Journal of Southeast University (English Edition) Vol. 29, No. 1, pp. 106 – 110 Mar. 2013 ISSN 1003—7985

Diagnostics in generalized nonlinear models based on maximum L_q-likelihood estimation

Xu Weijuan Lin Jinguan

(Department of Mathematics, Southeast University, Nanjing 211189, China)

Abstract: In order to detect whether the data conforms to the given model, it is necessary to diagnose the data in the statistical way. The diagnostic problem in generalized nonlinear models based on the maximum L_q-likelihood estimation is considered. Three diagnostic statistics are used to detect whether the outliers exist in the data set. Simulation results show that when the sample size is small, the values of diagnostic statistics based on the maximum L_q-likelihood estimation are greater than the values based on the maximum likelihood estimation. As the sample size increases, the difference between the values of the diagnostic statistics based on two estimation methods diminishes gradually. It means that the outliers can be distinguished easier through the maximum L_q-likelihood method than those through the maximum likelihood estimation method.

Key words: maximum L_q-likelihood estimation; generalized nonlinear regression model; case-deletion model; generalized Cook distance; likelihood distance; difference of deviance

doi: 10.3969/j.issn.1003 – 7985.2013.01.022

Linear regression diagnostics have developed well in the past three decades. A comprehensive study on this topic can be found, for instance, in Refs. [1 – 2]. However, there has not been much published work on regression diagnostics for the models outside of linear regression[3-4]. Only a few papers are related to the diagnostics for exponential family nonlinear models[4-6]. Wei[7] described the diagnostics for generalized nonlinear models systematically in chapter 6.

All the discussion above is based on the maximum likelihood method. Standard large sample theory guarantees that the maximum likelihood estimator (MLE) is asymptotically efficient. It means that when the sample is large, the MLE is at least as accurate as any other estimator. However, when the sample is moderate or small, the properties of the MLE may not be very good. So, we need a new estimation method, which can make the property better.

In this paper, we propose a modified diagnostic method for a generalized nonlinear model based on the maximum L_q-likelihood estimator[8]. Diagnostics for regression parameters and dispersion parameters are considered, and three kinds of diagnostic statistics are proposed. For small and modest sample sizes, the proposed diagnostic method is still available when q is properly chosen. Some simulations are performed to investigate the behavior of the proposed methods for different sample sizes. We also show that the modified diagnostic statistics method outperform the classical diagnostic statistics method when the sample sizes are modest or even small.

1 Generalized Entropy and Maximum L_q-Likelihood Estimator

The Kullback-Leibler (KL) divergence[9] is one of the most popular quantities employed to measure the distance of a distribution with respect to a "true" distribution. Consider a σ-finite measure μ on a measurable space Ω and let M be the set of all probability distribution functions f. The expectation with respect to f is denoted by E_f. The KL divergence between two density functions g and f is defined as

$$\Delta(f \parallel g) = E_f \log\left(\frac{f(X)}{g(X)}\right) = \int_\Omega \log\left(\frac{f(x)}{g(x)}\right) f(x)\, \mathrm{d}\mu(x) \tag{1}$$

Note that finding the density g that minimizes $\Delta(f \parallel g)$ is equivalent to minimizing Shannon's entropy[10]:

$$H(f \parallel g) = -E_f \log g(X) \tag{2}$$

Definition 1 Let f and g be two density functions; and the q-entropy is defined as

$$H_q(f, g) = -E_f[L_q(g(X))] \qquad q > 0 \tag{3}$$

where

$$L_q(u) = \begin{cases} \dfrac{u^{1-q} - 1}{1 - q} & \text{if } q \neq 1 \\ \log u & \text{if } q = 1 \end{cases} \tag{4}$$

The function L_q represents a Box-Cox transformation in statistics. In other fields it is often called a deformed logarithm. The above characterization emphasizes the simi-

Received 2012-03-09.

Biographies: Xu Weijuan (1977—), female, graduate; Lin Jinguan (corresponding author), male, doctor, professor, jglin@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 11171065), the Natural Science Foundation of Jiangsu Province (No. BK2011058).

Citation: Xu Weijuan, Lin Jinguan. Diagnostics in generalized nonlinear models based on maximum L_q-likelihood estimation [J]. Journal of Southeast University (English Edition), 2013, 29(1): 106 – 110. [doi: 10.3969/j.issn.1003 – 7985.2013.01.022]

larity to the classical Shannon's entropy: if $q \to 1$, then $L_q(u) \to \log(u)$ and the usual definition of Shannon's entropy is recovered.

Recently, Ferrari and Yang[8] introduced an estimator inspired by Havrda, and Charvát[11] generalized information measures (usually called α-order entropies or q-entropies in physics), the maximum L_q-likelihood estimator.

Definition 2 Let $X_1, \ldots, X_n$ be an i. i. d. sample from $f(x;\theta_0)$, $\theta_0 \in \Theta$. The maximum L_q-likelihood estimator of θ_0 is defined as

$$\tilde{\theta}_n = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} L_q[f(X_i;\theta)] \qquad q > 0 \tag{5}$$

where L_q is the q-logarithmic function defined in Eq. (4) with $q > 0$. The L_q-likelihood equation is

$$\frac{\partial}{\partial\theta} \sum_{i=1}^{n} L_q[f(X_i;\theta)] = 0 \tag{6}$$

Note that when the distortion parameter q tends to 1, $L_q(\cdot) \to \log(\cdot)$ and the usual MLE is recovered. In this sense, the maximum L_q-likelihood estimation extends the classic method, resulting in a general inferential procedure that inherits most of the desirable features of traditional maximum likelihood, and at the same time gains some new properties that can be exploited in particular estimation settings.

2 Generalized Nonlinear Model

Suppose that the components of $\boldsymbol{Y} = \{y_1, \cdots, y_n\}^{\mathrm{T}}$ are independent random variables, in which each y_i may depend on an independent known variable $x_i (i = 1, 2, \cdots, n)$. The parameters of interest are $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_p\}^{\mathrm{T}}$ defined in a subset B of $R^p (p < n)$ and the distribution of y_i depending on x_i satisfies the following constraint conditions:

$$g(\mu_i) = f(x_i, \boldsymbol{\beta}),\ y_i \sim ED(\mu_i, \sigma^2) \qquad i = 1, 2, \cdots, n \tag{7}$$

where $g(\cdot)$ is a known monotonic link function; $f(\cdot\,;\cdot)$ is a known function with an unknown vector parameter $\boldsymbol{\beta}$ and a known explanatory variable of q-vector x_i; $E(y_i) = \mu_i$ and $\mathrm{ED}(\mu_i, \sigma^2)$ is the exponential family distribution, which has the density function in the following form as

$$p(y_i;\theta_i,\varphi) = \exp\left\{\varphi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}s(\varphi, y_i)\right\}$$
$$\mu_i = \mu_i(\boldsymbol{\beta}) \text{ or } \theta_i = \theta_i(\boldsymbol{\beta}) \tag{8}$$

where θ_i is the natural parameter; and $\sigma^2 = \varphi^{-1}$ is the dispersion parameter. According to the property of the exponential family distribution[7], we have

$$\mu_i = E(y_i) = \dot{b}(\theta_i), \quad \mathrm{var}(y_i) = \sigma^2 \ddot{b}(\theta_i) \tag{9}$$

From model (8), the log-likelihood of $\boldsymbol{Y}$ for the parameter $\boldsymbol{\beta}$ and φ is usually denoted by

$$L(\boldsymbol{\beta}, \varphi) = \sum_{i=1}^{n} \left\{\varphi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}s(\varphi, y_i)\right\} \tag{10}$$

Then the L_q-likelihood of $\boldsymbol{Y}$ is

$$L_{q_n}(\boldsymbol{\beta}, \varphi) = \frac{\exp\left[\sum_{i=1}^{n} (1 - q_n)\left\{\varphi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}s(\varphi, y_i)\right\}\right] - 1}{1 - q_n} \tag{11}$$

Now let $\tilde{\boldsymbol{\beta}}_n$ and $\tilde{\varphi}_n$ be the maximum L_q-likelihood estimators of $\boldsymbol{\beta}$ and φ for the exponential family nonlinear model (or generalized nonlinear model) (7), which satisfies

$$L_{q_n}(\tilde{\boldsymbol{\beta}}_n, \tilde{\varphi}_n) = \max L_{q_n}(\boldsymbol{\beta}, \varphi) \tag{12}$$

3 Diagnostics for Regression Parameter and Dispersion Parameter

A fundamental approach of influence diagnostics is based on the comparison of parameter $\tilde{\boldsymbol{\beta}}_n$ and $\tilde{\varphi}_n$ with parameter estimates $\tilde{\boldsymbol{\beta}}_{(i)}$ and $\tilde{\varphi}_{(i)}$ that correspond to the so-called case deletion model (CDM):

$$\begin{aligned} y_i &\sim \mathrm{ED}(\mu_i, \sigma^2) \qquad i = 1, 2, \cdots, n \\ g(\mu_j) &= f(x_j;\boldsymbol{\beta}) \qquad j = 1, 2, \cdots, n;\ j \neq i \end{aligned} \tag{13}$$

This is just model (7) with the i-th case deleted.

To find the influence points, we compute a certain "distance" between $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ and $(\tilde{\boldsymbol{\beta}}_{(i)}, \tilde{\varphi}_{(i)})$ or $\tilde{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_{(i)}$. The latter is often used in practice because if a case is influential to $\tilde{\boldsymbol{\beta}}_n$; then this must be influential to $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$. Here we introduce three kinds of distance which are very often used in influence diagnostics and can be used for generalized nonlinear models.

1) Generalized Cook distance

This is a norm of $\tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)}$ with respect to a certain weight matrix $\boldsymbol{M} > \boldsymbol{0}$ and defined as

$$\mathrm{GD}_i = \| \tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)} \|_{\boldsymbol{M}}^2 = (\tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)})^{\mathrm{T}} \boldsymbol{M} (\tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)})$$

It is very natural to choose $\boldsymbol{M} = J(\boldsymbol{\beta})$, the Fisher information matrix of $\boldsymbol{Y}$ for $\boldsymbol{\beta}$. Since the Fisher information matrix of $\tilde{\boldsymbol{\beta}}_n$ is difficult to calculate, we use the observed information matrix $[-\ddot{L}_{q_n}(\tilde{\boldsymbol{\beta}}_n, \tilde{\varphi}_n)]^{-1}$ to replace $J(\boldsymbol{\beta})$. So the generalized Cook distance is

$$\mathrm{GD}_i = (\tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)})^{\mathrm{T}} [-\ddot{L}_{q_n}(\tilde{\boldsymbol{\beta}}_n, \tilde{\varphi}_n)]^{-1} (\tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_{(i)}) \tag{14}$$

2) Likelihood distance

The likelihood distance is defined as[1]

$$\mathrm{LD}_i(\boldsymbol{\beta}) = 2\{L_{q_n}(\tilde{\boldsymbol{\beta}}_n) - L_{q_n}(\tilde{\boldsymbol{\beta}}_{(i)})\} \tag{15}$$

We can calculate LD_i directly based on $\tilde{\boldsymbol{\beta}}_n$, $\tilde{\boldsymbol{\beta}}_{(i)}$ and the L_q-likelihood function L_{q_n}. Because $L_{q_n}(\tilde{\boldsymbol{\beta}}_n)$ is the maximum of $L_{q_n}(\boldsymbol{\beta})$ for all $\boldsymbol{\beta}$, which is defined in a subset B of R^p, $\mathrm{LD}_i(\boldsymbol{\beta}) \geqslant 0$ is always true. LD_i means the diversification of $L_{q_n}(\boldsymbol{\beta})$, which is removed from the i-th object before and after.

3) Difference of deviance

The differences of deviance has the form as

$$\Delta_i D = D(\tilde{\boldsymbol{\beta}}_n) - D_{(i)}(\tilde{\boldsymbol{\beta}}_{(i)}) \tag{16}$$

where $D(\boldsymbol{\beta}) \triangleq \sum_{j=1}^{n} d_j(y_j, \mu_j(\boldsymbol{\beta}))$, $D_{(i)}(\boldsymbol{\beta}) = \sum_{j \neq i} d_j(y_j, \mu_j(\boldsymbol{\beta}))$, $d_j(y_j, \mu_j(\boldsymbol{\beta})) = -2\{y_j\theta_j - b(\theta_j) - c(y_j)\} + 2\{y_j\theta_j - b(\theta_j) - c(y_j)\}_{\mu_j = y_j}$. $\Delta_i D$ actually reflects the differences of the maximum likelihood estimators of σ^2 or φ under the original model and under the CDM.

4 Numerical Simulations

We obtain the formula to the maximum L_q-likelihood estimators of $\boldsymbol{\beta}$ and φ under Eq. (12) in Section 2. In Section 3, three kinds of distance, which are used for the diagnostics, are introduced. Now let us look at a numerical simulation example for computing diagnostic statistics.

Suppose that the components of $\boldsymbol{Y} = \{y_1, \cdots, y_n\}^{\mathrm{T}}$ are independent random variables, in which each y_i may depend on an independent known variable x_i ($i = 1, 2, \cdots, n$). Assume that data is fitted by a Gamma nonlinear model[7], that is, $y_i \sim \mathrm{GA}(\mu_i, \sigma^2)$. The probability density function is presented in Eq. (8), where $\sigma^2 = \varphi^{-1}$, $\mu_i = \dfrac{1}{\beta_0 + \beta_1 x_i}$, $\theta_i = -\mu_i^{-1}$, $b(\theta_i) = -\log(-\mu_i^{-1})$, $c(y_i) = -\log(y_i)$, $s(y_i, \varphi) = -2(\varphi\log\varphi - \log\Gamma(\varphi)) + 2\log y_i$. Fig. 1 is the GD_i based on MLE when $n = 30, 50$ and 80. Fig. 2 is the GD_i based on the maximum L_q-likelihood estimator when $n = 30, 50, 80$. Fig. 3 is the LD_i based on MLE when $n = 30, 50, 80$. Fig. 4 is the LD_i based on the maximum L_q-likelihood estimator when $n = 30, 50, 80$.

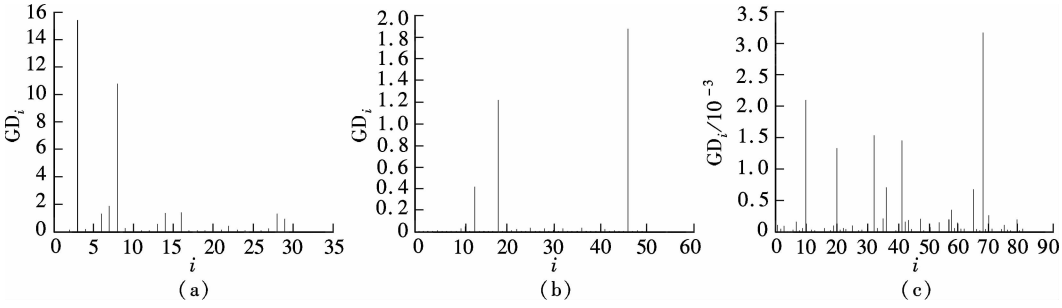

Fig. 1 GD_i based on MLE. (a) $n = 30$; (b) $n = 50$; (c) $n = 80$

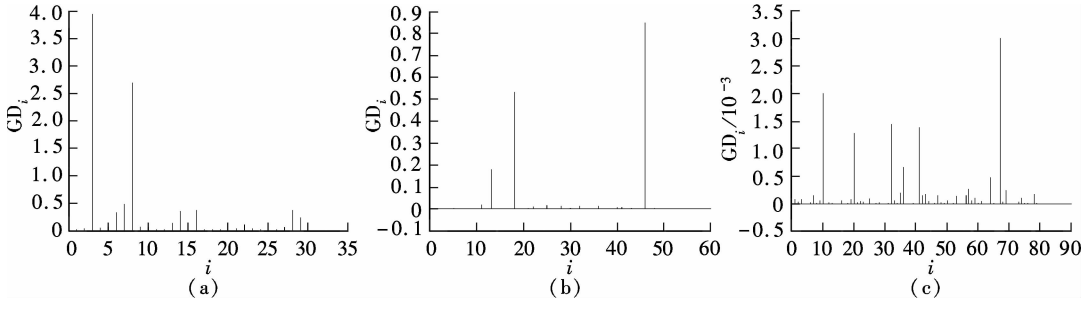

Fig. 2 GD_i based on the maximum L_q-likelihood estimator. (a) $n = 30$; (b) $n = 50$; (c) $n = 80$

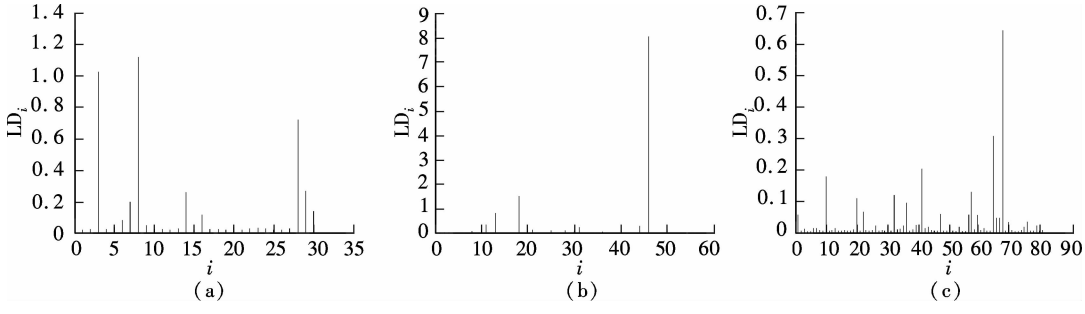

Fig. 3 LD_i based on MLE. (a) $n = 30$; (b) $n = 50$; (c) $n = 80$

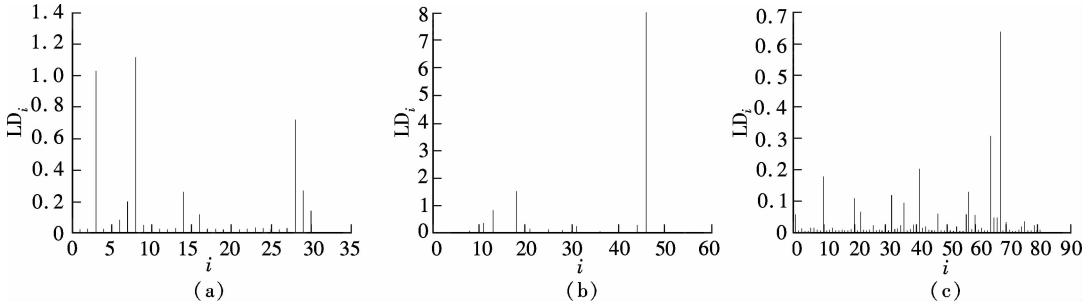

Fig. 4 LD_i based on the maximum L_q-likelihood estimator. (a) $n=30$; (b) $n=50$; (c) $n=80$

From the above figures, we can obtain the same strong impact points through calculating the value about GD_i and LD_i. The value of GD_i about the strong impact point which obtains from the maximum L_q-likelihood method is bigger than the value obtained from the maximum likelihood method when the sample size is small. With the increase in the sample size, the difference between them is small. It means that the maximum L_q-likelihood estimation method is more effective than the MLE method when the sample size is small.

5 Illustrative Example

We have the formula to the maximum L_q-likelihood estimators of $\boldsymbol{\beta}$ and φ in Section 3 and three diagnostic statistics in Section 4. Now let us look at an example for computing diagnostic statistics.

Example 1 Product sales data

These data were given by Whitmore[12], which is shown in Tab. 1. We call them the "Produce sales data" for short. In this dataset, x_i represents the projected sales amounts of the i-th product reported by a market survey organization and y_i is the corresponding actual sales amounts of a company ($i=1,2,\cdots,20$). Whitmore suggested an inverse Gaussian fit by using

$$y_i \sim \mathrm{IG}(\beta x_i^{\gamma},\ \kappa^{-1}x_i^{-\rho}) \qquad i=1,2,\cdots,20 \qquad (17)$$

that is $E(y_i)=\mu_i=\beta x_i^{\gamma}$, $\mathrm{var}(y_i)=\sigma_i^2\,V(\mu_i)$, where $\sigma_i^{-2}=\kappa x_i^{\rho}$ and $V(\mu_i)=\mu_i^3$. For ease of calculation, Wei[7] set $\rho=0$. In this case, $\sigma_i^2=\kappa^{-1}$ for all i. Then (17) becomes $y_i \sim \mathrm{IG}(\beta x_i^{\gamma},\ \kappa^{-1})$ which is an inverse Gaussian nonlinear model with $\mu_i=\beta x_i^{\gamma}$ and $\mathrm{var}(y_i)=\sigma_i^2\,\mu_i^3$ ($\sigma_i^2=\kappa^{-1}$). We obtain the maximum L_q-likelihood estimators from Eq. (12) directly, that is $\tilde{\beta}_n=1.181\ 3$, $\tilde{\kappa}_n=6\,543.9$, $\tilde{\gamma}_n=0.987\ 3$.

Then, we obtain the values of the diagnostic statistics (see Tab. 2).

All the results indicate that case 10 may be an outstanding outlier and the influence of case 20 is also obvious. Through the comparison between our results and the diagnostic results obtained through MLE[13], we find that No. 10 and 20 are two strong impact points, and the value of GD_i and LD_i using the maximum L_q-likelihood estimation method is greater than the value using the MLE method. It means that the result we arrived at is more significant. So it is feasible to diagnose using the maximum L_q-likelihood method.

Tab. 1 Product sales data

i	x_i	y_i	i	x_i	y_i
1	5 950	5 673	11	3 534	3 659
2	2 641	2 565	12	1 965	2 182
3	1 738	1 839	13	1 182	1 236
4	667	918	14	613	902
5	610	756	15	549	500
6	527	487	16	353	463
7	331	225	17	290	257
8	253	311	18	193	212
9	156	166	19	133	123
10	122	198	20	114	99

Tab. 2 Some diagnostic statistics for product sales data

i	GD_i	LD_i	$\Delta_i D$
1	0.025 6	0.025 2	$1.833\,7\times10^{-6}$
2	0.026 4	0.025 8	$3.458\,4\times10^{-6}$
3	0.025 8	0.025 4	$1.337\,4\times10^{-7}$
4	0.055 8	0.063 0	$8.493\,8\times10^{-5}$
5	0.033 4	0.033 4	$2.754\,8\times10^{-5}$
6	0.038 4	0.032 6	$5.129\,8\times10^{-5}$
7	0.704 4	0.622 5	$6.710\,6\times10^{-4}$
8	0.038 2	0.036 3	$4.651\,0\times10^{-5}$
9	0.034 6	0.032 2	$1.120\,2\times10^{-5}$
10	15.839 4	16.677 7	$1.607\,1\times10^{-3}$
11	0.025 6	0.025 3	$2.165\,3\times10^{-7}$
12	0.025 9	0.025 5	$6.286\,6\times10^{-7}$
13	0.026 1	0.025 6	$8.728\,7\times10^{-7}$
14	0.087 4	0.110 2	$1.522\,2\times10^{-4}$
15	0.040 7	0.034 0	$5.795\,1\times10^{-5}$
16	0.050 2	0.051 1	$8.951\,0\times10^{-5}$
17	0.059 0	0.054 2	$1.546\,9\times10^{-4}$
18	0.026 2	0.025 9	$2.231\,8\times10^{-7}$
19	0.637 1	0.346 6	$2.700\,2\times10^{-4}$
20	3.790 2	1.467 3	$6.123\,2\times10^{-4}$

6 Conclusion

In this paper, we consider the diagnostics of generalized nonlinear regression models. Three diagnostic statistics and a new estimation method, the maximum L_q-likelihood estimator, are introduced. Through the value of

the diagnostic statistics, we can make a decision of whether the individual point of the data is a strong impact one or not. By comparing the results obtained from the maximum likelihood estimator and the maximum L_q-likelihood estimator, we find that the method of the maximum L_q-likelihood estimation is more effective when the sample size is small. When the sample size is bigger, the values about the diagnostic statistics are almost the same. In other words, the performance of our proposed method is equivalent to that of the classical diagnostic method for large sample sizes. The same conclusion can also be obtained from the example in Section 5.

This paper discusses some diagnostic problems for the generalized nonlinear model by using the maximum L_q-likelihood estimation method and compares the results with those of the classical diagnostic method. However, consistency and asymptotic normality of the maximum L_q-likelihood estimation for the generalized nonlinear model have not been proved, which is obviously of great significance. Thus, more research on this issue will be valuable and we believe this is an interesting direction for further exploration.

References

[1] Cook R D, Weisberg S. *Residual and influence in regression* [M]. London: Chapman and Hall, 1982.

[2] Chatterjee S, Hadi A S. Influential observations, high leverage points and outliers in linear regression (with discussion) [J]. *Statistical Science*, 1986, **1**(3): 379 - 416.

[3] McCullagh P, Nelder J A. *Generalized linear models* [M]. London: Chapman and Hall, 1989.

[4] Davison A C, Tsai C L. Regression model diagnostics [J]. *International Statistical Review*, 1992, **60**(3): 337 - 355.

[5] Gay D M, Welsch R E. Maximum likelihood and quasi-likelihood for nonlinear exponential family regression models [J]. *Journal of the American Statistical Association*, 1988, **83**(404): 990 - 998.

[6] Wei B C and Shi J Q. On statistical models in regression diagnostics [J]. *Ann Inst Statist Math*, 1994, **46**(2): 267 - 278.

[7] Wei B C. *Exponential family nonlinear models* [M]. Singapore: Springer, 1998.

[8] Ferrari D, Yang Y. Estimation of tail probability via the maximum L_q-likelihood method [R]. Minneapolis, MN, USA: School of Statistics, University of Minnesota, 2007.

[9] Kullback S. *Information theory and statistics* [M]. Wiley: New York, 1959.

[10] Shannon C E. A mathematical theory of communication [J]. *Bell System Technical Journal*, 1948, **27**: 379 - 423.

[11] Havrda J, Charvát F. Quantification method of classification processes: concept of structural entropy [J]. *Kibernetika*, 1967, **3**: 30 - 35.

[12] Whitmore D A. Inverse Gaussian ratio estimation [J]. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 1986, **35**(1): 8 - 15.

[13] Wei B C, Lin J G, Xie F C. *Statistical diagnosis* [M]. Beijing: Higher Education Press, 2009. (in Chinese).

基于最大L_q似然估计的广义非线性模型的统计诊断

徐伟娟　　林金官

(东南大学数学系, 南京 211189)

摘要: 为了检测数据是否符合给定的模型,需要对数据进行统计诊断. 研究了基于最大 L_q 似然估计的广义非线性模型的统计诊断问题. 利用 3 个统计诊断量来检验数据中是否都存在异常点. 模拟结果显示,当样本容量较小时,使用最大 L_q 似然估计方法得到的诊断统计量的结果要比使用极大似然估计(MLE)方法得到的结果大;随着样本容量的增加,它们之间的区别逐渐减小. 因此,使用最大 L_q 似然估计方法比用 MLE 方法更容易找到数据中的异常点.

关键词: 最大 L_q 似然估计;广义非线性回归模型;数据删除模型;广义 Cook 距离;似然距离;偏差度

中图分类号: O212.4