DOI: 10.13718/j.cnki.xdzk.2017.09.013

Adaptive Elastic Net 方法在 Cox 模型变量选择中的研究①

韦新星[1]，　李春红[2]，　戴洪帅[3]

1. 河池学院 数学与统计学院，广西 宜州 546300；2. 广西大学 数学与信息科学学院，南宁 530004；
3. 山东财经大学 统计学院，济南 250014

摘要： 将 Adaptive Elastic Net 方法运用于 Cox 模型的变量选择中，证明了在一定条件下，Cox 模型的 Adaptive Elastic Net 估计具有组效应性质. 数值模拟和具体实例验证了该估计的组效应性质，表明 Cox 模型的 Adaptive Elastic Net 方法优于 Lasso 方法、Adaptive Lasso 方法和 Elastic Net 方法.

关　键　词： Adaptive Elastic Net 方法；Cox 模型；变量选择；组效应性质

中图分类号：O212.1　　**文献标志码：A**　　**文章编号：** 1673-9868(2017)09-0088-07

Cox 模型[1] 是处理生存数据的一种经典方法，常被广泛运用于医学、生物学、经济学、保险学等众多领域. 尽管 Cox 模型是目前为止最有用的生存分析方法，但它却要求自变量间相互独立，至少不存在强相关的情况. 此外，它还要求数据是大 n 小 p 类型. 于是，经典 Cox 模型在处理强相关及大 p 小 n 问题时，就不再适用了.

Lasso 方法为解决此类问题提供了新的思路. 1997 年 Tibshirani 成功将该方法应用于 Cox 模型[2]，进一步验证了它的实用性. 针对 Lasso 估计在某些情况下不相合的问题，Zou 于 2006 年提出了具有 Oracle 性质的 Adaptive Lasso 方法[3-5]，很好克服了 Lasso 的不足. 对于存在组效应的数据结构，Zou 和 Hastie 在 Lasso 的基础上，提出了 Elastic Net[6-8]，防止了模型的过于稀疏，且有效处理了大 p 小 n 问题. 同样，Elastic Net 估计也不具有 Oracle 性质[9-10]，而 Zou 和 Zhang 在 Elastic Net 的基础上，对 l_1 惩罚部分加权，提出了具有 Oracle 性质的 Adaptive Elastic Net 方法[11].

在 Cox 模型的诸多变量选择方法[7,12] 中，对于存在强相关性的变量，Elastic Net 方法较 Lasso 方法有更好的拟合效果和更高的预测能力，能将强相关变量全部选入或全部剔除模型. 但美中不足的是，在模型精确度方面，Elastic Net 方法对于零变量的估计却不及 Adaptive Elastic Net 方法. 为此，将 Adaptive Elastic Net 方法应用于 Cox 模型的变量选择中，研究在该模型下 Adaptive Elastic Net 方法的相关性质是一件有意义的工作.

1　Cox 模型 AEN 估计的定义

对于第 i 个个体，Cox 模型的表达式为：

---

① 收稿日期：2016-05-06
基金项目：国家自然科学基金项目(11361007，71671104)；教育部人文社科规划基金项目(16YJA910003)；广西自然科学基金(2014GXNSFCA118001)；河池学院硕士专业学位建设基金课题 2016YT004；山东省高等学校优势学科人才团队培育计划；山东财经大学优势学科人才团队孵化计划项目.
作者简介：韦新星(1990-)，女，广西柳州人，助教，主要从事应用统计研究.

$$h(t_i)=h_0(t_i)\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)$$

其中，样本容量为 n，预测变量个数为 p，协变量矩阵为 $\boldsymbol{X}=(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)$，$\boldsymbol{X}_i=(x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathrm{T}}$ 为第 i 个个体的 p 个协变量，回归向量为 $\boldsymbol{\beta}=(\beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}}$，$h_0(t_i)$ 为第 i 个个体的基准风险率，$i=1, 2, \cdots, n$.

现记观测数据为 $(Z_i, \delta_i, \boldsymbol{X}_i)$，$Z_i$ 为第 i 个个体的研究时间，令 $h_0(t)$ 恒定，则似然函数[13]为：

$$L(\boldsymbol{\beta})=\prod_{i=1}^{n}\left\{\frac{\exp(\sum_{k=1}^{p}\beta_k x_{ik})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}\right\}^{\delta_i}$$

其中：δ_i 为示性函数，事件删失时 $\delta_i=0$，事件发生时 $\delta_i=1$；R_i 为 t_i 时刻个体的风险集；$k=1,2,\cdots,p$.

于是，借鉴 Tibshirani[2] 及 Fan[10] 提出的处理思想，极小化偏对数似然函数的相反数并添上适当的惩罚项便可定义 Cox 模型的 Elastic Net 估计：

$$\hat{\beta}_{(EN)}=\operatorname{argmin}\left\{\frac{1}{n}\sum_{i=1}^{n}\{-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i+\ln[\sum_{j\in R_i}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_j)]\}+\lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|^2\right\}$$

进一步由

$$\|\boldsymbol{\beta}\|^2=\sum_{k=1}^{p}\beta_k^2,\quad \|\boldsymbol{\beta}\|_1=\sum_{k=1}^{p}|\beta_k|$$

可得

$$\hat{\beta}_{(EN)}=\operatorname{argmin}\left\{\frac{1}{n}\sum_{i=1}^{n}\{-\sum_{k=1}^{p}\beta_k x_{ik}+\ln[\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})]\}+\lambda_1\sum_{k=1}^{p}|\beta_k|+\lambda_2\sum_{k=1}^{p}\beta_k^2\right\} \tag{1}$$

其中 λ_1 和 λ_2 为调整参数，且满足 $\lambda_1\geqslant 0$，$\lambda_2\geqslant 0$.

进一步，借鉴普通线性模型中 Adaptive Elastic Net 估计的定义思想[7, 11]，在(1)式的基础上，对 l_1 惩罚部分加权，便可定义 Cox 模型的 Adaptive Elastic Net 估计：

$$\hat{\beta}_{(AEN)}=\operatorname{argmin}\left\{\frac{1}{n}\sum_{i=1}^{n}\{-\sum_{k=1}^{p}\beta_k x_{ik}+\ln[\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})]\}+\lambda_1^*\sum_{k=1}^{p}\hat{\omega}_k|\beta_k|+\lambda_2\sum_{k=1}^{p}\beta_k^2\right\} \tag{2}$$

其中：$\hat{\omega}_k=(|\hat{\beta}_{(EN)k}|)^{-\gamma}$，$\gamma$ 为一正常数.

2 Cox 模型 AEN 估计的性质

现研究 Cox 模型 Adaptive Elastic Net 估计的组效应性质.

定理 1 对 Cox 模型，给定数据 $(Z_i, \delta_i, \boldsymbol{X}_i)$ 及参数 (λ_1^*, λ_2)，响应变量已经中心化且自变量已经标准化. 令 $\boldsymbol{x}_a=(x_{1a}, x_{2a}, \cdots, x_{na})$ 为 n 个个体的第 a 个协变量，$\boldsymbol{x}_b=(x_{1b}, x_{2b}, \cdots, x_{nb})$ 为 n 个个体的第 b 个协变量，$a, b=1,2,\cdots,p$. $\hat{\beta}(\lambda_1^*, \lambda_2)$ 表示 AEN 估计，其中 $\hat{\beta}_a(\lambda_1^*, \lambda_2)$ 和 $\hat{\beta}_b(\lambda_1^*, \lambda_2)$ 是任意一组强相关变量 $\boldsymbol{x}_a$ 和 $\boldsymbol{x}_b$ 的系数. 假设 $\hat{\beta}_a(\lambda_1^*, \lambda_2)\hat{\beta}_b(\lambda_1^*, \lambda_2)>0$.

定义

$$D_{\lambda_1^*, \lambda_2}(a, b)=|\hat{\beta}_a(\lambda_1^*, \lambda_2)-\hat{\beta}_b(\lambda_1^*, \lambda_2)|$$

则

$$D_{\lambda_1^*, \lambda_2}(a, b)\rightarrow 0$$

证 由于 $\hat{\beta}_a(\lambda_1^*, \lambda_2)\hat{\beta}_b(\lambda_1^*, \lambda_2)>0$，故符号函数 $\operatorname{sgn}\{\hat{\beta}_a(\lambda_1^*, \lambda_2)\}=\operatorname{sgn}\{\hat{\beta}_b(\lambda_1^*, \lambda_2)\}$，且 $\hat{\beta}_a(\lambda_1^*, \lambda_2)\neq 0$，$\hat{\beta}_b(\lambda_1^*, \lambda_2)\neq 0$.

现令 $\hat{\beta}_m(\lambda_1^*,\lambda_2)\neq 0$，$\hat{\beta}(\lambda_1^*,\lambda_2)$ 满足

$$\frac{\partial L(\lambda_1^*,\lambda_2,\boldsymbol{\beta})}{\partial\beta_m}\Big|_{\boldsymbol{\beta}=\hat{\beta}(\lambda_1^*,\lambda_2)}=0$$

其中

$$L(\lambda_1^*,\lambda_2,\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\{-\sum_{k=1}^{p}\beta_k x_{ik}+\ln[\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})]\}+\lambda_1^*\sum_{k=1}^{p}\hat{\omega}_k|\beta_k|+\lambda_2\sum_{k=1}^{p}\beta_k^2$$

则由于 $\hat{\beta}_a(\lambda_1^*,\lambda_2)\neq 0$，故有

$$\frac{\partial L(\lambda_1^*,\lambda_2,\boldsymbol{\beta})}{\partial\beta_a}\Big|_{\boldsymbol{\beta}=\hat{\beta}(\lambda_1^*,\lambda_2)}=0$$

成立.

即

$$-\frac{1}{n}\sum_{i=1}^{n}x_{ia}+\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{ja}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}+\lambda_1^*\hat{\omega}_a\operatorname{sgn}\{\hat{\beta}_a(\lambda_1^*,\lambda_2)\}+2\lambda_2\hat{\beta}_a(\lambda_1^*,\lambda_2)=0$$

于是

$$\hat{\beta}_a(\lambda_1^*,\lambda_2)=\frac{1}{2\lambda_2}\left\{\frac{1}{n}\sum_{i=1}^{n}x_{ia}-\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{ja}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}-\lambda_1^*\hat{\omega}_a\operatorname{sgn}\{\hat{\beta}_a(\lambda_1^*,\lambda_2)\}\right\}\tag{3}$$

同理

$$\hat{\beta}_b(\lambda_1^*,\lambda_2)=\frac{1}{2\lambda_2}\left\{\frac{1}{n}\sum_{i=1}^{n}x_{ib}-\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in R_i}x_{jb}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}-\lambda_1^*\hat{\omega}_b\operatorname{sgn}\{\hat{\beta}_b(\lambda_1^*,\lambda_2)\}\right\}\tag{4}$$

将(3),(4) 式相减得到

$$\hat{\beta}_a(\lambda_1^*,\lambda_2)-\hat{\beta}_b(\lambda_1^*,\lambda_2)=\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}\left[x_{ia}-\frac{\sum_{j\in R_i}x_{ja}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}-x_{ib}+\frac{\sum_{j\in R_i}x_{jb}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}{\sum_{j\in R_i}\exp(\sum_{k=1}^{p}\beta_k x_{jk})}\right]+$$

$$\frac{\lambda_1^*}{2\lambda_2}\operatorname{sgn}\{\hat{\beta}_b(\lambda_1^*,\lambda_2)\}(\hat{\omega}_b-\hat{\omega}_a)\tag{5}$$

由于 Cox 模型的偏残差[14-15] 为

$$\hat{r}_{ir}=x_{ir}-E(x_{ir}\mid R_i)=x_{ir}-\frac{\sum_{l\in R_i}x_{lr}\exp(\boldsymbol{x}_l\boldsymbol{\beta})}{\sum_{l\in R_i}\exp(\boldsymbol{x}_l\boldsymbol{\beta})}\tag{6}$$

其中 $r=1,2,\cdots,p$，故(5) 式可变形为

$$\hat{\beta}_a(\lambda_1^*,\lambda_2)-\hat{\beta}_b(\lambda_1^*,\lambda_2)=\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}(\hat{r}_{ia}-\hat{r}_{ib})+\frac{\lambda_1^*}{2\lambda_2}\operatorname{sgn}\{\hat{\beta}_b(\lambda_1^*,\lambda_2)\}(\hat{\omega}_b-\hat{\omega}_a)$$

从而

$$|\hat{\beta}_a(\lambda_1^*,\lambda_2)-\hat{\beta}_b(\lambda_1^*,\lambda_2)|\leqslant\frac{1}{2\lambda_2 n}\sum_{i=1}^{n}|\hat{r}_{ia}-\hat{r}_{ib}|+\frac{\lambda_1^*}{2\lambda_2}|\hat{\omega}_b-\hat{\omega}_a|\tag{7}$$

于是，对于强相关变量 $\boldsymbol{x}_a$ 和 $\boldsymbol{x}_b$，由于 $\boldsymbol{x}_a$ 和 $\boldsymbol{x}_b$ 强相关，即 $E[\boldsymbol{x}_a\boldsymbol{x}_b^{\mathrm{T}}]\to 1$，故对第 i 个个体，有

$$|x_{ia}-x_{ib}|\to 0 \qquad |E(x_{ia})-E(x_{ib})|\to 0$$

从而

$$|[x_{ia}-E(x_{ia}\mid R_i)]-[x_{ib}-E(x_{ib}\mid R_i)]|\leqslant$$
$$|x_{ia}-x_{ib}|+|E(x_{ia}\mid R_i)-E(x_{ib}\mid R_i)|\to 0 \tag{8}$$

由(6),(8)式，有

$$|\hat{r}_{ia}-\hat{r}_{ib}|\to 0 \qquad |E(\hat{r}_{ia})-E(\hat{r}_{ib})|\to 0 \tag{9}$$

故

$$|\hat{\beta}_{(EN)_a}-\hat{\beta}_{(EN)_b}|=$$
$$\frac{1}{2\lambda_2}|[E(\hat{r}_{ia})-E(\hat{r}_{ib})]|\to 0$$

由

$$\hat{\omega}_k=(|\hat{\beta}_{(EN)_k}|)^{-\gamma}$$

可知

$$\hat{\omega}_a=(|\hat{\beta}_{(EN)_a}|)^{-\gamma} \qquad \hat{\omega}_b=(|\hat{\beta}_{(EN)_b}|)^{-\gamma}$$

于是

$$|\hat{\omega}_a-\hat{\omega}_b|\to 0 \tag{10}$$

由(7),(9),(10)式，得到

$$|\hat{\beta}_a(\lambda_1^*,\lambda_2)-\hat{\beta}_b(\lambda_1^*,\lambda_2)|\to 0$$

即

$$D_{\lambda_1^*,\lambda_2}(a,b)\to 0$$

证毕.

$D_{\lambda_1^*,\lambda_2}(a,b)$ 刻画了两个变量系数估计之间的差距，这表明若 $\boldsymbol{x}_a$ 和 $\boldsymbol{x}_b$ 高度相关，则对应的系数估计之间的差距将趋于0. 也就是说，Cox 模型的 AEN 估计具有组效应性质，即强相关变量得到的系数估计大致相同.

3　数值模拟

上节从理论上揭示了 Cox 模型 Adaptive Elastic Net 估计的组效应性质. 现通过数值模拟加以验证.

设 $x_i\sim N(0,1)$，$i=1,2,\cdots,10$，其中 $x_3=x_2$，$x_7=x_6$，$x_4=2x_1+\frac{1}{3}x_2+\frac{1}{3}x_3$. 则 x_3 与 x_2 强相关，x_7 与 x_6 强相关，且 x_4 与 x_1，x_2 及 x_3 之间存在共线性. 考虑 Cox 模型 $h(t)=h_0(t)\exp(\sum_{i=1}^{10}\beta_i x_i)$，$t\sim U[0,1]$，且真实参数为 $(-1,3,3,0,\frac{1}{2},2,2,0,0,0)^{\mathrm{T}}$，同时将该模型模拟1 000次，得到 $n=1\ 000$，$p=10$ 的样本数据.

分别运用 Lasso 方法、Adaptive Lasso(ALasso) 方法、Elastic Net(EN) 方法及 Adaptive Elastic Net(AEN) 方法对上述数据进行变量筛选[16-18]，其中后3种方法的系数估计值可先转化为 Lasso 方法的形式，再利用 Lars 算法[19] 得到. 取 $\lambda_2=\frac{1}{3}$，$\gamma=3$，而其他参数由交叉验证方法[20] 选出，重复计算50次，取系数估计值的平均值，得到的系数估计值见下表1.

表 1　数值模拟得到的系数估计值

	x_1	x_2	x_3	x_4	x_5
Lasso	−0.993 945	5.897 354	0.076 631	0	0.494 791
ALasso	−0.996 327	5.919 980	0.094 632	0	0.495 517
EN	−0.995 228	2.994 628	2.994 628	0	0.497 010
AEN	−0.999 330	2.999 382	2.999 382	0	0.499 934
真值	−1	3	3	0	0.500 000

	x_6	x_7	x_8	x_9	x_{10}
Lasso	3.873 642	0.083 943	0.005 477	0.003 621	0.004 615
ALasso	3.916 843	0.075 518	0	0.000 126	0
EN	1.993 058	1.993 058	−0.005 790	0.002 857	−0.004 665
AEN	1.999 576	1.999 576	0.000 161	0.000 177	0
真值	2	2	0	0	0

由表 1 可知：

1）对与 x_1,x_2,x_3 存在共线性的 x_4，4 种方法均没有将其选入模型，说明 Lasso 方法、Adaptive Lasso 方法、Elastic Net 方法及 Adaptive Elastic Net 方法均能处理共线性问题.

2）比较 Lasso 方法和 ALasso 方法：在对 x_8,x_9 及 x_{10} 这 3 个零变量的处理上，ALasso 方法比 Lasso 方法精确. 这体现了 Adaptive Lasso 方法在零变量的处理方面优于 Lasso 方法.

3）比较 EN 方法和 AEN 方法：在对变量 x_8,x_9 及 x_{10} 这 3 个零变量的处理上，AEN 方法比 EN 方法精确. 这体现了 Adaptive Elastic Net 方法在零变量的处理方面优于 Elastic Net 方法.

4）比较 ALasso 方法和 AEN 方法：在对 x_2 与 x_3，x_6 与 x_7 这两组强相关变量的处理上，AEN 方法能将强相关变量 x_2 与 x_3，x_6 与 x_7 全部选入模型，且这两组强相关变量的系数估计值相同，而 ALasso 方法只能选择强相关变量组中的一个变量. 这体现了 AEN 方法具有组效应性质.

4　实例分析

接下来，我们通过电信客户的实际数据来验证 Cox 模型 Adaptive Elastic Net 估计的优越性.

本实例来自对某高校在校大学生手机卡使用情况的调查. $x_1,x_2,\cdots,x_{10}$ 分别表示性别、年级、是否学生干部、是否少数民族、是否农业户口、是否生源地就读、是否移动用户、月均电话费用、售后服务质量、月均生活费用 10 个变量. 调查时间从 2007 年 1 月开始至 2014 年 1 月结束，最终得到 380 份有效问卷.

对数据进行简单统计分析后发现，大多数变量间存在较高的相关性，故经典 Cox 模型不再适用. 接下来，我们分别将 Lasso 方法、Adaptive Lasso(ALasso) 方法、Elastic Net(EN) 方法及 Adaptive Elastic Net(AEN) 方法运用于 Cox 模型中，得到的变量选择结果见表 2.

表 2　实例分析得到的系数估计值

	x_1	x_2	x_3	x_4	x_5
Lasso	0	−0.092	0.005	−0.011	0.125
ALasso	0	−0.042	0	0	0.096
EN	0	−0.112	0.004	−0.023	0.124
AEN	0	−0.120	0.002	0	0.125

	x_6	x_7	x_8	x_9	x_{10}
Lasso	−0.246	0	0	0.277	0.012
ALasso	0	0	0	0.018	0.047
EN	−0.284	0	0.019	0.313	0.026
AEN	−0.301	0	0.030	0.329	0.036

由表 2 可知：

1）4 种方法均没有将变量 x_1 和 x_7 选入模型，这说明将 Adaptive Elastic Net 方法运用于 Cox 模型是可行的.

2）对于具有较强相关性的 x_8 与 x_{10}，Lasso 和 ALasso 只选择了 x_{10}，而 EN 和 AEN 则把这两个强相关变量同时选入了模型，这表明 Cox 模型的 Elastic Net 方法和 Adaptive Elastic Net 方法能把强相关变量组中的变量全部选出；此外，在所有系数的估计值中，二者系数的差距最小. 这表明 Elastic Net 方法和 Adaptive Elastic Net 方法能体现变量间的相关性，且相关系数越大，它们系数估计的差距就越小，这体现了 Elastic Net 方法和 Adaptive Elastic Net 方法的组效应性质.

3）对与手机卡的流失无影响的 x_3 和 x_4，AEN 在对这两个零变量的处理上，比 EN 精确得多，这体现了 Cox 模型的 Adaptive Elastic Net 估计在零变量的处理方面优于 Elastic Net.

综上，Cox 模型的 Adaptive Elastic Net 估计优于其它 3 种估计.

5 结 论

Cox 模型的 Elastic Net 方法在处理具有强相关性的生存数据方面，优于 Cox 模型的 Lasso 方法. 但在模型精确度方面，Elastic Net 方法对零变量的估计却不太理想.

为克服这一缺陷，本文将 Adaptive Elastic Net 方法运用于 Cox 模型的变量选择中，证明了在一定条件下，Cox 模型的 Adaptive Elastic Net 估计具有组效应性质，即 Adaptive Elastic Net 方法能将强相关变量全部选入 Cox 模型. 此外，数值模拟和具体实例既验证了其组效应性质，也表明了 Cox 模型的 Adaptive Elastic Net 估计对零变量的处理更准确. 这表明，Cox 模型的 Adaptive Elastic Net 方法优于其他 3 种方法.

参考文献：

[1] COX D R. Regression Models and Life Tables [J]. Journal of Royal Statistical Society，1972(34)：187－220.

[2] TIBSHITANI R. The Lasso Method for Variable Selection in the Cox Model [J]. Statistics in Medicine，1997，16(4)：385－395.

[3] ZHANG H H，LU W. Adaptive Lasso for Cox's Proportional Hazards Model [J]. Biometrika，2007，94(3)：691－703.

[4] ZOU H. The Adaptive Lasso and Its Oracle Properties [J]. Journal of the American Statistical Association，2006，101(476)：1418－1429.

[5] HUANG J，MA S G，Zhang C H. Adaptive Lasso for Sparse High-Dimensional Regression Models [J]. Statistica Sinica，2008，18(4)：1603－1618.

[6] ZOU H，HASTIE T. Regularization and Variable Selection via the Elastic Net [J]. Journal of the Royal Statistical Society，Series B，2005，67(1)：301－320.

[7] 卢 颖. 广义线性模型基于 Elastic Net 的变量选择方法研究 [D]. 北京：北京交通大学，2011.

[8] 闫丽娜. 惩罚 Cox 模型和弹性网技术在高维数据生存分析中的应用 [D]. 太原：山西医科大学，2011.

[9] FAN J，LI R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties [J]. Journal of the American Statistical Association，2001，96(456)：1348－1360.

[10] FAN J，LI R. Variable Selection for Cox's Proportional Hazards Model and Frailty Model [J]. Annals of Statistics，2002，30(1)：74－99.

[11] ZOU H，ZHANG H H. On the Adaptive Elastic Net with a Diverging Number of Parameters [J]. Annals of Statistics，2009，37(4)：1733－1751.

[12] 毕伯竹. 高维多重共线性数据的变量选择问题 [D]. 济南：山东大学，2011.

[13] 王启华. 生存数据统计分析 [M]. 北京：科学出版社，2006：232－237.

[14] 郜艳晖，何大卫. Cox 模型的残差分析和影响诊断 [J]. 现代预防医学，2000，27(1)：48－50.

[15] SCHOENFELD D. Partial Residuals for the Proportional Hazards Regression Model [J]. Biometrika，1982，69(1)：

239－241.

［16］吴喜之. 复杂数据统计方法——基于 R 的应用［M］. 北京：中国人民大学出版社，2012.

［17］王斌会. 多元统计分析及 R 语言建模［M］. 广州：暨南大学出版社，2010.

［18］董　英，黄品贤. Cox 模型及预测列线图在 R 软件中的实现［J］. 数理医药学杂志，2012，25(6)：711－713.

［19］EFRON B，HASTIE T，JOHNSTONE I，et al. Least Angle Regression［J］. Technical Report，Stanford University，2004，32(2)：407－451.

［20］VERWEIJ P J. Cross-Validation in Survival Analysis［J］. Statist Med，1993，12(24)：2305－2314.

The Study of the Adaptive Elastic Net Method in the Variable Selection of the Cox Model

WEI Xin-xing[1]，　LI Chun-hong[2]，　DAI Hong-shuai[3]

1. School of Mathematics and Statistics，Hechi University，Yizhou Guangxi 546300，China；

2. School of Mathematics and Information Science，Guangxi University，Nanning 530004，China；

3. School of Statistics，Shandong University of Finance and Economics，Jinan 250014，China

Abstract：In this paper，we study the adaptive elastic net method applied in variable selection of the Cox model. We prove the grouping effect property of its estimators under certain conditions. Finally，we show the grouping effect property by a numerical simulation and a real case，demonstrating that for the Cox model，the adaptive elastic net method performs better than the Lasso method，the adaptive Lasso method and the elastic net method.

Key words：adaptive elastic net method；Cox model；variable selection；grouping effect property

责任编辑　张　构