DOI:10.3969/j.issn.1001-5337.2024.3.041

基于 DP 算法的 Poisson 回归模型的变量选择

王秀丽, 姜 喆

(曲阜师范大学统计与数据科学学院,273165,山东省曲阜市)

摘要:利用已有的 DP 算法,对 Poisson 回归模型的似然函数进行优化估计出模型的参数,分别写出 Lasso、Adaptive Lasso、SCAD、MCP 4 种惩罚似然函数,用 DP 算法进行求解,实现变量选择.为了实现算法的自动变量选择,在对惩罚参数的选择上,使用了 AIC、BIC 信息准则.为了验证 DP 算法的可行性,通过随机模拟的方式生成数据进行求解.在各项指标下,将 DP 算法与已有的相关算法进行对比.

关键词:Poisson 回归模型;变量选择;DP 算法;BIC;AIC

中图分类号:O211.6 **文献标识码**:A **文章编号**:1001-5337(2024)03-0041-08

0 引 言

EM 算法是 1997 年 Dempster 等[1]为了解决缺失参数或者数据截尾问题首次提出的,1983 年 Wu[2]证明了 EM 算法的收敛性质.而 DP 算法是由 De Pierro[3]1995 年在 EM 算法的基础上进行改进提出的,并对算法的收敛性进行证明.EM 算法的限制是不能解决缺少潜变量的情况,DP 算法的改进使得在最大化似然函数的过程中不需要引入潜变量,主要包括转换步(T 步)和求目标函数的极大值点(M 步).DP 算法的另一个优势是可以把高维目标函数转化为多个一维目标函数依次迭代求极值,使得计算上的难度大大降低.

传统的 Poisson 回归模型的估计参数是对似然函数通过 Newton-Raphson 算法或者利用梯度下降算法进行模型的参数估计.DP 算法是一种基于坐标下降的算法,可以用来迭代求解 Poisson 模型的参数.1996 年 Tibshirani[4]在岭回归的基础上提出了 Lasso 方法进行选择参数,开辟了压缩系数变量选择的先河.之后的变量选择方法在 Lasso 的基础上提出改进,本文主要讨论 4 种惩罚函数:Lasso、Adaptive Lasso[5]、SCAD[6]和 M CP[7].随着高维数据的兴起,针对 Poisson 回归模型的变量选择提出了不少有效的算法.Hunter 和 Li[8]在 2005 年提出了 MM 算法来求解惩罚似然函数的极大值点.Zou 和 Li[9]在 2008 年使用 LLA 算法求解惩罚的似然函数.杨成敏[10]在 2015 年以 Poisson 回归模型为例对 3 种惩罚函数 Lasso、SCAD、MCP 进行蒙特卡洛模拟,得到了很好的实验效果.

本文的第 1 部分以 Poisson 回归模型为例,根据极大似然的思想写出似然函数和相应的惩罚似然函数,推导出 DP 算法在 Poisson 回归模型下的具体迭代形式.第 2 部分给出数值模拟,验证 DP 算法的可行性与准确性,借鉴文献[11]中的评价指标来评估不同方法与算法的性能.此外还调用 R 语言的 glm 函数以及 ncvreg 函数包,计算各项指标与 DP 算法的结果进行对比.第 3 部分给出总结.

1 模型介绍与算法推导

1.1 模型介绍

考虑线性泊松(Poisson)回归模型 $E(Y\mid \boldsymbol{x})=\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}}$,其中 Y 为响应变量,$\boldsymbol{x}$ 是 $q+1$ 维协变量,其中第 1 个分量为 1,$\boldsymbol{\beta}$ 是未知参数(含常数项).

收稿日期:2022-10-17

基金项目:山东省自然科学基金(ZR2020QA021).

通信作者:王秀丽,女,1985-,博士,副教授,硕士生导师;研究方向:大数据抽样;E-mail:stat_sci@126.com.

在估计参数 $\boldsymbol{\beta}$ 时,可以使用极大似然估计的方法来估计.虽然泊松回归模型的似然函数没有办法直接求出显示表达式来获得解析解,但是可以利用该模型似然函数的凹性(二阶导数小于 0),通过 Newto-Raphson 算法或者其他的梯度下降对模型的参数进行求解.

1.2 目标函数的构造

对 Poisson 回归模型,Y 的密度函数为

$$p(y \mid \boldsymbol{x};\boldsymbol{\beta})=\frac{[E(Y \mid \boldsymbol{x})]^{y}\mathrm{e}^{-E(Y|\boldsymbol{x})}}{y!}=\frac{\mathrm{e}^{y\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}}\mathrm{e}^{-\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}}}}{y!}.$$

在 Poisson 总体中,抽样出 n 组观测样本.当观测数据已知时,给出联合概率

$$p(y_1\cdots y_n,\boldsymbol{x}_1\cdots\boldsymbol{x}_n;\boldsymbol{\beta})=\prod_{i=1}^{n}\frac{\mathrm{e}^{y_i\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i}\mathrm{e}^{-\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i}}}{y_i!}.$$

相应的对数似然函数为 $L(\boldsymbol{\beta})=\sum_{i=1}^{n}[y_i\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i-\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i}-\ln(y_i!)]$,进一步简化为

$$L(\boldsymbol{\beta})=\sum_{i=1}^{n}(y_i\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i-\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i}).\tag{1}$$

令 $f_i(\boldsymbol{x})=y_i\boldsymbol{x}-\mathrm{e}^{x}$,则式(1)可化为 $L(\boldsymbol{\beta})=\sum_{i=1}^{n}f_i(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i)$.

1.3 DP 算法介绍

当模型出现多余参数或者数据截尾时,估计参数将面临巨大挑战,Dempster 等[1] 首次提出 EM 算法,也称作最大期望值算法.De Pierro[3] 提出了修正的EM算法.该算法将高维度的优化问题转换为一元函数的优化问题,此方法的优势为降低计算机的计算复杂度且有解、迭代过程收敛、在迭代过程中不需要其他维度的参数更新.本文称此方法为 DP 算法.

在推导泊松模型的算法之前,需要构建非负权重矩阵,根据 Becker 等[12] 的建议,定义 $\lambda_{ij}=\frac{|x_{ij}|}{\sum_{j=1}^{q+1}|x_{ij}|}$,显然,这里对于任意固定的 i,有 $\sum_{j=1}^{q+1}\lambda_{ij}=1$.

DP算法主要分 2 步.第 1 步,根据 DP算法中的 T 步转换似然函数,带入泊松回归模型的似然函数,从而定义出一个替代函数,

$$\begin{aligned}Q(\boldsymbol{\beta}\mid\boldsymbol{\beta}^{(t)})&=\sum_{i=1}^{n}\sum_{j=1}^{q+1}\lambda_{ij}f_i(\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})\\&=\sum_{i=1}^{n}\sum_{j=1}^{q+1}\lambda_{ij}[y_i(\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})-\exp(\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})].\end{aligned}\tag{2}$$

第 2 步,DP算法的 M 步,求解极大值点推导迭代公式.可以通过最大化替代函数(2)来进行迭代求解,即 $\boldsymbol{\beta}^{(t+1)}=\operatorname*{argmax}_{\boldsymbol{\beta}}Q(\boldsymbol{\beta}\mid\boldsymbol{\beta}^{(t)})$,但是,此时带入泊松回归模型的似然函数很难求得显式解,所以,可以对固定的 j,解下面一元方程求出极大值点进行迭代,

$$\begin{aligned}&\sum_{i=1}^{n}\nabla f_i(\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})x_{ij}\\=&\sum_{i=1}^{n}x_{ij}[(y_i-\exp(\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})x_{ij})]=0.\end{aligned}$$

由于一元函数极值点问题的牛顿法具有很好的估计性质.本文则直接使用牛顿法来近似(2)式中每一个 j 的解,得到下面近似迭代公式,

$$\beta_j^{(t+1)}=\beta_j^{(t)}+\frac{\sum_{i=1}^{n}\nabla f_i(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})x_{ij}}{\sum_{i=1}^{n}\{-\nabla^2 f_i(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})x_{ij}^2\}/\lambda_{ij}}$$

$$=\beta_j^{(t)}+\frac{\sum_{i=1}^{n}\{y_i-\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})\}x_{ij}}{\sum_{i=1}^{n}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})x_{ij}^2/\lambda_{ij}}. \tag{3}$$

下面总结上述 DP 算法的主要过程.

第 1 步：验证算法的条件，所要优化求解的似然函数需要满足严格连续二阶可导凹函数的条件. 该模型的似然函数公式(1) 显然满足该条件.

第 2 步：DP 算法的 T 步，转化似然函数写出泊松回归模型的替代函数，如公式(2).

第 3 步：DP 算法的 M 步，利用一维牛顿的算法计算极大值点更新每一维度的参数，如公式(3).

第 4 步：给定 DP 算法迭代停止条件，给定误差 ε（本文取为10^{-6}），当 $\|\beta^{(t+1)}-\beta^{(t)}\|_1\leqslant\varepsilon$ 时，算法停止迭代.

1.4　变量选择算法介绍

变量选择是统计模型的一个重要任务，传统的变量选择模型有向前、向后回归和所有子集法，但是都存在很大的局限性. 在岭回归的基础上提出的通过压缩系数来进行变量选择的方法在 1996 年被提出来. 具有代表性的 4 种惩罚函数是 Lasso、Adaptive Lasso、SCAD 和 MCP. 本节分别推导 4 种惩罚函数下 DP 算法的迭代形式.

因为在惩罚变量时不对常数项进行惩罚，基于上述算法对优化函数(1) 去除截距项后加入不同的惩罚项，再通过 DP 算法迭代以达到变量选择的目的，得到如下的目标函数

$$L^P(\boldsymbol{\beta})=\sum_{i=1}^{n}(y_i\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i-\mathrm{e}^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i})-n\sum_{j=1}^{q}p_\lambda(|\beta_j|).$$

下面分别是DP算法的T步和M步，根据惩罚似然写出替代函数. 为了简化替代函数的公式表达，令 $z=\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}$，则得到如下替代函数

$$\begin{aligned}Q^P(\boldsymbol{\beta}\mid\boldsymbol{\beta}^{(t)})&=\sum_{i=1}^{n}\sum_{j=1}^{q}[\lambda_{ij}f_i(z)-p_\lambda(|\beta_j|)]=\sum_{i=1}^{n}\sum_{j=1}^{q}\{\lambda_{ij}[y_iz-\mathrm{e}^z]-p_\lambda(|\beta_j|)\}\\&=\sum_{i=1}^{n}\sum_{j=1}^{q}\left\{\lambda_{ij}\left[y_iz-\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}(1+\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})+\frac{[\lambda_{ij}^{-1}x_{ij}(\beta_j-\beta_j^{(t)})]^2}{2})\right]-p_\lambda(|\beta_j|)\right\}.\end{aligned}\tag{4}$$

下面分别对 4 种惩罚函数求极大值点，写出迭代公式.

(1) 当惩罚函数为 $p_\lambda(|\beta_j|)=\lambda|\beta_j|$ 时，称之为 Lasso. 将惩罚函数代入式(4)，并关于变量 β_j 求极大值点更新 $\beta_j^{(t+1)}$，迭代关系如下

$$\beta_j^{(t+1)}=\begin{cases}\dfrac{t_j-n\lambda}{\sum_{i=1}^{n}\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}x_{ij}^2/\lambda_{ij}}, & t_j>n\lambda,\\ 0, & |t_j|\leq n\lambda,\\ \dfrac{t_j+n\lambda}{\sum_{i=1}^{n}\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}x_{ij}^2/\lambda_{ij}}, & t_j<-n\lambda,\end{cases}$$

其中 $t_j=\sum_{i=1}^{n}x_{ij}(y_i-\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}+\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}\lambda_{ij}^{-1}x_{ij}\beta_j^{(t)})$. 为了进一步简化结果的表达，定义

$$F(u,v)=\begin{cases}u-v, u>v,\\ 0, \quad |u|\leq v,\\ u+v, u<-v.\end{cases}\qquad h_j=\sum_{i=1}^{n}\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)}}x_{ij}^2/\lambda_{ij},$$

则可以整理得到 $\beta_j^{(t+1)}=\dfrac{F(t_j,n\lambda)}{h_j}$.

(2) 当惩罚函数为 $p_\lambda(|\beta_j|)=\lambda|\tilde{\beta}_j|^{-\eta}|\beta_j|$ 时，称之为 Adaptive Lasso，其中 $\tilde{\boldsymbol{\beta}}=(\widetilde{\beta_1},\cdots,\widetilde{\beta_p})^{\mathrm{T}}$ 为参数 $\boldsymbol{\beta}$ 的初始估计. 一般取 $\eta=1$，则有

$$\beta_j^{(t+1)}=\frac{F(t_j,n\lambda|\widetilde{\beta_j}|^{-\eta})}{h_j}.$$

(3) 当惩罚函数为

$$p_\lambda(|\beta_j|,a)=\lambda|\beta_j|I(|\beta_j|<\lambda)+\{(a^2-1)\lambda^2-[a\lambda-|\beta_j|]_+^2\}I(\lambda\leqslant|\beta_j|)/[2(a-1)]$$

时,称之为 SCAD. 根据文献[6],取 $a=3.7$,则有

$$\beta_j^{(t+1)}=\begin{cases}\dfrac{F(t_j,n\lambda)}{h_j}, & |t_j|<n\lambda(1+h_j),\\ \dfrac{F(t_j,n\lambda a)/(a-1)}{h_j-1/(a-1)}, & n\lambda(1+h_j)\leqslant|t_j|\leqslant an\lambda h_j,\\ \dfrac{t_j}{h_j}, & |t_j|>an\lambda h_j.\end{cases}$$

(4) 当惩罚函数为

$$p_\lambda(|\beta_j|,\gamma)=[\lambda|\beta_j|-\beta_j^2/(2\gamma)]I(|\beta_j|<\gamma\lambda)+\gamma\lambda^2/2I(|\beta_j|\geq\gamma\lambda)$$

时,称之为 MCP. 根据文献[7],取 $\gamma=3$,则有

$$\beta_j^{(t+1)}=\begin{cases}\dfrac{F(t_j,n\lambda)}{h_j-1/\gamma}, & |t_j|\leqslant n\lambda\gamma h_j,\\ \dfrac{t_j}{h_j}, & |t_j|>n\lambda\gamma h_j.\end{cases}$$

1.5 调整参数的选择标准

惩罚参数 λ 的选择对惩罚变量选择方法起着至关重要的作用. Schwarz[13] 在 1978 年提出的贝叶斯信息准则,简称 BIC 准则. 相比较与 AIC 准则,BIC 准则加大了变量个数的惩罚. 下面给出 2 种准则的定义式,其中 AIC 的定义为 $\mathrm{AIC}=2k-2\ln(L)$. 在 AIC 中,模型中变量的个数惩罚权重较小,BIC 信息准则加大了对参数个数的惩罚. BIC 的定义为 $\mathrm{BIC}=k\ln(n)-2\ln(L)$,其中 k 是最终估计出的模型中非零参数的个数,L 是似然函数,n 是样本的观测数量. 在惩罚似然函数的变量选择中,结合模型分别得到下面准则

$$\mathrm{AIC}_\lambda=2k_\lambda-2\ln(L),\ \mathrm{BIC}_\lambda=k_\lambda\ln(n)-2\ln(L),$$

其中 k_λ 是在固定 λ 时最终估计出的模型中非零参数的个数,L 是公式(1) 中的似然函数. 遍历所有大于零的 λ 的值可求得合适的 λ,进而得到模型中参数估计,实现选择变量.

2 数值模拟

例 1 响应变量服从的分布为

$$Y\sim\mathrm{Pois}(\lambda(\boldsymbol{x})),\lambda(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta},$$

其中 $\boldsymbol{x}=(x_1,x_2,\ldots,x_{15})^{\mathrm{T}},x_j\sim N(0,0.01),j=1,2,\cdots,15$. 并且 15 个变量之间相互独立,样本量 $n=2\ 000$. 设置如下真实的稀疏参数

$$\beta_1=\beta_2=\beta_3=\beta_4=1,\beta_4=\beta_5=\cdots=\beta_{15}=0.$$

为了检验 DP 算法的可行性与模型性能,下面设置要做的任务.

任务 1:使用 DP 算法估计上述模型的参数,并与内置的 glm 函数比较.

任务 2:使用 DP 算法用 Lasso、Adaptive Lasso、SCAD 和 MCP 4 种惩罚函数,分别做出画出参数 β 随着 λ 变化的参数估计轨迹图. 用 AIC、BIC 准则选择调整参数,实现自动变量选择.

任务 3:调用 R 语言内置的 ncvreg 包,分别用 Lasso、SCAD 和 MCP 3 种惩罚函数利用一次实验数据进行变量选择,对比分析所得结果.

对于此例题的任务 1,首先使用 DP 算法进行估计参数,然后使用 R 语言内置的 MASS 包中的 glm 函数估计参数,整理出如下结果,结果保留 3 位小数(四舍五入).

表 1　DP 算法与内置函数的参数结果

算法 / 系数	截距项	β_1	β_2	β_3	β_4	β_5
DP 算法	−0.062	0.709	1.132	0.683	0.958	0.207
内置函数	−0.062	0.709	1.132	0.683	0.958	0.207
算法 / 系数		β_6	β_7	β_8	β_9	β_{10}
DP 算法		0.074	0.061	0.596	0.154	−0.113
内置函数		0.074	0.061	0.596	0.154	−0.113
算法 / 系数		β_{11}	β_{12}	β_{13}	β_{14}	β_{15}
DP 算法		−0.188	−0.643	−0.207	0.417	−0.227
内置函数		−0.188	−0.643	−0.207	0.417	−0.227

DP 算法迭代次数:35

由表 1 可以看出,DP 算法在泊松模型中估计参数的时候与内置的 glm 函数一致,可以完成似然函数求极值并且达到估计参数的目的.

对于此例题的任务 2,本文对 λ 从固定区间内取 200 个点,对不同的惩罚函数的估计,分别记录每个 λ 对应的系数,画出变量参数轨迹图 1.

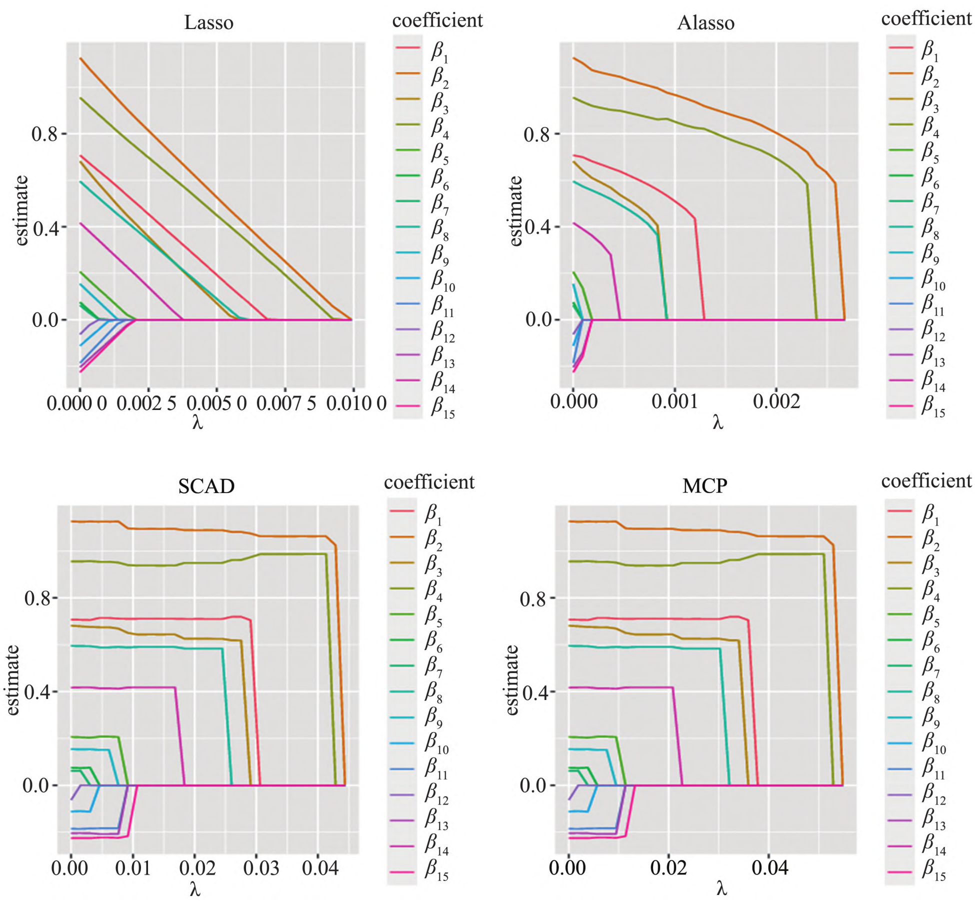

图 1　4 种方法的变量参数轨迹图对比

从图 1 中可以看出,随着 λ 的增加系数都趋于零,并且除了与响应变量有关的 β_1、β_2、β_3 和 β_4 外,其他变量被快速压缩为零,能看出 DP 算法可以处理带有惩罚项的似然函数的极值问题,很好地达到变量选择的效果. 为了能够自动选择模型变量,在模型中引入 AIC、BIC 的信息准则. 这里给出 Lasso、SCAD 和 MCP 3 种方

法的惩罚结果.

表 2 自动选择模型参数

参数	真实参数	AIC 信息准则			BIC 信息准则		
		Lasso	SCAD	MCP	Lasso	SCAD	MCP
常数	0	−0.05	−0.06	−0.06	−0.05	−0.06	−0.06
β_1	1	0.50	0.71	0.71	0.50	0.71	0.71
β_2	1	0.86	1.10	1.10	0.86	1.10	1.09
β_3	1	0.41	0.62	0.48	0.41	0.30	0.26
β_4	1	0.74	0.95	0.95	0.74	0.96	0.97
β_5	0	0	0	0	0	0	0
β_6	0	0	0	0	0	0	0
β_7	0	0	0	0	0	0	0
β_8	0	0.39	0.59	0.59	0.39	0.34	0.32
β_9	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0
β_{11}	0	0	0	0	0	0	0
β_{12}	0	0	0	0	0	0	0
β_{13}	0	0	−0.01	−0.01	0	0.07	0
β_{14}	0	0	0.21	0.14	0.19	0	0
β_{15}	0	0.19	−0.00	0	0	0	0
迭代次数		30	35	34	30	34	34

表 3 ncvreg 包的变量选择

参数/方法	真实参数	Lasso	SCAD	MCP
常数	0	−0.05	−0.06	−0.06
β_1	1	0.49	0.58	0.66
β_2	1	0.85	1.10	1.10
β_3	1	0.40	0.43	0.53
β_4	1	0.74	0.96	0.95
β_5	0	0	0	0
β_6	0	0	0	0
β_7	0	0	0	0
β_8	0	0.37	0.37	0.48
β_9	0	0	0	0
β_{10}	0	0	0	0
β_{11}	0	0	0	0
β_{12}	0	0	0	0
β_{13}	0	0	0	0
β_{14}	0	0.18	0.16	0.20
β_{15}	0	0	0	0

由表 2 可以看出,运用 DP 算法和信息准则的方式去选择变量可以选出全部的真实变量,且与表 3 中内置的 ncvreg 包的结果相符合. 甚至在某些情况下要优于内置的 ncvreg 包,但只是在一次的数值模拟所得到的结果.

在例 2 中将设置 5 个指标,并随机的进行 100 次模拟. 对 DP 算法与内置的 ncvreg 包的结果比较.

例 2 为了更容易看出不同方法的效果,定义 5 种模型评价指标如表 4.

表 4 5 种模型评价指标

符号	含义
PE	泊松回归的预测误差
loss1	估计参数与真实参数的 1 范数
loss2	估计参数与真实参数的 2 范数
FN	最终模型参数的个数(不含截距项)
MRN	遗漏模型真实参数的个数

表 4 是对定义的 5 种指标的解释,其中 PE、loss1 和 loss2 的计算公式为

$$\mathrm{PE}=E(Y-\mathrm{e}^{x^{\mathrm{T}}\boldsymbol{\beta}})^2,\ \mathrm{loss1}=\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|,\ \mathrm{loss2}=\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|_2.$$

分别用 DP 算法与内置的 ncvreg 包对 Lasso、SCAD 和 MCP 3 种惩罚函数进行 100 数值模拟并计算 3 种惩罚函数的各项指标. 为了直观的表达结果,分别对实验结果可视化,画出箱线图 2.

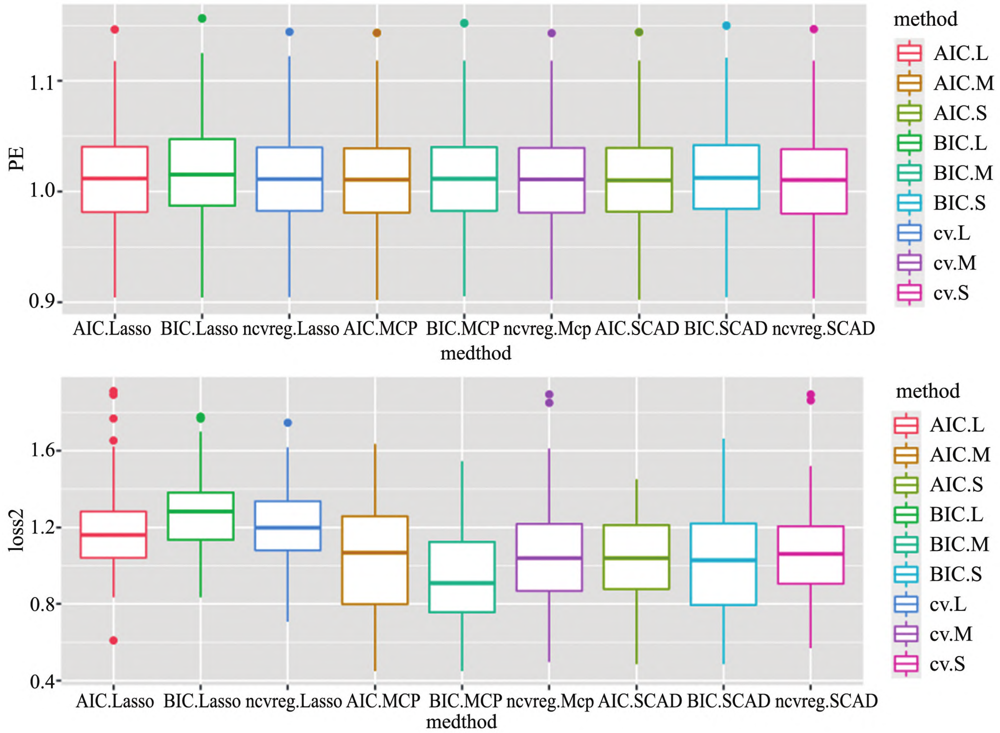

图 2　100 次模拟的 PE 和 loss2

图 2 展示了 DP 算法的 AIC、BIC 准则以及内置的 ncvreg 包下泊松模型的 PE 和 loss2 2 种指标. 从 PE 结果来看,3 种方法的差距不大. 从 loss2 结果来看,对 Lasso 方法,DP 算法的 AIC 准则损失最小;对 SCAD 方法,结果基本一致;对 MCP 方法,DP 算法的 BIC 准则损失明显小于其他 2 种.

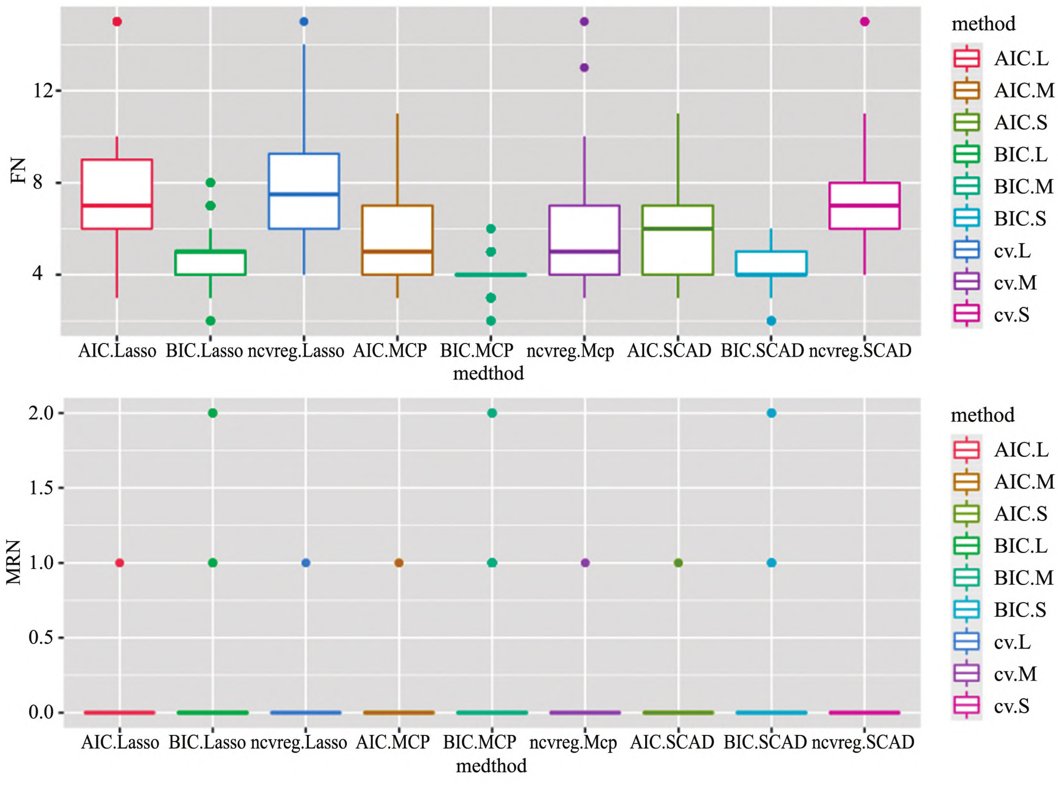

图 3　100 次模拟结果的最终参数个数和错过的真实参数的个数

图 3 展示了 DP 算法的 AIC、BIC 准则以及内置的 ncvreg 包得到的 FN 和 MRN 2 个指标的结果. 从 MRN 来看,所有估计方法及惩罚函数基本没有遗漏真实参数. 但对 FN 来说,DP 算法的 BIC 准则要明显得一致优于 DP 算法的 AIC 准则和内置的 ncvreg 包,更加接近真实值.

3 结 论

DP 算法可以实现凹函数优化问题,对似然函数进行求解,并且误差可以控制的比较小. 但是由于初始参数的设置,迭代次数比较多,有进一步优化的空间. 本文将 DP 算法应用到泊松回归模型的似然函数求解中,并与内置的 ncvreg 包进行对比,在某些方面使用 DP 算法自编的函数要优于内置的 ncvreg 包中的函数.

参考文献:

[1] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. J R Stat Soc B, 1977, 39(1):1-22.

[2] WU C J. On the convergence properties of the EM algorithm[J]. Ann Stat, 1983, 11(1):95-103.

[3] DE PIERRO A R. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography[J]. IEEE T Med Imaging, 1995, 14(1):132-137.

[4] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. J R Stat Soc B, 1996, 58(1):267-288.

[5] ZOU H. The adaptive lasso and its oracle properties[J]. J Am Stat Assoc, 2006, 101(476):1418-1429.

[6] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. J Am Stat Assoc, 2001, 96(456):1348-1360.

[7] ZHANG C H. Nearly unbiased variable selection under minimax concave penalty[J]. Ann Stat, 2010, 38(2):894-942.

[8] HUNTER D R, LI R. Variable selection using MM algorithms[J]. Ann Stat, 2005, 33(4):1617-1642.

[9] ZOU H, LI R. One-step sparse estimates in nonconcave penalized likelihood models[J]. Ann Stat, 2008, 36(4):1509-1533.

[10] 杨成敏. 广义线性模型中的参数估计及变量选择方法研究[D]. 重庆:重庆大学, 2015.

[11] FAN J, LV J. Nonconcave penalized likelihood with NP-dimensionality[J]. IEEE T Inform Theory, 2011, 57(8):5467-5484.

[12] BECKER M P, YANG I, LANGE K. EM algorithms without missing data[J]. Stat Methods Med Res, 1997, 6(1):38-54.

[13] SCHWARZ G. Estimating the dimension of a model[J]. Ann Stat, 1978, 6(2):461-464.

Variable selection of Poisson regression models based on DP algorithm

WANG Xiuli, *JIANG Zhe*

(School of Statistics and Data Science, Qufu Normal University, 273165, Qufu, Shandong, PRC)

Abstract: Using the existed DP algorithm, we optimize the likelihood function of the Poisson regression model to estimate the parameters of the model. According to the four penalty likelihood functions of the Lasso, Adaptive Lasso, SCAD and MCP, we use the DP algorithm to solve the optimization. AIC and BIC information criteria are used to select the penalty parameters. Simulation studies are conducted to verify the feasibility of the DP algorithm. Under Various indicators, the DP algorithm is compared with the existed related algorithms.

Key words: Poisson regression model; variable selection; DP algorithm; BIC; AIC