doi:10.3969/j.issn.1673-9833.2019.01.009

拟似然非线性模型的统计诊断

宋　蕾，冯　予

（南京理工大学 理学院，江苏 南京 210094）

摘　要：研究了经验似然估计下的拟似然非线性模型的统计诊断问题。首先给出了模型经验似然比函数，进而求出模型的经验似然估计；再基于数据删除模型推导出参数的一阶近似公式，并提出了经验 Cook 距离；最后通过对实例的分析，验证了该统计方法的有效性和合理性。

关键词：拟似然非线性模型；经验似然；数据删除模型；经验 Cook 距离；强影响点

中图分类号：O212　　**文献标志码**：A　　**文章编号**：1673-9833(2019)01-0054-06

引文格式：宋　蕾，冯　予 . 拟似然非线性模型的统计诊断 [J]. 湖南工业大学学报，2019，33(1)：54-59.

A Statistical Diagnosis for Quasi-Likelihood Nonlinear Models

SONG Lei，FENG Yu

（School of Science，Nanjing University of Science and Technology，Nanjing 210094，China）

Abstract：A research has been conducted on the statistical diagnosis of quasi-likelihood nonlinear models based on empirical likelihood estimation. Firstly, the empirical likelihood ratio function of the model has been provided, thus obtaining the empirical likelihood estimation of the model. Then, based on the case deletion model, the first-order approximation formula of parameters is to be derived, with the empirical Cook distance subsequently proposed. Finally, the validity and rationality of the statistical method can be verified by an analysis of a case study.

Keywords：quasi-likelihood nonlinear model；empirical likelihood；case-deletion model；empirical Cook distance；influential point

1　研究背景

在实际的统计分析中，模型的诊断是一个非常重要的环节，其主要任务是通过诊断统计量检测已知观测数据在用既定模型拟合时的合理性。自 R. D. Cook[1] 开始对统计模型的诊断研究起，许多统计学家在此方面做了大量的研究工作。如文献 [2] 介绍了统计模型的局部影响分析方法，这种方法被广泛地应用到统计模型的诊断中。文献 [3] 研究了非线性回归模型的经验似然诊断，用经验似然的方法得到了参数的极大经验似然估计和 3 种不同的影响曲率度量。

自从 R. W. M. Wedderburn 提出拟似然概念以来，拟似然非线性模型的研究受到了学者们的广泛关注。文献 [4] 在欧氏内积空间建立了拟似然非线性模型的几何结构，并导出了参数和子集参数的与统计曲率有关的 3 种近似置信域。对于参数估计问题，文献 [5] 首次研究了拟似然非线性模型中极大似然估计的大样本性质，并证明了该估计的存在性、相合性以及渐进正态性。文献 [6] 讨论了拟似然非线性模型中的极大似然估计的弱相合性，所用的矩的条件比强相合性

收稿日期：2018-11-12
基金项目：国家自然科学基金资助项目（11271189）
作者简介：宋　蕾（1993-），女，山东日照人，南京理工大学硕士生，主要研究方向为概率论与数理统计，
E-mail：354071001@qq.com

的条件要弱。文献 [7] 证明了拟似然非线性模型的拟似然方程解的渐进存在性，并且求出了该解收敛于真值的速度。

虽然近年来有不少学者对拟似然非线性模型进行了研究，但对拟似然非线性模型的统计诊断方面并没有涉及。而在实际应用中，观测数据可能会包含异常点和强影响点等，这时就需要通过统计诊断检测出严重偏离既定模型的数据点，确定模型的合理性。因此研究拟似然非线性模型的统计诊断是非常必要的。

基于以上考虑，本文研究了拟似然非线性模型的统计诊断，并且给出了该模型参数的经验似然估计。通过对数据删除模型的统计诊断，比较删除模型与未删除模型相应统计量之间的差异，检测出数据中存在的强影响点。再通过实例分析来检验该方法的有效性和合理性。

2　模型的参数估计

2.1　拟似然非线性模型

1974 年，R. W. M. Wedderburn 注意到，如果响应变量的均值函数和方差函数被正确给定，许多似然的方法仍可适用，由此提出了拟似然的概念，并且建立了拟似然模型。

假设响应变量 $\boldsymbol{Y}=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ 的各个分量相互独立，其均值向量和协方差矩阵 [8] 满足

$$\begin{cases} E(\boldsymbol{Y})=\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^{\mathrm{T}}, \\ Cov(\boldsymbol{Y})=\sigma^2\boldsymbol{V}(\boldsymbol{\mu})。\end{cases}$$

式中：$\mu_i=\mu(\boldsymbol{x}_i,\boldsymbol{\beta})$，其中 $\boldsymbol{x}_i=(x_{i1}, x_{i2}, \cdots, x_{iq})^{\mathrm{T}}(i=1, 2, \cdots, n)$ 为已知的非随机 q 维设计变量，其定义域为 $\chi\subset\mathbf{R}^q$，$\boldsymbol{\beta}=(\beta_1, \beta_2, \cdots, \beta_p)$ 是 p 维未知参数，其定义域为 $B\subseteq\mathbf{R}^p$，$\mu(\boldsymbol{x}_i,\boldsymbol{\beta})$ 是定义在 $\chi\times B$ 上的已知函数；

σ^2 是不依赖于 $\boldsymbol{\beta}$ 的未知参数，为了简化通常令 $\sigma^2=1$；

$\boldsymbol{V}(\boldsymbol{\beta})\triangleq\boldsymbol{V}(\boldsymbol{\mu})=\mathrm{diag}\left(V_1(\mu_1),\ V_2(\mu_2),\ \cdots,V_n(\mu_n)\right)$ 为已知的对称正定矩阵，其对角元 $V_i=(\mu_i)$ 为定义在 Ω 上的已知函数。

文献 [9] 提出的拟似然函数为

$$Q(\boldsymbol{\beta},\boldsymbol{Y})=\sum_{i=1}^{n}\int_{y_i}^{\mu_i}\frac{y_i-t}{\sigma^2V_i(\mu_i)}\mathrm{d}t\ ,\qquad(1)$$

两边微分，得拟得分函数为

$$q(\boldsymbol{\beta})=\sigma^{-2}\sum_{i=1}^{n}\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(V_i(\mu_i)\right)^{-1}\left(y_i-\mu_i(\boldsymbol{\beta})\right)。\qquad(2)$$

由式（1）~（2）所定义的模型称为拟似然非线性模型。

2.2　模型的极大经验似然估计

由拟得分函数可以得到拟似然非线性模型的估计方程 [10] 为

$$\sum_{i=1}^{n}\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(V_i(\mu_i)\right)^{-1}\left(y_i-\mu_i(\boldsymbol{\beta})\right)=0\ 。\qquad(3)$$

因此，拟似然非线性模型的截面经验似然比函数 [11] 可以定义为

$$R_E(\boldsymbol{\beta})=\sup\left\{\prod_{i=1}^{n}np_i\,\middle|\,p_i\geqslant0,\ \sum_{i=1}^{n}p_i=1,\ \sum_{i=1}^{n}p_ig_i(\boldsymbol{\beta})=0\right\},$$

其中，$p_i=\Pr(x=x_i)$ 为点 x_i 出现的概率。

估计函数为

$$g_i(\boldsymbol{\beta})=\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(V_i(\mu_i)\right)^{-1}\left(y_i-\mu_i(\boldsymbol{\beta})\right)。$$

在约束条件下，最大化截面经验似然比函数所得到的解就是经验似然估计量，即最大化 $\prod_{i=1}^{n}np_i$ 。

为计算简便，对截面经验似然比函数进行对数变换，将乘法运算转换成加法运算，得

$$\sum_{i=1}^{n}\ln p_i+n\ln n\ 。$$

对给定的样本观测，$n\ln n$ 为常数。由此可以定义截面对数似然函数为

$$l_E(\boldsymbol{\beta})=\sup\left\{\sum_{i=1}^{n}\ln p_i\,\middle|\,p_i\geqslant0,\ \sum_{i=1}^{n}p_i=1,\ \sum_{i=1}^{n}p_ig_i(\boldsymbol{\beta})=0\right\}。\qquad(4)$$

由拉格朗日法，构造拉氏函数 [12]

$$H=\sum_{i=1}^{n}\ln p_i+\lambda\left(1-\sum_{i=1}^{n}p_i\right)-n\boldsymbol{t}^{\mathrm{T}}\sum_{i=1}^{n}p_ig_i(\boldsymbol{\beta})\ ,\qquad(5)$$

式中，λ 和 $\boldsymbol{t}=(t_1, t_2, \cdots, t_r)^{\mathrm{T}}$ 是拉格朗日乘子。

式（5）对 p_i 求偏导，可得

$$\frac{\partial H}{\partial p_i}=\frac{1}{p_i}-\lambda-n\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})=0\ ,\qquad(6)$$

化简后可得

$$p_i=\frac{1}{\lambda+n\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}\ 。\qquad(7)$$

将式（6）等号两边同乘 p_i 并对 i 累计求和得

$$\sum_{i=1}^{n}\left(p_i\cdot\frac{\partial H}{\partial p_i}\right)=\sum_{i=1}^{n}p_i\left(\frac{1}{p_i}-\lambda-n\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right)=n-\lambda=0\ ,$$

从而得 $\lambda=n$，故

$$p_i=\frac{1}{n}\cdot\frac{1}{1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}\ 。\qquad(8)$$

由此，拟似然非线性模型的经验似然比函数可以

定义为

$$L_E(\boldsymbol{\beta})=\prod_{i=1}^{n}\left\{\frac{1}{1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}\right\}。\tag{9}$$

经验对数似然比函数为

$$l_E(\boldsymbol{\beta})=-\sum_{i=1}^{n}\ln\left[1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right],\tag{10}$$

极小化经验对数似然比 $l_E(\boldsymbol{\beta})$，可得到 $\boldsymbol{\beta}$ 的经验似然估计 $\hat{\boldsymbol{\beta}}$，即

$$\hat{\boldsymbol{\beta}}=\arg\min l_E(\boldsymbol{\beta})=\arg\max\left\{\sum_{i=1}^{n}\ln\left[1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right]\right\},$$

其中，$\boldsymbol{t}$ 可由 $\frac{1}{n}\sum_{i=1}^{n}\frac{g_i(\boldsymbol{\beta})}{1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}=0$ 确定。

若记

$$Q_n(\boldsymbol{t},\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\ln\left[1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right],\tag{11}$$

则参数 $\boldsymbol{\beta}$ 的经验似然估计 $\hat{\boldsymbol{\beta}}$ 和 $\boldsymbol{t}$ 的估计 $\hat{\boldsymbol{t}}$ 为方程组

$$\begin{cases}Q_{1n}(\boldsymbol{t},\boldsymbol{\beta})=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{g_i(\boldsymbol{\beta})}{1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}=0,\\ Q_{2n}(\boldsymbol{t},\boldsymbol{\beta})=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(\frac{\boldsymbol{t}^{\mathrm{T}}}{1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})}\cdot\frac{\partial g_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)=0。\end{cases}\tag{12}$$

的解，其中，

$$g_i(\boldsymbol{\beta})=\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(V_i(\mu_i)\right)^{-1}\left(y_i-\mu_i(\boldsymbol{\beta})\right);$$

$$\frac{\partial g_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\frac{\partial^2\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\left(V_i(\mu_i)\right)^{-1}\left(y_i-\mu_i(\boldsymbol{\beta})\right)+\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\times\frac{\partial V_i^{-1}(\mu_i)}{\partial\boldsymbol{\beta}}\left(y_i-\mu_i(\boldsymbol{\beta})\right)-\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(V_i(\mu_i)\right)^{-1}\frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}。$$

3　模型诊断

3.1　数据删除模型

为了研究第 i 个数据点对模型统计推断的影响，考虑将该点删除，比较删除前后统计推断结果的变化，以此来判断该点是否为异常点或强影响点 [13-15]。删除第 i 个数据点以后的模型称为数据删除模型。

将模型（1）~（2）中删除第 i 个数据点 (x_i, y_i) 以后的删除模型表示为

$$\begin{cases}Q(\boldsymbol{\beta},\boldsymbol{Y})=\displaystyle\sum_{j\neq i}\int_{y_j}^{\mu_j}\frac{y_j-t}{\sigma^2V_j(\mu_j)}\mathrm{d}t,\\ q(\boldsymbol{\beta})=\sigma^{-2}\displaystyle\sum_{j\neq i}\left(\frac{\partial\mu_j(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\cdot\frac{y_j-\mu_j(\boldsymbol{\beta})}{V_j(\mu_j)}\right)。\end{cases}\tag{13}$$

记 $Q_{n[i]}(\boldsymbol{t},\boldsymbol{\beta})=\frac{1}{n}\sum_{j\neq i}\ln\left[1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right]$ 为删除第 i 个数据点 (x_i, y_i) 后，模型的最优经验似然函数，$\hat{\boldsymbol{\beta}}_{(i)}^{I}$ 为删除第 i 个数据点 (x_i, y_i) 后参数 $\boldsymbol{\beta}$ 的估计值。

定理 1　对于数据删除模型，$\hat{\boldsymbol{\beta}}_{(i)}^{I}$ 和 $\hat{\boldsymbol{t}}_{(i)}^{I}$ 的一阶近似可以表示为

$$\hat{\boldsymbol{\beta}}_{(i)}^{I}=\hat{\boldsymbol{\beta}}-n^{-1}\left(S_{22.1}^{-1}S_{21}S_{22}^{-1}T-S_{22.1}^{-1}N\right)\left[1+o_p(1)\right],$$

$$\hat{\boldsymbol{t}}_{(i)}^{I}=\hat{\boldsymbol{t}}+n^{-1}\left[\left(S_{11}^{-1}+S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\right)T-S_{11}^{-1}S_{12}S_{22.1}^{-1}N\right]\left[1+o_p(1)\right]。$$

其中：$N=\dfrac{g_i(\hat{\boldsymbol{\beta}})}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}$；

$$T=\frac{\hat{\boldsymbol{t}}^{\mathrm{T}}}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}\cdot\frac{\partial g_i(\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}};$$

$$\boldsymbol{S}_{n[i]}=\begin{bmatrix}S_{11}&S_{12}\\S_{21}&S_{22}\end{bmatrix}=\begin{bmatrix}\dfrac{\partial Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}}&\dfrac{\partial Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}\\ \dfrac{\partial Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}}&\dfrac{\partial Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}\end{bmatrix};$$

$S_{22.1}=S_{22}-S_{21}S_{11}^{-1}S_{12}$。

证 明　将 $Q_{1n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)=\dfrac{\partial Q_{n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)}{\partial\boldsymbol{t}^{\mathrm{T}}}=0$ 和 $Q_{2n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)=\dfrac{\partial Q_{n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)}{\partial\boldsymbol{\beta}}=0$ 分别在 $\left(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}}\right)$ 处 Tayor 展开，得

$$Q_{1n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)=Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})+\frac{\partial Q_{1n[i]}}{\partial\boldsymbol{\beta}}\left[\hat{\boldsymbol{\beta}}_{(i)}^{I}-\hat{\boldsymbol{\beta}}\right]+\frac{\partial Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}}\left[\hat{\boldsymbol{t}}_{(i)}^{I}-\hat{\boldsymbol{t}}\right]+o_p(1)=0,$$

$$Q_{2n[i]}\left(\hat{\boldsymbol{t}}_{(i)}^{I},\hat{\boldsymbol{\beta}}_{(i)}^{I}\right)=Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})+\frac{\partial Q_{2n[i]}}{\partial\boldsymbol{\beta}}\left[\hat{\boldsymbol{\beta}}_{(i)}^{I}-\hat{\boldsymbol{\beta}}\right]+\frac{\partial Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}}\left[\hat{\boldsymbol{t}}_{(i)}^{I}-\hat{\boldsymbol{t}}\right]+o_p(1)=0。$$

因为

$$Q_{1n}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})=\frac{1}{n}\sum_{i=1}^{n}\frac{g_i(\hat{\boldsymbol{\beta}})}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}=0,$$

$$Q_{2n}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\boldsymbol{t}}^{\mathrm{T}}}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}\cdot\frac{\partial g_i(\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}=0,$$

所以

$$Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})=-\frac{1}{n}\cdot\frac{g_i(\hat{\boldsymbol{\beta}})}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})},$$

$$Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})=-\frac{1}{n}\cdot\frac{\hat{\boldsymbol{t}}^{\mathrm{T}}}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}\cdot\frac{\partial g_i(\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}。$$

由于 $N=\dfrac{g_i(\hat{\boldsymbol{\beta}})}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}$，$T=\dfrac{\hat{\boldsymbol{t}}^{\mathrm{T}}}{1+\hat{\boldsymbol{t}}^{\mathrm{T}}g_i(\hat{\boldsymbol{\beta}})}\cdot\dfrac{\partial g_i(\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}$，

$$\boldsymbol{S}_{n[i]}=\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}=\begin{bmatrix}\dfrac{\partial Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}} & \dfrac{\partial Q_{1n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}\\ \dfrac{\partial Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{t}^{\mathrm{T}}} & \dfrac{\partial Q_{2n[i]}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}\end{bmatrix}，所以$$

上述 Tayor 展开式化简成矩阵形式为

$$\begin{bmatrix}\hat{\boldsymbol{t}}_{(i)}^{I}-\hat{\boldsymbol{t}}\\ \hat{\boldsymbol{\beta}}_{(i)}^{I}-\hat{\boldsymbol{\beta}}\end{bmatrix}=\frac{1}{n}\boldsymbol{S}_{n[i]}^{-1}(\hat{\boldsymbol{t}},\hat{\boldsymbol{\beta}})\begin{bmatrix}T\\ N\end{bmatrix}。$$

根据分块矩阵的求逆公式 [16]，可以得到 $\hat{\boldsymbol{\beta}}_{(i)}^{I}$ 和 $\hat{\boldsymbol{t}}_{(i)}^{I}$ 的一阶近似公式为

$$\hat{\boldsymbol{\beta}}_{(i)}^{I}=\hat{\boldsymbol{\beta}}-n^{-1}\left(S_{22.1}^{-1}S_{21}S_{22}^{-1}T-S_{22.1}^{-1}N\right)\left[1+o_p(1)\right],$$

$$\hat{\boldsymbol{t}}_{(i)}^{I}=\hat{\boldsymbol{t}}+n^{-1}\left[\left(S_{11}^{-1}+S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\right)T-S_{11}^{-1}S_{12}S_{22.1}^{-1}N\right]\left[1+o_p(1)\right]。$$

证毕。

3.2　经验 Cook 距离

为了研究第 i 个数据点 (x_i, y_i) 对模型的影响，考虑数据删除模型。由于按上述方法得到的 $\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}$ 是一个向量，不便于进行定量的比较，因此需要引入一个度量。类似于 Cook 距离 [17−19]，Zhu H. T. 等 [20] 提出了经验 Cook 距离：

$$ECD_i(\boldsymbol{M})=\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathrm{T}}\boldsymbol{M}\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right),$$

其中 $\boldsymbol{M}$ 可取某一正定矩阵，例如 $\boldsymbol{M}=\dfrac{\partial^2 l_E(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}$。

由上式可以看出，$ECD_i(\boldsymbol{M})$ 越大，则 $\hat{\boldsymbol{\beta}}_{(i)}$ 与 $\hat{\boldsymbol{\beta}}$ 之间的差异越大。在进行具体的数据分析时，先用有关软件逐点分别计算出经验 Cook 距离，再通过画图或者列表，找到一个或几个特别大的 ECD_i，则相应的数据点就是对参数 $\boldsymbol{\beta}$ 的估计影响特别大的点，称为强影响点。

对 ECD_i 进行分析时，若存在一个判别的界限 D，当经验 Cook 距离大于这个界限时，就认为第 i 个数据点为强影响点。遗憾的是，这个问题目前还没有明确的结论，只能人为地进行判断。

4　实例分析

为了验证似然非线性模型的统计诊断的有效性，对下述实例进行分析。

例 1　对表 1[21] 中数据进行分析，判断其中的异常点或强影响点。表中 y 表示 1 周内损坏的电子元件数，试验共进行了 20 周，每周试验分为 2 组，每组所用时间分别为 x_1 和 x_2。

表 1　1 周内损坏的电子元件数和时间

Table 1　Number and time of electronic components damaged in a week

序号	x_1	x_2	y	序号	x_1	x_2	y
1	33.3	25.3	15	11	63.1	30.6	15
2	52.5	14.4	9	12	106.3	58.1	22
3	64.7	32.5	14	13	113.7	57.3	26
4	137.0	20.5	24	14	32.6	46.8	12
5	125.9	97.6	27	15	87.3	61.5	28
6	116.3	53.6	27	16	90.4	43.9	21
7	131.7	56.6	23	17	128.4	94.6	24
8	85.0	87.3	18	18	102.1	39.5	23
9	91.9	47.8	22	19	73.5	47.8	18
10	32.0	23.4	12	20	71.3	69.2	16

4.1　模型分析

假设 $\sigma^2=1$，均值向量为 $E(\boldsymbol{Y})=\boldsymbol{\mu}=(\mu_1, \mu_2, \cdots, \mu_{20})$，其中，$\mu_i=\exp\{\beta_0+x_{i1}\beta_1+x_{i2}\beta_2\}$，方差为 $\boldsymbol{V}(\boldsymbol{\mu})=\boldsymbol{\mu}$，此时拟似然非线性模型可以表示为

$$\begin{cases}Q(\boldsymbol{\beta},\boldsymbol{Y})=\displaystyle\sum_{i=1}^{20}\left[y_i\ln\mu_i-\mu_i\right],\\ q(\boldsymbol{\beta})=\displaystyle\sum_{i=1}^{20}\left[\frac{\partial\mu_i}{\partial\boldsymbol{\beta}}\frac{1}{\mu_i}(y_i-\mu_i)\right]。\end{cases}$$

由拟似然函数可以看出响应变量 y_i 服从 Poisson 分布，故该模型为 Poisson 回归模型。

由拟得分函数可以得到模型的估计方程为

$$g_i(\boldsymbol{\beta})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Z},$$

式中：$\boldsymbol{X}^{\mathrm{T}}=(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{20})$，$\boldsymbol{X}_i=(1, x_{i1}, x_{i2})$；$\boldsymbol{Z}=(z_1, z_2, \cdots, z_{20})$，$z_i=y_i-\mu_i$。

结合 2.1 节可知，模型的经验对数似然比函数为

$$l_E(\boldsymbol{\beta})=-\sum_{i=1}^{n}\ln\left[1+\boldsymbol{t}^{\mathrm{T}}g_i(\boldsymbol{\beta})\right],$$

极小化经验对数似然比 $l_E(\boldsymbol{\beta})$，可以得到 $\boldsymbol{\beta}$ 的经验似然估计 $\hat{\boldsymbol{\beta}}$。

4.2　参数估计

由 4.1 节的分析和已知数据，利用 R 软件即可以求出参数 $\boldsymbol{\beta}$ 的估计值

$$\hat{\boldsymbol{\beta}}=(2.256\,336, 0.007\,007, 0.001\,769)^{\mathrm{T}}。$$

4.3　影响分析

在参数估计的基础上，首先对参数进行残差分析，然后计算经验 Cook 距离。图 1 和图 2 分别是标准化残差和经验 Cook 距离的散点图。

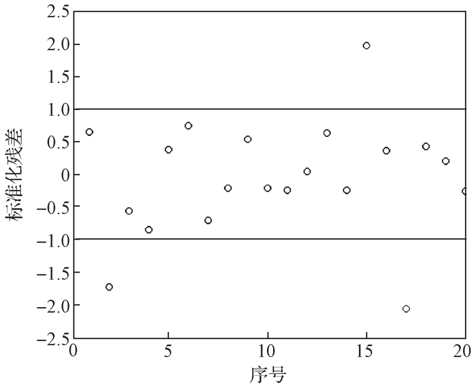

图 1　残差图

Fig. 1　Distribution map of the residual plot

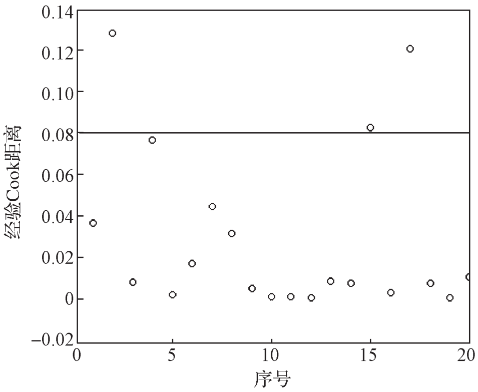

图 2　经验 Cook 距离图

Fig. 2　Diagram of empirical Cook distance

由图 1 可以明显地看出，第 2 号点、15 号点和 17 号点明显异于其它点，可以认为这 3 个点是“异常”的，但并不能判断这 3 个点到底是异常点、强影响点还是高杠杆点，于是利用经验 Cook 距离图进行下一步的判断。

从图 2 也可以看出第 2 号点、15 号点和 17 号点的数值高于其他点，且在横线之上，故可以认为第 2 号点、15 号点和 17 号点为强影响点。

5　结语

本文研究了拟似然非线性模型在经验似然估计下的统计诊断问题。首先利用经验似然比函数求出模型的经验似然估计；再研究了数据删除模型并进行了影响分析，给出了模型的统计诊断量和近似公式；最后通过对实例分析，验证了该统计方法的有效性和合理性。

参考文献：

[1] COOK R D. Detection of Influential Observation in Linear Regression[J]. Technometrics，1977，19(1)：15–18.

[2] THOMAS W，COOK R D. Assessing Influence on Regression Coefficients in Generalized Linear Models[J]. Biometrika，1989，76(4)：741–749.

[3] 丁先文，徐　亮，林金官 . 非线性回归模型的经验似然诊断 [J]. 应用数学学报，2012，35(4)：693–702.
DING Xianwen，XU Liang，LIN Jinguan. Diagnostic Measures for Nonlinear Regression Models Based on Empirical Likelihood Method[J]. Acta Mathematicae Applicatae Sinica，2012，35(4)：693–702.

[4] TANG N S，WANG X R. Confidence Regions in Quasi-Likelihood Nonlinear Models：A Geometric Approach[J]. Journal of Biomathematics，2000，15(1)：55–64.

[5] 夏　天 . 拟似然非线性模型参数估计的大样本理论 [D]. 昆明：云南大学，2006.
XIA Tian. Theory of Large Sample of Parametric Estimator in Quasi-Likelihood Nonlinear Models[D]. Kunming：Yunnan University，2006.

[6] 夏　天，李友光，王学仁 . 拟似然非线性模型中 MQLE 的弱相合性的充分条件 [J]. 数学的实践与认识，2011，41(17)：168–173.
XIA Tian，LI Youguang，WANG Xueren. Suffucuent Conditions of Weak Consistency of MQLE in Quasi-Likelihood Nonlinear Models[J]. Mathematics in Practice and Theory，2011，41(17)：168–173.

[7] 张　戈 . 拟似然非线性模型中极大拟似然估计的强收敛速度 [J]. 喀什师范学院学报，2015，36(6)：8–10.
ZHANG Ge. Strong Convergence Rates of Maximum Quasi-Likelihood Estimation in Quasi-Likelihood Nonlinear Models[J]. Journal of Kashgar Teachers College，2015，36(6)：8–10.

[8] 夏　天，王学仁 . 带自适应设计的拟似然非线性模型中极大拟似然估计的强相合性 [J]. 数学的实践与认识，2015，45(21)：252–258.
XIA Tian，WANG Xueren. Strong Consistency of Maximum Quasi-Likelihood Estimator in Quasi-Likelihood Nonlinear Models with Adaptive Designs[J]. Mathematics in Practice and Theory，2015，45(21)：252–258.

[9] WEDDERBURN R W M. Quasi-Likelihood Function Generalized Linear Models，and the Gauss-Newton Method[J]. Biometrika，1974，61(3)：439–447.

[10] 陈显彬，林　松，尹长明 . 两阶段 Logit 广义估计方程的渐近性质 [J]. 重庆理工大学学报 (自然科学)，2018，32(5)：221–224.
CHEN Xianbin，LIN Song，YIN Changming. Asymptotic Properties of Two-Step Logit Generalized Estimating Equation[J]. Journal of Chongqin University of Technology (Natural Science)，2018，32(5)：221–224.

[11] 王淑玲，冯　予，刘易达 . 纵向数据下部分线性模

型的经验似然诊断 [J]. 黑龙江大学自然科学学报，2016，33(4)：450−455.
WANG Shuling，FENG Yu，LIU Yida. Empirical Likelihood Diagnosis for Partially Linear Models with Longitudinal Data[J]. Journal of Natural Science of Heilongjiang University，2016，33(4)：450−455.

[12] LI Yan，XIA Chen. Empirical Likelihood for Generalized Linear Models with Fixed and Adaptive Designs[J]. Statistics，2015，49(5)：978−988.

[13] 韦博成，林金官，解锋昌 . 统计诊断 [M]. 北京：高等教育出版社，2009：25−31.
WEI Bocheng，LIN Jinguan，XIE Fengchang. Statistical Diagnostics[M]. Beijing：Higher Education Press，2009：25−31.

[14] 王淑玲， 谢　凤， 朱倩倩 . 某型飞机燃油消耗随机森林模型的统计诊断 [J]. 黑龙江大学自然科学学报，2018，35(1)：61−64.
WANG Shuling，XIE Feng，ZHU Qianqian. Statistical Diagnosis for Random Forest Model of Aircraft Fuel Consumption Data[J]. Journal of Natural Science of Heilongjiang University，2018，35(1)：61−64.

[15] 仝　倩，冯　予 . 带缺失数据的半参数非线性模型基于经验似然的统计诊断 [J]. 贵州师范大学学报（自然科学版），2017，35(4)：89−94.
TONG Qian，FENG Yu. Diagnostic and Empirical Likelihood for Semi：Parametric Nonlinear Model with Missing Data[J]. Journal of Guizhou Normal University (Natural Sciences)，2017，35(4)：89−94.

[16] 马　骏，李　萌，靳丹丹，等 . 几类 3×3 分块矩阵的逆矩阵的求法 [J]. 曲阜师范大学学报，2016，42(3)：54−58.
MA Jun，LI Meng，JIN Dandan，et al. On the Inverse Matrix of Certain 3×3 Block Matrices[J]. Journal of Qufu Normal University，2016，42(3)：54−58.

[17] COOK R D，WEISBERG S. Residuals and Influence in Regression[J]. Technometrics，1984，39(4)：413−415.

[18] 朱　宁，刘庆华 . 修正 LIU 估计下数据删除模型的强影响分析 [J]. 汕头大学学报（自然科学版），2017，32(1)：30−37.
ZHU Ning，LIU Qinghua. Strong Impact Analysis of Data Delete Model Based on Modified LIU Estimator[J]. Journal of Shantou University (Natural Science)，2017，32(1)：30−37.

[19] ZHANG Xiao，ZHANG Boqiao，WU Hongya，et al. Effect of High-Molecular-Weight Glutenin Subunit Deletion on Soft Wheat Quality Properties and Sugar-Snap Cookie Quality Estimated Through Near-Isogenic Lines[J]. Journal of Integrative Agriculture，2018，17(5)：1066–1073.

[20] ZHU H T，IBRAHIM J G，TANG N S，et al. Diagnostic Measures for Empirical Likelihood of General Estimating Equations[J]. Biometrika，2008，95(2)：489−507.

[21] JORGERSEN D W. Multiple Regression Analysis of Poisson Process[J]. Journal of the American Statistical Association，1961，56：235−245.

（责任编辑：邓光辉）