Applied Mathematics, 2012, 3, 1065-1067
http://dx.doi.org/10.4236/am.2012.39156 Published Online September 2012 (http://www.SciRP.org/journal/am)

Local Influence Analysis of Generalized Linear Model

Shuling Wang, Man Liu, Qianqian Zhu, Ting Wang
Basic Courses Department, Xuzhou Air Force College, Xuzhou, China
Email: wangshuling2007@yahoo.com.cn

Received May 6, 2012; revised July 30, 2012; accepted August 6, 2012

ABSTRACT

Nearly thirty years, the diagnosis and influence analysis of linear regression model has been fully developed. So far the local influence analysis of the generalized linear model has not yet seen in the literature. In this paper, local influence is discussed. Then, concise influence matrix is obtained. At last, an example is given to illustrate our results.

Keywords: Exponential Distributions; Local Influence Analysis; Influence Matrix

1. Introduction

Local influence analysis is proposed from the viewpoint of differential geometry [1]. Nearly thirty years, the diagnosis and influence analysis of linear regression model has been fully developed [2,3]. Regarding the generalized linear model, diagnosis has some results [4]. So far the local influence analysis of the generalized linear model has not yet seen in the literature, this paper attempts to study it.

2. Local Influence

Let α be an unknown *k*-dimensional parameter, whose domain is an open subset of Euclidean space R^k. $l(\alpha)$ is a object function(for example, likelihood function, punishment log-likelihood function). ω is a *n*-vector which denotes disturbed factor for example weightd or tiny shift. Let $M(\omega)$ is the disturbed model, whose object function is $l(\alpha|\omega)$. $\hat{\alpha}_\omega$ is the estimate which is from $M(\omega)$.Given ω_0 makes $l(\alpha|\omega_0)=l(\alpha)$ and $\hat{\alpha}=\hat{\alpha}_{\omega_0}$, where $l(\alpha|\omega)$ has continuous second-order partial derivatives, $l(\hat{\alpha}_\omega)$ is the function of ω. In geometry, $l(\hat{\alpha}_\omega)$ denotes *n*-dimentional surface

$$\eta(\omega)=\left(\omega^T, l(\hat{\alpha}_\omega)\right)^T \tag{1}$$

This image is called influence image, which vary with ω. The variation rate in ω_0 of influence image reflects that the sensitivity of model where ω_0 corresponds to the primary model. This method is called local influence [5]. Cook advanced that utilize influence curvature to measure the change of influence image near ω_0. Cook (1986) and Wei Bocheng (1990) pointed out that the influence curvature of $\eta(\omega)$ is given by

$$C_d = 2\left|d^T D^T \ddot{l} D d\right| = 2\left|d^T \Delta^T \ddot{l}^{-1} \Delta d\right| \tag{2}$$

where $\ddot{l}$ is the second derivatives of $l(\alpha)$ with respect to α, and

$$D=\frac{\partial \hat{\alpha}_\omega}{\partial \omega}, \Delta=\frac{\partial^2 l(\alpha|\omega)}{\partial \alpha \partial \omega^T} \tag{3}$$

D and Δ is $k\times n$ matrix, where $\alpha=\hat{\alpha}$, $\omega=\omega_0$. The influence matrix is given by

$$F = D^T \ddot{l} D = \Delta^T \ddot{l}^{-1} \Delta \tag{4}$$

Formula (2.4) shows that the maximal influence curvature $C_{\max}=2\lambda_1$, where λ_1 is the eigenvalue of $\ddot{F}$ whose absolute value is maximal, and $d_{\max}$ is the corresponding eigenvector which is called the direction of maximal influence curvature. Escobar and Meeker (1992) pointed out that the diagonal value of influence matrix also is the important diagnostic statistics.

3. Local Influence Analysis of Model

Considering non-parametric regression model

$$\eta_i = g(\mu_i) = x_i^T \beta, EY_i=\mu_i=b'(\theta_i), Y_i \sim ED(\theta_i,\phi), \quad i=1,2,\cdots,n \tag{1}$$

where $(y_i, x_i^T)^T$ $(i=1,2,\cdots,n)$ is the measure observations. $x_i^T=(x_i 1, x_i 2,\cdots,x_i p)$, Y_i $(i=1,2,\cdots,n)$ is i.i.d.. $Y_i \sim ED(\theta_i,\phi)$ denote that Y_i submit to exponential distributions, the corresponding density function is

$$p(y_i;\theta_i,\phi)=\exp\left\{y_i\theta_i - b(\theta_i)/a_i(\phi)+c(y_i,\phi)\right\} \tag{2}$$

where $\theta\in\Theta$, $a_i(\phi)$, $b(\theta_i)$ and $c(y_i,\phi)$ are known functions, $EY_i=\mu_i=b'(\theta_i)$, $Var(Y_i)=\sigma^2=b''(\theta_i)a_i(\phi)$, $b'(\theta_i)$ and $b''(\theta_i)$ are the first and second derivatives of $b(\theta_i)$, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is *p*-dimension parameter, η_i is the linear predict vector, $g(.)$ is univariate increasing function. Let $l(\beta)$ is the log-likelihood function of β.

$$l(\beta) = \phi^{-1} \sum_{i=1}^{n} \left[y_i \theta_i - b(\theta_i) \right] \tag{3}$$

Let $\dot{l}(\beta)$ and $\ddot{l}(\beta)$ are the first and second derivatives of $l(\beta)$ with respect to β, then

$$\dot{l}(\beta) = \phi^{-1} \sum_{i=1}^{n} \left[y_i - b'(\theta_i) \right] \frac{\partial \theta_i}{\partial \beta} \tag{4}$$

$$\ddot{l}(\beta) = \phi^{-1} \sum_{i=1}^{n} \left[\left(y_i - b'(\theta_i) \right) \frac{\partial^2 \theta_i}{\partial \beta^2} - b''(\theta_i) \left(\frac{\partial \theta_i}{\partial \beta} \right)^2 \right] \tag{5}$$

Supposed that the MLE of β in (3.1) is $\hat{\beta}$, and $\hat{\beta}$ submits to

$$\varnothing^{-1} \sum_{i=1}^{n} \left[y_i - b'(\theta_i) \right] \frac{\partial \theta_i}{\partial \beta} = 0$$

3.1. Weighted Perturbation Model

Suppose that $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^T$, $\omega_0 = (1, 1, \cdots, 1)^T$, then the weighted perturbation model can be shown that

$$l(\beta|\omega) = \phi^{-1} \sum_{i=1}^{n} \omega_i \left[y_i \theta_i - b(\theta_i) \right] \tag{3.1.1}$$

Substituting this result into (2.3) yields

$$\Delta = \left. \frac{\partial^2 l(\beta|\omega)}{\partial \beta \partial \omega^T} \right|_{\hat{\beta}, \omega_0} = \phi^{-1} \theta' T \tag{3.1.2}$$

where $T = diag\left(y_1 - b'(\theta_1), y_2 - b'(\theta_2), \cdots, y_n - b'(\theta_n) \right)$, $\theta' = \left(\frac{\partial \theta_1}{\partial \beta}, \frac{\partial \theta_2}{\partial \beta}, \cdots, \frac{\partial \theta_n}{\partial \beta} \right)$. The second derivatives of $l(\beta|\omega)$ with respect to β is given by

$$\ddot{l} = \phi^{-1} \sum_{i=1}^{n} \left[\left(y_i - b'(\theta_i) \right) \frac{\partial^2 \theta_i}{\partial \beta^2} - b''(\theta_i) \left(\frac{\partial \theta_i}{\partial \beta} \right)^2 \right] \tag{3.1.3}$$

where $b'(\theta_i) b''(\theta_i) \frac{\partial \theta_i}{\partial \beta}$ and $\frac{\partial^2 \theta_i}{\partial \beta^2}$.

Substituting (3.1.2) and (3.1.3) into (2.4), we obtain the corresponding influence matrix

$$F_\omega(\beta) = \phi^{-1} (\theta' T)^T \cdot \left\{ \sum_{i=1}^{n} \left[\left(y_i - b'(\theta_i) \right) \frac{\partial^2 \theta_i}{\partial \beta^2} - b''(\theta_i) \left(\frac{\partial \theta_i}{\partial \beta} \right)^2 \right] \right\}^{-1} \theta' T \tag{3.1.4}$$

Here d_ω denotes the direction of maximal influence curvature.

3.2. Response Variable Perturbation Model

Suppose that $Y_\omega = Y + \omega$, $\omega_0 = (0, 0, \cdots, 0)^T$, then the response variable perturbation model can be shown that

$$l(\beta|\omega) = \phi^{-1} \sum_{i=1}^{n} \left[(y_i + \omega_i) \theta_i - b(\theta_i) \right] \tag{3.2.1}$$

Substituting this result into (2.3) yields

$$\Delta = \left. \frac{\partial^2 l(\beta|\omega)}{\partial \beta \partial \omega^T} \right|_{\hat{\beta}, \omega_0} = \phi^{-1} \theta' \tag{3.2.2}$$

The second derivatives of $l(\beta|\omega)$ with respect to β is given by

$$\ddot{l} = \phi^{-1} \sum_{i=1}^{n} \left[\left(y_i - b'(\theta_i) \right) \frac{\partial^2 \theta_i}{\partial \beta^2} - b''(\theta_i) \left(\frac{\partial \theta_i}{\partial \beta} \right)^2 \right] \tag{3.2.3}$$

Substituting (3.2.2) and (3.2.3) into (2.4), we obtain the corresponding influence matrix

$$F_r(\beta) = \phi^{-1} (\theta')^T \cdot \left\{ \sum_{i=1}^{n} \left[\left(y_i - b'(\theta_i) \right) \frac{\partial^2 \theta_i}{\partial \beta^2} - b''(\theta_i) \left(\frac{\partial \theta_i}{\partial \beta} \right)^2 \right] \right\}^{-1} \theta' \tag{3.2.4}$$

Here d_r denotes the direction of maximal influence curvature.

4. An Illustrative Example

(Kyphosis Data) Now we consider an example as the illustration for the above results. Considering a kyphosis data (see [6]). There are 81 patients who have been treated with chiropractic. There are four variables: kyphosis, Age, Number and Start. Wang xiaoming (2005) ultized a linear semi-parametric model to fit this test data. The regression analysis of kyphosis data are as follows (**Table 1**).

The local influence analysis results of kyphosis data are as follows (from **Figures 1-3**).

Figures 1 and **2** show that the sixth, forty-third, fifty-third and the eightieth data are influential points, **Figure 3** shows that the first, second, third and fourth data are influential points. Actually, the direction of maximal influence curvature d_r also shows that the first, second,

Table 1. The regression analysis of kyphosis data.

Coefficients	β_0	β_1	β_2	β_3
Value	−2.0369	0.0109	0.4106	−0.2605
Std. error	1.4492	0.0064	0.2248	0.0677
T value	−1.4056	1.6962	1.8266	−3.0510

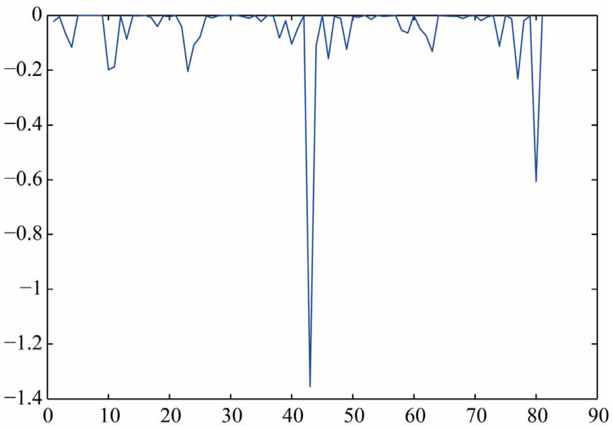

Figure 1. The diagonal value of influence matrix F_ω.

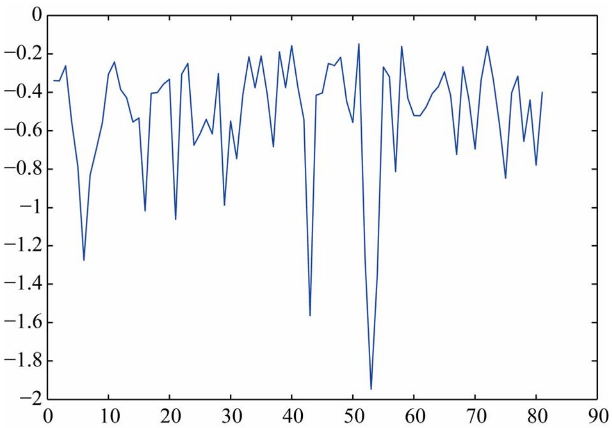

Figure 2. The diagonal value of influence matrix F_r.

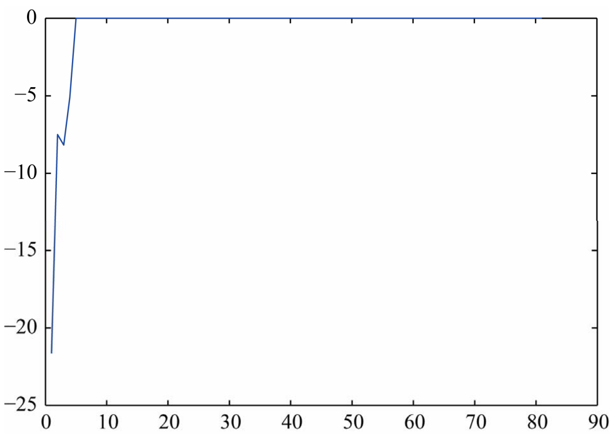

Figure 3. The direction of maximal influence curvature d_ω.

third and fourth data are influential points. This also proves that the above method is effective.

REFERENCES

[1] R. D. Cook, "Assessment of Local Influence (with Discussion)," *Journal of the Royal Statistical Society* (*Series B*), Vol. 48, 1986, pp. 133-169.

[2] R. D. Cook and S. Weisberg, "Residuals and Influence in Regression," Chapman and Hall, New York, 1982.

[3] B. C. Wei, G. B. Lu and J. Q. Shi, "Statistical Diagnostics," Publishing House of Southeast University, Nanjing, 1990.

[4] Y. Zhou, "Diagnostic and Numerical Example Analysis for Generalized Linear Model," *Journal of Sichuan University* (*Natural Science Edition*), Vol. 44, 2007, pp. 1163-1168.

[5] L. A. Escobar and W. Q. Meeker, "Assessing Influence in Regression Analysis with Censored Data," *Biometrics*, Vol. 48, No. 2, 1992, pp. 507-528. doi:10.2307/2532306

[6] X. M. Wang and X. L. Han, "The Course of S-plus Applied Statistics," Publishing House of Shanghai University of Finance and Economics, Shanghai, 2005.