应用概率统计　第二十六卷
第三期　2010年6月

Chinese Journal of Applied Probability
and Statistics Vol.26 No.3 Jun. 2010

On Confidence Region of Nonlinear Models for Failure Time Data *

ZONG XUPING

(*School of Mathematics, Yangzhou University, Yangzhou, 225002*)

Abstract

This paper proposes a differential geometric framework for nonlinear models for Failure Time Data. The framework may be regarded as an extension of that presented by Bates & Wates for nonlinear regression models. As an application, we use this geometric framework to derive three kinds of improved approximate confidence regions for parameter and subset parameter in terms of curvatures. Several results such as Bates and Wates (1980), Hamilton (1986) and Wei (1998) are extended to our models.

Keywords: Confidence regions, curvature array, Fisher information, Score function, nonlinear models.

AMS Subject Classification: 62F25.

§1. Introduction

The problem of analyzing time to event data arises in a number of applied fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography. A common feature of these data sets is that they contain censored observations, especially right censored data. And there is an enormous literature on dealing with these data sets. One simple method proposed by Aitkin (1981), in this paper, we use this method to deal with nonlinear model for right censored data. This type of model has been used for analyzing correlated survival observations (Hougaard, 1986). For Cox's proportional hazards model (Cox, 1972) with a gamma-fraility, inference procedures has been proposed by Klein (1992). Zhang et. al (1998) consider inference for a semiparametric stochastic mixed model for longitudinal data, and Cai, Cheng and Wei (2002) discussed Cox model for analyzing univariate failure time data by using semiparametric mixed-models.

Bates & Wates (1980) proposed a differential geometric framework for analyzing statistical problems related to ordinary nonlinear regression models. In this paper, we modify

*The project supported by National Science Plan (2007BAC29B01).
Received February 27, 2009. Revised July 2, 2009.

the Bates & Wates geometric framework by using the inner product given by Fisher information so that the modified Bates & Wates geometric framework can be used for nonlinear models for Failure Time Data. We use this geometric framework to derive three kinds of improved approximate confidence regions for parameter and parameter subsets in terms of curvatures from a geometric viewpoint. Several results such as Bates and Wates (1980), Hamilton (1986), Lee and Nelder (1996) and Wei (1998) are extended.

§2. Life Time Nonlinear Regression Model

Consider a life-testing experiment in which $n+m$ items are put on test and m items still survive at the conclusion of the test. Suppose that Y is an $(m+n)\times 1$ observed vector of y_i. Let the life times y_i $(i=1,\cdots,n+m)$ be independently and normally distributed with mean μ_i and common variance σ^2, and without loss of generality that the last m lifetimes are censored because of termination of the experiment.

The means μ_i are related to design or explanatory variables x_i by

$$\mu_i = f(x_i,\beta), \tag{2.1}$$

where β is a $p\times 1$ unknown parameter vector defined in $\mathcal{B}$, $x_i^T=(x_{i1},\cdots,x_{ip})$ is observed vector. Denote $f(x,\beta)=(\mu_1,\cdots,\mu_{n+m})^T$.

Let

$$\phi(y)=\frac{1}{\sqrt{2\pi}}e^{-1/2y^2},\qquad \varphi(y)=\int_y^\infty \phi(t)\mathrm{d}t,\qquad t_i=\frac{y_i-\mu_i}{\sigma},\qquad S(y)=\frac{\phi(y)}{\varphi(y)}.$$

The joint likelihood function is

$$L=\frac{1}{\sigma^n}\prod_{i=1}^n \phi(t_i)\prod_{i=n+1}^{n+m}\varphi(t_i),$$

then the joint log-likelihood function can be written as

$$l(\beta)=-\frac{n}{2}\log 2\pi\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-f(x_i,\beta))^2+\sum_{i=n+1}^{m+n}\log\varphi(t_i). \tag{2.2}$$

Differentiating the above formula to β, we get

$$\begin{aligned}\frac{\partial l}{\partial\beta} &= \frac{1}{\sigma^2}\sum_{i=1}^n(y_i-f(x_i,\beta))\frac{\partial f(x_i,\beta)}{\partial\beta}+\frac{1}{\sigma}\sum_{i=n+1}^{m+n}S(t_i)\frac{\partial f(x_i,\beta)}{\partial\beta}\\ &= \frac{1}{\sigma^2}\Big\{\sum_{i=1}^n\frac{\partial f(x_i,\beta)}{\partial\beta}(y_i-f(x_i,\beta))+\frac{1}{\sigma^2}\sum_{i=n+1}^{m+n}\frac{\partial f(x_i,\beta)}{\partial\beta}[\sigma S(t_i)+f(x_i,\beta)-f(x_i,\beta)]\Big\}\\ &= \frac{1}{\sigma^2}\Big[\sum_{i=1}^{n+m}\frac{\partial f(x_i,\beta)}{\partial\beta}(z_i-f(x_i,\beta))\Big],\end{aligned}$$

where

$$z_i = \begin{cases} y_i, & i = 1, 2, \cdots, n; \\ \sigma S(t_i) + f(x_i, \beta), & i = n+1, \cdots, n+m. \end{cases}$$

$$\dot{l}_\beta = \frac{1}{\sigma^2} D^T e, \tag{2.3}$$

where $e = Z - f(x, \beta)$, Z is the $n+m$ vector with the element z_i, $D = \partial^2 f(x, \beta)/(\partial\beta\partial\beta^T)$.

$$\begin{aligned} \ddot{l}_{\beta\beta} &= \frac{\partial^2 l}{\partial\beta\partial\beta^T} \\ &= -\frac{1}{\sigma^2}\Big\{\sum_{i=1}^{n} \frac{\partial f(x_i,\beta)}{\partial\beta}\frac{\partial f(x_i,\beta)}{\partial\beta^T} + \sum_{i=n+1}^{m+n} \frac{\partial f(x_i,\beta)}{\partial\beta} \cdot \frac{\phi'(t_i)\Phi(t_i) + \phi^2(t_i)}{\Phi^2(t_i)} \frac{\partial f(x_i,\beta)}{\partial\beta^T}\Big\} \\ &\quad + \frac{1}{\sigma^2}\Big\{\sum_{i=1}^{n} (y_i - f(x_i,\beta))\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T} + \sum_{i=n+1}^{m+n} \frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}\sigma S(t_i)\Big\}. \end{aligned}$$

$$\ddot{l}_{\beta\beta} = -\frac{1}{\sigma^2} D^T \Omega^{-1} D + \frac{1}{\sigma^2}[e^T][W], \tag{2.4}$$

where $\Omega = \text{diagonal}(v_1, v_2, \cdots, v_{n+m})$,

$$v_i = \begin{cases} 1, & i = 1, 2, \cdots, n; \\ \dfrac{\phi'(t_i)\Phi(t_i) + \phi^2(t_i)}{\Phi^2(t_i)}, & i = n+1, \cdots, n+m, \end{cases} \qquad W = \frac{\partial^2 f(x,\beta)}{\partial\beta\partial\beta^T},$$

and $[\cdot][\cdot]$ denotes the array multiplication, see Wei (1998) Appendix A for details.

We assume that regular conditions such as Wei (1998) for our model are satisfied, in particular, we assume that

$$\lim_n \frac{D^T \Omega^{-1} D}{n} = K(\beta).$$

Let $\widehat{\beta}$ be the maximum likelihood estimate of β, then it follows from (2.3) that $\widehat{\beta}$ satisfies

$$D^T(\widehat{\beta})\Omega^{-1}\widehat{e}^* = 0,$$

where $\widehat{e}^* = \Omega\widehat{e}$, $\widehat{e} = Z - f(x, \beta)|_{\widehat{\beta}}$.

The above equation shows that in Euclidean space R^{n+m}, the "residual vector" $\widehat{e}^*$ is orthogonal to the space spanned by column vectors of $D(\widehat{\beta})$ with respect to the matrix Ω^{-1} inner product. Combining this geometric interpretation with the geometric framework of nonlinear regression models presented by Bates and Wates (1980), we can introduce a modified BW geometric framework for our models (2.1) as follows.

Take $\eta = f(x, \beta)$ as a coordinate in Euclidean space R^n, then $\eta = f(x, \beta)$ may be called solution locus. It is easily seen that the tangent space T_β is spanned by the columns of $D(\beta)$. For any two vectors a and b in R^n, we define an inner product as $< a, b >= a^T\Omega^{-1}b$. Under this inner product, the corresponding normal space is denoted by T'_β. We can define curvature arrays for the solution locus $\eta = f(\beta)$, connected with the model (2.1). To this aim, we choose the orthogonal basis for spaces T_β. Suppose that the QR decomposition of $D(\beta)$ under inner product is given by

$$D(\beta) = (Q, N)\begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{2.5}$$

where R and $L = R^{-1}$ are $p \times p$ nonsingular upper triangular matrices and the columns of Q and N are orthogonal basis for the tangent space and the normal space of solus $\eta = f(\beta)$ at β. The matrices Q and N satisfy $Q^T\Omega^{-1}Q = I_p$, $Q^T\Omega^{-1}N = 0$, $N^T\Omega^{-1}N = I_{n+m-p}$, where I_p and I_{n+m-p} are identity matrices of order p and $n+m-p$, respectively.

Now we define the intrinsic curvature array A^I and parameter-effects curvature array A^P as:

$$A^I = [N^T\Omega^{-1}][U], \qquad A^P = [Q^T\Omega^{-1}][U], \qquad U = L^TWL. \tag{2.6}$$

Note that for an inner product space with weight Ω^{-1}, the projection operator of matrix D is $P_D = D(D^T\Omega^{-1}D)^{-1}D^T\Omega^{-1}$ and satisfies that $P_D^2 = P_D$ and $\Omega^{-1}P_D = P_D^T\Omega^{-1}$. Therefore $P_T = QQ^T\Omega^{-1}$ and $P_N = NN^T\Omega^{-1}$ are orthogonal projection operators of tangent space T_β and normal space T'_β, respectively. It is also easy to show that $U = [N][A^I] + [Q][A^P]$.

The geometric framework introduced above seems similar to that defined by Bates and Wates (1980), so it may be called the modified BW (MBW) geometric framework. But there are some differences between our MBW framework and BW framework.

§3. Confidence Regions in Terms of Curvature

Hamilton (1986) studied confidence regions for parameters in normal nonlinear models based upon the Bates & Watts (1980) geometric framework. They obtained quadratic approximations for the inference in terms of curvature measures. It is interesting that we can completely extend all the results to the nonlinear models with random effects based upon the modified Bates & Watts curvature measures.

A usual approximate confidence region of β for our models (2.1) is based on the likelihood ratio static

$$LR(\beta) = -2\{l(\beta) - l(\widehat{\beta})\}, \tag{3.1}$$

which is a function of β and asymptotically has a χ_p^2 for β.

To derive improved approximate projections of the solution locus inference region onto the tangent space, we introduce a nonlinear transformation for β as follows. In parameter space, the point β and $\widehat{\beta}$ map to vectors $\eta(\beta)$ and $\eta(\widehat{\beta})$, respectively. The projection of $\eta(\beta)-\eta(\widehat{\beta})$ onto the tangent space at $\widehat{\beta}$ is $t = QQ^T\Omega^{-1}\{\eta(\beta)-\eta(\widehat{\beta})\}$, where Q is evaluated at $\widehat{\beta}$, all the quantities such as Q, D, R, L are evaluated at $\widehat{\beta}$ which we omit. If the columns of Q are taken as an orthogonal basis for the tangent space at $\widehat{\beta}$, then the coordinates of projection t in the tangent space are

$$\tau(\beta) = Q^T\Omega^{-1}\{\eta(\beta)-\eta(\widehat{\beta})\}. \tag{3.2}$$

As a new parameter, $\tau=\tau(\beta)$ represents a nonlinear mapping from the parameter space to the tangent space and connects the solution locus and the tangent space. The coordinates τ provide a natural reference system for the solution locus and approximations to it. We may construct confidence regions for the parameter β in terms of coordinates τ by using quadratic approximations. The analogous transformation to (3.2) has been used for nonlinear regression models by Hamilton (1986). Notice that the transformation (3.2) gives an one-to-one mapping between β and τ in some neighborhood of $\widehat{\beta}$, and $\beta=\widehat{\beta}$ corresponding to $\tau=0$. We denote the inverse of $\tau=\tau(\beta)$ by $\beta=\beta(\tau)$.

3.1 Likelihood Region of Parameter

We can derive an improved approximate projection of the solution locus likelihood region onto the tangent space using transformation (3.2). For simplicity, we denote the log-likelihood $l(\beta)$ and the likelihood ratio statistic $LR(\beta)$ by $l(\tau)$ and $LR(\tau)$, respectively when $\beta=\beta(\tau)$ is considered. Similarly, we denote $l(\widehat{\beta})$, $\dot{l}(\widehat{\beta})$ and $\ddot{l}(\widehat{\beta})$ by $l(0)$, $\dot{l}(0)$ and $\ddot{l}(0)$, respectively, when $\beta=\beta(\tau)$ at $\widehat{\beta}$ (i.e. $\tau=0$). We may derive a quadratic approximation for $LR(\tau)$ in terms of the parameter $\tau=\tau(\beta)$ instead of the parameter β. To do so, we need the following lemma.

Lemma 3.1 For the models (2.1), the derivatives at $\widehat{\beta}$ of the functions of $\tau(\beta)$ and $\beta(\tau)$ defined (3.2) are given by

$$\frac{\partial\tau}{\partial\beta^T}=R, \qquad \frac{\partial^2\tau}{\partial\beta\partial\beta^T}=R^TA^PR. \tag{3.3}$$

$$\frac{\partial\beta}{\partial\tau^T}=L, \qquad \frac{\partial^2\beta}{\partial\tau\partial\tau^T}=-[L][A^P]. \tag{3.4}$$

Proof It is easy to get (3.3) from (3.2), so we just prove (3.4). Since $\partial\tau/\partial\tau^T = (\partial\tau/\partial\beta^T)(\partial\beta/\partial\tau) = I_p$, we have $\partial\beta/\partial\tau^T = R^{-1} = L$ and

$$\begin{aligned}\frac{\partial^2\tau}{\partial\tau\partial\tau^T} &= \Big(\frac{\partial\beta}{\partial\tau^T}\Big)^T\Big(\frac{\partial^2\tau}{\partial\beta\partial\beta^T}\Big)\Big(\frac{\partial\beta}{\partial\tau^T}\Big) + \Big[\frac{\partial\tau}{\partial\beta^T}\Big]\Big[\frac{\partial^2\beta}{\partial\tau\partial\tau^T}\Big] \\ &= L^T(R^TA^PR)L + [R]\Big[\frac{\partial^2\beta}{\partial\tau\partial\tau^T}\Big] = 0,\end{aligned}$$

which gives the second equation of (3.4). □

Theorem 3.1 For the model (2.1), the approximate tangent space projection of the solution locus likelihood region of β with level $100(1-\alpha)\%$ can be represented as

$$\tau^T(\beta)(I_p - B)\tau(\beta) \le \sigma^2\chi^2(p,\alpha), \tag{3.5}$$

where $B = [\widehat{e}^TN][A^I]$, and Q, N, A^I are all evaluated at $\widehat{\beta}$.

Proof Under the transformation (3.2), (3.1) can be represented as

$$LR(\beta) = -2\{l(\tau) - l(0)\} \approx -\tau^T(\beta)\ddot{l}(0)\tau(\beta), \tag{3.6}$$

where $\ddot{l}(0) = \partial^2 l/\partial\tau\partial\tau^T$ evaluated at $\tau = 0$ (i.e. $\beta = \widehat{\beta}$). It follows from Lemma 3.1 that

$$\ddot{l}(\tau) = \Big\{\Big(\frac{\partial\beta}{\partial\tau}\Big)^T\Big(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big)\Big(\frac{\partial\beta}{\partial\tau}\Big) + \Big[\Big(\frac{\partial l}{\partial\beta}\Big)^T\Big]\Big[\frac{\partial^2\beta}{\partial\tau\partial\tau^T}\Big]\Big\}. \tag{3.7}$$

From $D = QR$, $D^T\Omega^{-1}D = R^TR$, and Lemma 3.1 that

$$-\ddot{l}(\widehat{\beta}) = \frac{1}{\sigma^2}R^T\{I_p - [\widehat{e}^T][L^TWL]\}R,$$

and from (2.4)-(2.6), we have $[\widehat{e}^T][L^TWL] = [\widehat{e}^T][U] = [\widehat{e}][(NN^T\Omega^{-1} + QQ^T\Omega^{-1})U] = B$, hence

$$-\ddot{l}(\widehat{\beta}) = \sigma^{-2}R^T(I_p - B)R. \tag{3.8}$$

Substituting this equation into $LR(\beta)$ gives

$$LR(\beta) = \sigma^{-2}\tau^T(\beta)(I_p - B)\tau(\beta). \tag{3.9}$$

It follows from equation (3.8) that $I_p - B \ge 0$, therefore expression (3.5) shows that the approximate tangent space projection of the solution locus likelihood region is an ellipsoid which does not depend on the parameterizations. Our Theorem 3.1 for nonlinear models for failure time data is similar to the result obtained by Wei (1998) for the embedded models. If σ^2 is unknown, then (3.5) can be represented as $\tau^T(\beta)(I_p - B)\tau(\beta) \le \delta^2$, and δ^2 is decided by $F(p, n-p, 1-\alpha)$. □

3.2 Confidence Region for Parameter Subsets Based Upon Likelihood Ratio

If a subset of parameters is of primary interest as discussed by Hamilton, often the parameter vector β cab be partitioned as $\beta^T = (\beta_1^T, \beta_2^T)$, where the last k parameters β_2 are of interested. Further, parton $\tau^T = (\tau_1^T, \tau_2^T)$, $D = (D_1, D_2)$, $R = (R_{ij})$, and $B = (B_{ij})$ $(i, j = 1, 2)$, to confirm to the portioning of β, similar partings are used later.

The likelihood ratio statistic corresponding to β_2 is similar to (3.1) and given by

$$LR_s(\beta_2) = -2\{l(\widetilde{\beta}) - l(\widehat{\beta})\}, \tag{3.10}$$

where $\widetilde{\beta}^T = (\widetilde{\beta}_1^T(\beta_2), \beta_2^T)$ and $\widetilde{\beta}_1^T(\beta_2)$ maximizes $l(\beta)$ for each value of β_2. The function $LR_s(\beta_2)$, which is analogous to $LR(\beta)$, depends on parameters β_2 and asymptotically has distribution $\chi^2(k)$ for each β_2 (see [11]-[13]). To obtain an improved approximate likelihood region for the parameter subset β_2, the transformation (3.2) can also be used. In this case, (3.2) has the form

$$\widetilde{\tau} = \tau(\widetilde{\beta}) = Q^T\Omega^{-1}\{\eta(\widetilde{\beta}) - \eta(\widehat{\beta})\}, \tag{3.11}$$

where $\widetilde{\tau} = \tau(\widetilde{\beta})$ is a function of β_2. Form (3.11) we have Theorem 3.2.

Theorem 3.2 For the nonlinear models with random effects stated above, the approximate tangent space projection of the solution locus likelihood region of β with level $100(1-\alpha)\%$ can be represented as

$$\widetilde{\tau}_2^T(I_k - T)\widetilde{\tau}_2 \le \sigma^{-2}\chi^2(k, \alpha), \tag{3.12}$$

where $T = B_{22} + B_{21}(I_k - B_{11})^{-1}B_{12}$, $\widetilde{\tau}^T = (\widetilde{\tau}_1^T, \widetilde{\tau}_2^T)$.

Proof It is easily seen that equation (3.9) still holds at and it can be represented as

$$LR_s(\beta_2) = -2\{l(\widetilde{\beta}) - l(\widehat{\beta})\} \approx \widetilde{\tau}^T(I_p - B)\widetilde{\tau}/\sigma^2. \tag{3.13}$$

Form this equation, we may derive an approximate relationship between $\widetilde{\tau}_1$ and $\widetilde{\tau}_2$. In fact, the approximations to (3.2) and (3.11) give $\tau \approx Q^T\Omega^{-1}D(\beta - \widehat{\beta}) = R(\beta - \widehat{\beta})$ and $\widetilde{\tau} = R(\widetilde{\beta} - \widehat{\beta})$, respectively. Then the components of $\widetilde{\tau}$ are give by $\widetilde{\tau}_2 = R_{22}(\beta_2 - \widehat{\beta}_2)$ and $\widetilde{\tau}_1 = R_{11}(\widetilde{\beta}_1 - \widehat{\beta}_1) + R_{12}(\beta_2 - \widehat{\beta}_2)$, respectively. On the other hand, it follow from (3.8) that $\dot{l}(\widetilde{\beta}) \approx \ddot{l}(\widehat{\beta})(\widetilde{\beta} - \widehat{\beta}) = -\sigma^{-2}R^T(I - B)\widetilde{\tau}$, that is $-\sigma^{-2}L^T\dot{l}(\widetilde{\beta}) = \sigma^{-2}(I - B)\widetilde{\tau}$. Since $(\partial l/\partial\beta_1)_{\widetilde{\beta}} = 0$, this leads to

$$\widetilde{\tau}_1 = (I - B_{11})^{-1}B_{12}\widetilde{\tau}_2 \qquad (\widetilde{\tau}_2 = \tau_2). \tag{3.14}$$

Therefore we have

$$LR_s(\beta_2) = \widetilde{\tau}_2(I_k - T)\widetilde{\tau}_2/\sigma^2,$$

which implies (3.12) and the theorem is proved. □

Notice that both the expression and the geometric interpretation of (3.12) are very similar to those of Hamilton (1986) for normal nonlinear regression models. Our Theorem 3.2 can be used for general classes.

3.3 Confidence Region for Parameter Subsets Based on the Score Statistic

The score statistic can be used to construct confidence regions for parameter subsets as discussed by Hamilton (1986) for nonlinear regression models. For our model (2.1), the score statistic associated with β is

$$SC = \Big\{\Big(\frac{\partial l}{\partial \beta_2}\Big)^T J^{22}\Big(\frac{\partial l}{\partial \beta_2}\Big)\Big\},$$

where J^{22} is the lower right corner of parting $J^{-1}(Y) = (J^{ij})$ $(i,j = 1,2)$ and $J(Y) = \sigma^{-2}D^T\Omega^{-1}D$. SC asymptotically has the $\chi^2(k)$ distribution for each β_2. To get a quadratic approximation of SC in terms of the curvature, we first give a lemma.

Lemma 3.2 Let $P = D(D^T\Omega^{-1}D)^{-1}D^T\Omega^{-1}$, $P_1 = D_1(D_1^T\Omega^{-1}D_1)^{-1}D_1^T\Omega^{-1}$ and $\widetilde{e} = e(\widetilde{\beta})$, then

$$SC = \sigma^{-2}\widetilde{e}^T(\widetilde{P} - \widetilde{P}_1)\widetilde{e}, \tag{3.15}$$

where $\widetilde{\Omega}$, $\widetilde{P}$ and $\widetilde{P}_1$ are all evaluated at $\widetilde{\beta}$.

Proof It is easy to show from (2.3) and (2.4) that $\partial l/\partial\beta_2 = D_2^T e$, $J^{22} = (D_2^T\Omega^{-1} \cdot P_1' D_2)^{-1}\sigma^{-2}$ and

$$SC = \sigma^{-2}\widetilde{e}^T D_2(D_2^T\Omega^{-1}P_1'D_2)^{-1}D_2^T\widetilde{e}, \tag{3.16}$$

where $P_1' = I - P_1$. Since $D_1^T\widetilde{e} = 0$, $\widetilde{e}$ is orthogonal to the columns of D_1 and $P_1'\widetilde{e} = \widetilde{e}$ holds. Notice that $(P_1')^T\Omega^{-1} = \Omega^{-1}P_1'$ hold: substitution of these results into the above equation gives

$$SC = \sigma^{-2}\{e^T P_1'^T D_2(D_2^T\Omega^{-1}P_1'D_2)^{-1}D_2^T e\}_{\beta=\widetilde{\beta}} = \sigma^{-2}\{e^T(P - P_1)e\}_{\beta=\widetilde{\beta}},$$

where we use the fact that $P - P_1$ is equal to the projection operator of $P_1'D_2$. □

Lemma 3.3 If $\beta = \beta(\tau)$ is determined by (3.2), then the quadratic approximation of $\beta(\tau)$ can be represented as

$$\eta(\beta(\tau)) \approx \eta(\widehat{\beta}) + Q\tau + (1/2)N(\tau^T A^I \tau), \tag{3.17}$$

where Q, N and A^I are all evaluated at $\widehat{\beta}$.

Proof It follows from Lemma 1 that

$$\begin{aligned}
\frac{\partial^2 \eta}{\partial \tau \partial \tau^T} &= \Big(\frac{\partial \beta}{\partial \tau^T}\Big)^T \Big(\frac{\partial^2 \eta}{\partial \beta \partial \beta^T}\Big)\Big(\frac{\partial \beta}{\partial \tau^T}\Big) + \Big[\frac{\partial \eta}{\partial \beta^T}\Big]\Big[\frac{\partial^2 \eta}{\partial \tau \partial \tau^T}\Big] \\
&= L^T W L + [D][-[L][A^P]] \\
&= U - [Q][A^P] = [N][A^I],
\end{aligned}$$

then (3.17) can be obtained by using the second order Taylor series expansion for $\eta(\beta(\tau))$ at $\tau = 0$. □

Theorem 3.3 For the model (2.1) in Section 2, the approximate tangent space projection of the solution locus inference region of β based on the score statistic with level $100(1-\alpha)\%$ can be represented as

$$\widetilde{\tau}_2^T(\beta)(I_k - T)^2 \widetilde{\tau}_2(\beta) \leq \sigma^2 \chi^2(k, \alpha), \tag{3.18}$$

where T is defined in Theorem 3.2.

Proof By Lemma 3.2 and Lemma 3.3, we first use $\overline{D}$ to calculate

$$\widetilde{e}^{*T} \Omega^{-1} P \widetilde{e}^* = \widetilde{e}^{*T} \Omega^{-1} \overline{D}(\overline{D}^T \Omega^{-1} \overline{D})^{-1} \overline{D}^T \Omega^{-1} \widetilde{e}^*,$$

where $\overline{D} = D(\widetilde{\beta})$. It follows from (3.17) that at $\beta = \widetilde{\beta}$ (i.e. $\tau = \widetilde{\tau}$), we have

$$\overline{D} = Q + NM \quad (M = A^I \widetilde{\tau}) \qquad \text{and} \qquad (\overline{D}^T \Omega^{-1} \overline{D})^{-1} = (I + M^T M)^{-1},$$
$$\widetilde{e}^* = \widehat{e}^* - Q\widetilde{\tau} - \frac{1}{2} N \widetilde{\tau^T} A^I \widetilde{\tau}.$$

By a little calculation, we can get (3.18). □

The expression and geometric interpretation of Theorem 3.3 are very similar to those given by Halmilton (1986, P.60) for nonlinear models. For the case when there is no nuisance parameter under consideration, then $\tau_2 = \tau$, $T = B$ and (3.18) reduces to

$$\tau^T(\beta)(I_p - B)^2 \tau(\beta) \leq \sigma^2 \chi^2(p, \alpha),$$

which may be considered as an alternative confidence region for β.

The geometric framework is introduced in this paper while the parameter β and parameter subsets are interested, and some asymptotic inference may be studied based upon this framework.

References

[1] Aitkin, M., A note on the regression analysis of censored data, *Techometrics*, **23**(1981), 161–163.

[2] Bates, D.M. and Wates, D.G., Relative curvature measures of nonlinearity, *J. Roy. Statist. Soc. Ser. B*, **42**(1980), 1–25.

[3] Breslow, N.E. and Calyton, D.G., Approximate inference in generalized linear mixed models, *JASA*, **88**(1993), 9–25.

[4] Cai, T., Cheng, S.C. and Wei, L.J., Semiparametric mixed-effects models for cluster failure time data, *Journal of American Statistical Association*, **97**(2002), 514–522.

[5] Cook, R.D. and Tsai, C.L., Residuals in nonlinear regression, *Biometrika*, **72**(1985), 23–29.

[6] Cox, D.R., Regression models and life tables (with discussion), *J. Roy. Statist. Soc. Ser. B*, **34**(1972), 187–220.

[7] Efron, B., The geometry of exponential families, *Ann. Staist.*, **3**(1975), 1189–1242.

[8] Hamilton, B.C., Confidence regions for parameter subsets in nonlinear regression, *Biometrika*, **73** (1986), 57–64.

[9] Klein, J.P., Semiparametric estimation of random effects using Cox model based on EM algorithm, *Biometrics*, **48**(1992), 795–806.

[10] Lee, Y. and Nelder, N.A., Hierarchical generalized linear models, *J. Roy. Statist. Soc. Ser. B*, **58** (1996), 619–678.

[11] Robinson, G.K., That BLUP is a good thing: the estimation of random effects, *Statist. Sci.*, **6**(1991), 15–51.

[12] Wei, B.C., On confidence regions of embeded models in regular parametric families (A geometric approach), *Austl. J. Statist.*, **36**(1994), 327–338.

[13] Wei, B.C., Some asymptotic inference in multinomial nonlinear models, *Appl. Math. - JUC*, **10B** (1995), 273–284.

[14] Wei, B.C., *Exponential Family Nonlinear Models*, Singapore: Springer-Verlag, 1998.

带寿命数据非线性随机效应模型的置信域

宗 序 平

(扬州大学数学科学学院, 扬州, 225002)

本文对带寿命数据非线性随机效应模型, 建立了微分几何框架, 推广了Bates & Wates关于非线性模型几何结构. 在此基础上, 我们导出了关于固定效应参数和子集参数的置信域的曲率表示, 这些结果是Bates and Wates (1980), Hamilton (1986)和Wei (1998)等的推广.

关键词: 置信域, 曲率立体阵, Fisher信息, Score函数, 非线性模型.

学科分类号: O212.1.