值，SCAD 和 Bridge 方法得到的估计比较接近，均优于 Lasso，是具有 Oracle 性质的; 2) 当样本量增大时，SCAD 估计的 MRME 比 Bridge 小，但是均比 Lasso 小很多; 3) 4 种方法对于非零系数的选择都比较好，几乎没有出现错误选择的情况.

3.3　实例分析

这里讨论的是南非心脏病的数据[20]. 这组数据

表 1　非零系数的估计结果

Table 1　The estimate result of non-zero coefficient

(n,p_n)	r	方法	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
(200,17)	0.2	Lasso	0.987 8	1.983 2	2.984 6	3.971 7
		Bridge	0.995 5	1.997 1	3.001 7	4.001 1
		SCAD	0.995 4	2.001 7	2.999 7	3.999 2
		Oracle	1.000 2	1.999 9	3.002 9	3.999 5
	0.8	Lasso	0.998 7	1.985 2	2.988 7	3.993 2
		Bridge	0.994 6	2.007 6	2.995 4	4.001 0
		SCAD	1.000 1	2.001 6	2.997 7	4.000 3
		Oracle	1.004 0	2.000 5	2.994 1	3.999 7
(400,21)	0.2	Lasso	0.983 8	1.986 5	2.982 5	3.986 0
		Bridge	0.995 0	1.998 4	2.995 2	3.999 6
		SCAD	1.000 8	1.999 0	2.998 7	4.001 2
		Oracle	1.000 4	1.999 1	2.999 8	4.000 6
	0.8	Lasso	1.001 1	1.997 2	3.000 1	3.990 4
		Bridge	0.994 6	1.998 1	3.000 8	4.002 3
		SCAD	1.000 6	1.997 7	2.999 8	4.001 7
		Oracle	1.001 0	1.999 0	2.998 4	3.999 2
(600,24)	0.2	Lasso	0.981 5	1.990 1	2.988 6	3.985 0
		Bridge	0.994 9	1.997 5	2.997 5	3.998 1
		SCAD	0.999 5	2.000 1	2.999 9	4.001 5
		Oracle	1.000 5	2.000 9	3.001 8	3.999 3
	0.8	Lasso	0.997 0	2.001 2	2.998 7	3.994 7
		Bridge	0.997 3	2.002 6	2.998 8	3.999 2
		SCAD	0.997 8	2.002 1	3.000 1	3.998 2
		Oracle	0.999 9	1.999 8	2.999 6	3.998 9

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

来自南非西开普的三个农村地区. 研究的目的在于建立高发病率地区缺血性心脏病风险因素的强度. 这组数据包含了462个样本,其中有9个不同的变量和1个响应变量. 响应变量是用"presence"来表示的,它表示是否患有心脏病. 把响应变量0~1化,利用Logistic模型拟合这组数据. 在拟合中不仅采用了前述的3种方法,而且也通过普通极大似然估计(MLE)和最优子集方法对这组数据进行变量选择和参数估计. 这里Bridge的上指数取$\gamma=0.05$,调整参数λ_n由BIC准则进行选择. SCAD,Bridge,Lasso的λ_n分别为0.482 1,0.024 1,0.007 8. 估计结果及标准差在表3中给出.

表2 变量选择和模型比较模拟结果

Table 2 Simulation results of variable selection and model comparison

(n,p_n)	方法	$r=0.2$			$r=0.8$		
		MRME	C	I	MRME	C	I
(200,17)	Lasso	0.649 1	10.53 4	0	0.531 7	10.382	0
	Bridge	0.335 7	12.46	0	0.299 7	12.228	0
	Oracle	0.221 4	13	0	0.178 9	13	0
	SCAD	0.322 8	12.614	0	0.210 6	12.35	0
(400,21)	Lasso	0.533 6	14.388	0	0.673 9	13.638	0
	Bridge	0.245 2	16.732	0	0.284 6	16.692	0
	Oracle	0.175 8	17	0	0.186 8	17	0
	SCAD	0.208 7	16.896	0	0.212 8	16.776	0
(600,24)	Lasso	0.394 6	17.134	0	0.443 3	17.312	0
	Bridge	0.170 3	19.944	0	0.193 2	19.874	0
	Oracle	0.125 9	20	0	0.165 4	20	0
	SCAD	0.152 4	19.976	0	0.176 7	19.654	0

注: MRME这一列代表着ME与MEL比值的中位数,其中ME是选择模型的模型误差,MEL是在全模型下无惩罚函数普通极大似然估计的模型误差

表3 不同方法对南非心脏病数据分析的结果

Table 3 South African heart disease data analysis with different methods

变量	MLE		Lasso		Bridge		SCAD		最优子集	
	估计值	标准差	估计值	标准差	估计值	标准差	估计值	标准差	估计值	标准差
sbp	0.004 5	0.006 2	0	0	0	0	0	0	0	0
tob	0.076 4	0.028 4	0.079 0	0.010 0	0.076 9	0.010 6	0.079 5	0.001 0	0.089 9	0.028 6
ldl	0.205 4	0.068 3	0.170 5	0.024 3	0.161 1	0.026 9	0.163 5	0.024 7	0.183 4	0.058 7
adi	0.004 8	0.031 9	0.016 7	0.013 1	0	0	0	0	0	0
fam	0.939 5	0.249 9	0.928 0	0.096 0	0.873 3	0.092 0	0.912 6	0.089 6	0.931 4	0.246 7
typ	0.035 6	0.014 0	0.039 0	0.004 9	0.035 1	0.005 1	0.037 2	0.004 7	0.048 3	0.013 3
obe	−0.052 6	0.048 5	−0.058 0	0.02 0	0	0	0	0	0	0
alc	0.005 8	0.005 0	0	0	0	0	0	0	0	0
age	0.050 7	0.013 6	0.045 3	0.005 0	0.051 1	0.004 0	0.049 9	0.004 0	0.051 4	0.011 0

从表3中可以看到: 1) SCAD和Bridge成功实现了最简模型且接近于最优子集的选择结果,而SCAD的标准差略小于Bridge估计,Lasso选择的模型明显较复杂; 2) 用于分析数据的5种变量选择和估计方法得到的非零系数的估计也相差不大,且与文献[20]中的估计结果也是一致的.

4 结 论

本文主要研究在高维广义线性模型中,利用拟似然SCAD方法进行变量选择和参数估计的问题. 在不要求响应变量分布的情况下,给出了参数个数发散时,拟似然SCAD估计的渐近统计性质. 通过数值模拟和实例分析,得到了拟似然SCAD方法在参数估计和变量选择方面都具有良好的有限样本表现.

参考文献:

[1] NELDER J A, WEDDERBURN R W M. Generalized linear models [J]. *Journal of the Royal Statistical Society: Series A*, 1972, **135**(3): 370-384.

[2] WEDDERBURN R W M. Quasi-likelihood functions,

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

generalized linear models and Gauss-Newton method [J]. *Biometrika*, 1974, **61**(3): 439-447. DOI: 10.2307/2334725.

[3] CHIOU J M, MULLER H G. Nonparametric quasi-likelihood [J]. *The Annals of Statistics*, 1999, **27**(1): 36-64.

[4] CHEN X, CHEN X R. Adaptive quasi-likelihood estimate in generalized linear models [J]. *Science in China Series A: Mathematics*, 2005, **48**(6): 829-846.

[5] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its Oracle properties [J]. *Journal of the American Statistical Association*, 2001, **96**(456): 1348-1360.

[6] FRANK I E, FRIEDMAN J H. A Statistical view of some chemometrics regression tools [J]. *Technometrics*, 1993, **35**(2): 109-148.

[7] TIBSHIRANI R J. Regression shrinkage and selection via the Lasso [J]. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 1996, **58**(1): 267-288.

[8] ZOU H. The adaptive Lasso and its Oracle properties [J]. *Journal of the American Statistical Association*, 2006, **101**(476): 1418-1429.

[9] ZOU H, HASTIE T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 2005, **67**(2): 301-320.

[10] ZOU H, ZHANG H H. On the adaptive elastic-net with a diverging number of parameters [J]. *The Annals of Statistics*, 2009, **37**(4): 1733-1751. DOI: 10.1214/08-AOS625.

[11] YUAN M, LIN Y. Model selection and estimation in regression with grouped variables [J]. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 2006, **68**(1): 49-67.

[12] HUANG J, XIE H L. Asymptotic Oracle properties of SCAD-penalized least squares estimators [J]. *Lecture Notes—Monograph Series*, 2007, **55**: 149-166.

[13] KNIGHT K, FU W J. Asymptotics for Lasso-type estimators [J]. *The Annals of Statistics*, 2000, **28**(5): 1356-1378.

[14] HUANG J, HOROWITZ J, MA S. Asymptotic properties of Bridge estimators in sparse high-dimensional regression models [J]. *The Annals of Statistics*, 2008, **36**(2): 587-613. DOI: 10.1214/0090536070000 00875.

[15] WANG M Q, SONG L X, WANG X G. Bridge estimation for generalized linear models with a diverging number of parameters [J]. *Statistics & Probability Letters*, 2010, **80**(21): 1584-1596.

[16] XIE H L, HUANG J. SCAD-penalized regression in high-dimensional partially linear models [J]. *The Annals of Statistics*, 2009, **37**(2): 673-696. DOI: 10.1214/07-AOS580.

[17] FAN J Q, PENG H. Nonconcave penalized likelihood with a diverging number of parameters [J]. *The Annals of Statistics*, 2004, **32**(3): 928-961. DOI: 10.1214/009053604000000256.

[18] FAN J Q, LÜ J C. A Selective overview of variable selection in high-dimensional feature space [J]. *Statistica Sinica*, 2010, **20**(1): 101-148.

[19] WANG H S, LI R Z, TSAI C. Tuning parameter selectors for the smoothly clipped absolute deviation method [J]. *Biometrika*, 2007, **94**(3): 553-568. DOI: 10.1093/biomet/asm053.

[20] HASTIE T, TIBSHIRANI R, FRIEDMAN J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* [M]. New York: Springer-Verlag, 2001.

□

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net