第 34卷第 2期
2011年 6月
南京师大学报(自然科学版)
JOURNAL OF NANJING NORMAL UNIVERSITY (Natural Science Edition)
Vol 34 No. 2
June, 2011

Bayes条件下非线性模型的参数置信域的曲率表示

吴一凡

(江苏食品职业技术学院基础教学部,江苏 淮安 223003)

[摘要] 给出了 Bayes条件下非线性模型的参数估计,建立了该模型的几何结构,推广了 Bates & Wates关于非线性回归模型的几何框架,在此基础上,用几何方法探讨了关于参数与子集参数置信域的曲率表示.

[关键词] Bayes条件下非线性模型,参数估计,几何结构,置信域,曲率

[中图分类号] O212.5 [文献标志码] A [文章编号] 1001-4616(2011)02-0027-07

Confidence Regions of Parameters in Terms of Curvatures in Nonlinear Models With Bayes Conditions

Wu Yifan

(Department of Basic Courses, Jiangsu Food Science College, Huai'an 223003, China)

Abstract In this paper, we propose a differential geometric framework for nonlinear models with Bayes conditions. The framework may be regarded as an extension of that presented by Bates & Watts for nonlinear regression models. On this basis, we use this geometric framework to discuss about the confidence regions for parameter and parameter subset in terms of curvature.

Key words nonlinear model with Bayes conditions, estimates of parameters, geometric framework, confidence regions, curvature

Efron[1]首次利用几何方法研究了估计量的渐近理论,促使统计学在上世纪 80年代飞速发展. Bates and Watts[2]在 Euclid空间引入了非线性回归模型的曲率度量,由于 Euclid空间比较直观,回归模型更加重要,许多人沿着这一方向进行了卓有成效的研究工作. Hamilton等[3]利用 Bates和 Watts[2]的几何结构,研究了参数的置信域,获得了似然置信域的曲率表示. Hamilton[4]获得了基于似然比和子集参数置信域的曲率表示. 唐年胜、韦博成、王学仁[5]对非线性再生散度模型在 Euclid空间建立了几何结构,并研究了该模型参数和子集参数的 3种置信域. Zhong[6]对非线性随机效应模型在 Euclid空间建立了几何结构,并研究了该模型参数和子集参数的 3种置信域. 本文在 Bayes条件下对非线性回归模型建立了类似 Bates和 Watts的几何结构,在此基础上讨论了 Bayes条件下非线性模型的参数与子集参数的置信域的曲率表示.

1 Bayes条件下非线性模型的参数估计

Bayes条件下非线性回归模型可表示为:

$$\boldsymbol{Y} = f(\boldsymbol{X}, \beta) + \varepsilon, \ \varepsilon \sim N(0, \sigma^2 \boldsymbol{I}_n), \ \beta \sim N(0, \sigma^2 \Sigma), \tag{1}$$

其中, $\boldsymbol{Y} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ 是 $n \times 1$观测向量, Σ为 $p \times p$阶正定阵, $\beta = (\beta_0, \beta_1, \cdots, \beta_{p-1})^{\mathrm{T}}$ 为 $p \times 1$未知参数向量, ε为 $n \times 1$随机误差向量. 模型(1)关于 $\boldsymbol{Y}$和 β的联合概率密度函数可以表示为

$$f(\boldsymbol{Y}, \beta, \sigma^2) = f_{\boldsymbol{Y}}(\boldsymbol{Y} \mid \beta, \sigma^2) p_{\beta}(\beta \mid \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^{n+p} \mid \Sigma \mid^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2\sigma^2} \| \boldsymbol{Y} - f(\boldsymbol{X}, \beta) \|^2 - \frac{1}{2\sigma^2} \beta^{\mathrm{T}} \Sigma^{-1} \beta}. \tag{2}$$

收稿日期: 2011-02-25

通讯联系人: 吴一凡,讲师,研究方向:应用概率统计. E-mail: wyfdm_1@163.com

模型(1)的对数联合概率密度函数可以表示为

$$l(Y;\beta,\sigma^2)=-\ln\left((\sqrt{2\pi}\sigma)^{n+p}\mid\Sigma\mid^{\frac{1}{2}}\right)-\frac{1}{2\sigma^2}\left[\|Y-f(X,\beta)\|^2+\beta^{\mathrm{T}}\Sigma^{-1}\beta\right]=$$

$$-\ln\left((\sqrt{2\pi}\sigma)^{n+p}\mid\Sigma\mid^{\frac{1}{2}}\right)-\frac{1}{2\sigma^2}\left[(Y-f(X,\beta))^{\mathrm{T}}(Y-f(X,\beta))+\beta^{\mathrm{T}}\Sigma^{-1}\beta\right].\tag{3}$$

(3)式对 β 求导,得到模型的 Score函数为

$$\dot{l}(\beta)=\frac{1}{\sigma^2}(\boldsymbol{D}_\beta^{\mathrm{T}}[\boldsymbol{Y}-f(\boldsymbol{X},\beta)]-\Sigma^{-1}\beta),$$

进一步可写成

$$\dot{l}(\beta)=\frac{1}{\sigma^2}(\boldsymbol{D}_\beta^{\mathrm{T}}\Sigma^{-\frac{1}{2}})\begin{pmatrix}\boldsymbol{Y}-f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}\Bigg|_{\hat{\beta}},\tag{4}$$

其中 $\boldsymbol{D}_\beta=\frac{\partial f(\boldsymbol{X},\beta)}{\partial\beta}$. 模型(1)的观察信息阵为

$$-\ddot{l}(\beta)=\frac{1}{\sigma^2}(\boldsymbol{D}_\beta^{\mathrm{T}}\boldsymbol{D}_\beta+\Sigma^{-1})-\frac{1}{\sigma^2}[\boldsymbol{e}^{\mathrm{T}}]\cdot[\boldsymbol{W}],\tag{5}$$

其中 $\boldsymbol{e}=\begin{pmatrix}\boldsymbol{Y}-f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}$ 为 $n+p$ 维向量, $\boldsymbol{W}=\frac{\partial^2\eta(\beta)}{\partial\beta\partial\beta^{\mathrm{T}}}$ 为 $(n+p)\times p\times p$ 的立体矩阵, $\eta(\beta)=\begin{pmatrix}f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}$ 为 $n+p$ 维向量,方括号乘积 $[\cdot][\cdot]$ 可参见[7].

2　非线性回归模型的几何结构

对于模型(1),参数 β 的估计满足方程(4),若记 $\boldsymbol{D}=\begin{pmatrix}\boldsymbol{D}_\beta\\ \Sigma^{-\frac{1}{2}}\end{pmatrix}_{(n+p)\times p}$,且 $\boldsymbol{D}$ 在 β 处记值, $\boldsymbol{e}=\begin{pmatrix}\boldsymbol{Y}-f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}_{(n+p)\times 1}$,则(4)式变为:

$$\dot{l}(\beta)=\frac{1}{\sigma^2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}=0,$$

即有

$$\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}=0.\tag{6}$$

在 $\mathbf{R}^n$ 欧氏空间中,两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的内积为 $(\boldsymbol{a},\boldsymbol{b})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$, $\hat{\boldsymbol{e}}$ 垂直于 $\boldsymbol{D}(\hat{\beta})$ 列向量生成的子空间,因而垂直于曲面 $\eta=\begin{pmatrix}f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}$ 的切空间. 由此,我们提出以下几何结构:

在 $\mathbf{R}^{n+p}$ 欧氏空间中,定义 p 维曲面 π: $\eta=\begin{pmatrix}f(\boldsymbol{X},\beta)\\ \Sigma^{-\frac{1}{2}}\beta\end{pmatrix}_{(n+p)\times 1}$. 令

$$\boldsymbol{C}_a:\boldsymbol{C}_a(t)=\eta(\beta_1,\dots,\beta_{a-1},\beta_a+t,\beta_{a+1},\dots,\beta_p),\ a=1,\dots,p.$$

其中 t 是实数, $\boldsymbol{C}_a$ 称为 π 在 β 处的坐标曲线. 显然 $\boldsymbol{C}_a(t)$ 在 β(即 $t=0$)处的切向量为 $\boldsymbol{C}_a(t)$ 在 $t=0$ 处的导数 $\dot{\boldsymbol{C}}_a(0)$. 易见, $\dot{\boldsymbol{C}}_a(0)(a=1,2,\dots,p)$ 正是 $\boldsymbol{D}(\beta)=\frac{\partial\eta(\beta)}{\partial\beta^{\mathrm{T}}}$ 的列向量. 由坐标曲线 $\boldsymbol{C}_a(t)$ 的切向量 $\dot{\boldsymbol{C}}_a(0)(a=1,2,\dots,p)$ 生成的子空间称为 π 在 β 处的切空间,记 $\boldsymbol{T}_\beta$. 假定 $\boldsymbol{D}(\beta)$ 为列满秩阵,则 $\boldsymbol{T}_\beta$ 是 $\mathbf{R}^{n+p}$ 的 p 维线性子空间,并且可认为是 π 在 β 处的线性逼近. 令 $\boldsymbol{T}_\beta$ 的法空间为 $\boldsymbol{T}'_\beta$,则 $\boldsymbol{T}'_\beta$ 是 $\mathbf{R}^{n+p}$ 的 n 维线性子空间,我们能得到 $\boldsymbol{T}_\beta$ 和 $\boldsymbol{T}'_\beta$ 的正交基. 在上述内积意义下,设 $\boldsymbol{D}$ 的 $\boldsymbol{DR}$ 分解为

$$\boldsymbol{D}=(\boldsymbol{Q},\boldsymbol{N})\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}=\boldsymbol{QR},\tag{7}$$

其中 $\boldsymbol{R}$ 为 p 阶具有正对角元的上三角矩阵, $\boldsymbol{Q},\boldsymbol{N}$ 分别为 $(n+p)\times p$ 和 $(n+p)\times n$ 阶列正交矩阵,其列向量构成 π 在 β 处的切空间和法空间的标准正交基,即 $\boldsymbol{Q},\boldsymbol{N}$ 满足

$$Q^{\mathrm{T}}Q = I_p,\ Q^{\mathrm{T}}N = 0,\ N^{\mathrm{T}}N = I_n, \tag{8}$$

其中 I_p 和 I_n 分别表示 p 阶和 n 阶单位矩阵.

定义模型 (1) 的固有曲率立体阵 A^I 和参数效应曲率立体阵 A^P 分别为

$$A^I = [N^{\mathrm{T}}][U],\ A^P = [Q^{\mathrm{T}}][U], \tag{9}$$

其中 $L = R^{-1}$, $U = L^{\mathrm{T}}WL$, $W = \dfrac{\partial^2\eta(\beta)}{\partial\beta\partial\beta^{\mathrm{T}}}$, $[\cdot][\cdot]$ 表示矩阵与立体阵的方括号乘积.

由立体阵的乘法可知, 这个定义中的 A^I 和 A^P 分别是 $n\times p\times p$, $p\times p\times p$ 阶立体阵. 通过参数变换 $\gamma = R(\beta-\beta_0)$, 我们能清楚理解这个定义的几何意义.

令 $\beta = \beta(\gamma) = L\gamma + \beta_0$, 故 $\eta(\beta(\gamma))$ 关于 γ 的前二阶导数分别为

$$\frac{\partial\eta}{\partial\gamma^{\mathrm{T}}} = \frac{\partial\eta}{\partial\beta^{\mathrm{T}}}\frac{\partial\beta}{\partial\gamma^{\mathrm{T}}} = DL = Q,$$

$$\frac{\partial^2\eta}{\partial\gamma\partial\gamma^{\mathrm{T}}} = \left(\frac{\partial\beta}{\partial\gamma^{\mathrm{T}}}\right)^{\mathrm{T}}\left(\frac{\partial^2\eta}{\partial\beta\partial\beta^{\mathrm{T}}}\right)\left(\frac{\partial\beta}{\partial\gamma^{\mathrm{T}}}\right) = L^{\mathrm{T}}WL = U.$$

上式表示在 (9) 式中的立体阵 U 是 $\eta(\beta(\gamma))$ 关于 γ 的二阶导数, 而 $\eta(\beta(\gamma))$ 的一阶导数正是 Q 且构成 π 在 β 的切空间的标准正交基. 因而 (9) 式定义的 A^I 和 A^P 分别是二阶导数 U 在法空间 T'_β 和切空间 T_β 的投影.

令 $A^I = (A^I_{kij})$, $A^P = (A^P_{aij})$, $U = (U_{ij})$, $N = (N_k)$ 及 $Q = (Q_a)$, $k = 1, \cdots, n-p$; $a, i, j = 1, 2, \cdots, p$, 则 U_{ij} 表示立体阵 U 在 (i, j) 处的 $n\times 1$ 维向量, 及 N_k 和 Q_a 分别是法向量和切向量的第 k 个和第 a 个正交基向量. 因而 $A^I_{kij} = N_k^{\mathrm{T}}U_{ij}$ 和 $A^P_{aij} = Q_a^{\mathrm{T}}U_{ij}$ 是 U_{ij} 在 N_k 和 Q_a 的投影. 这个几何意义与通常的微分几何的曲率定义一致. 因而曲率立体阵 A^I 和 A^P 确定了 π 在 β 处的几何特征, 并且它们为模型的非线性强度提供了合适的度量.

正如 Bates and Watts[2] 指出的那样, 立体阵 A^I 和 A^P 是刻画模型非线性强度的主要指标, A^I 是模型固有的数量特征, 与参数的选择无关, 而 A^P 则依赖于参数. 模型的统计性质与两个曲率立体阵有紧密关系, 例如参数的置信域就可以由这两个立体阵表出. 曲率立体阵 A^I 和 A^P 满足

$$[N][A^I] + [Q][A^P] = U. \tag{10}$$

以后的讨论将以上面定义的几何结构为基础.

3 参数置信域的曲率表示

3.1 参数的似然置信域的曲率表示

对于模型 (1), 给定水平 α, 则参数 β 的水平为 $1-\alpha$ 的置信域就是参数空间 B 中的一个区域, 它满足

$$P_B\{\beta: \beta\in C(Y)\} \geqslant 1-\alpha. \tag{11}$$

关于参数的置信域, 常用基于似然比检验统计量的似然置信域, 对模型 (1) 参数的似然置信域可表示为

$$C(Y) = \{\beta: LR(\beta) \leqslant \rho^2(\alpha)\}, \tag{12}$$

其中 $LR(\beta) = 2\{l(\hat{\beta}) - l(\beta)\}$. $LR(\beta)$ 有渐近 χ^2 分布, 所以 $\rho^2(\alpha) = \chi^2(p, \alpha)$.

对模型 (1), 由于 $f(\beta)$ 是 β 的非线性函数, 置信域 $C(Y)$ 在参数空间中的形状是相当复杂的. 为了解置信域的性质, 正如 Hamilton 等[3] 指出的那样, 获得一个直观和容易理解的置信域的近似表达是相当有价值的. 为此常常需要进行参数变换. 例如, 令 $t = t(\beta)$, t 属于某一个空间 T, 使得参数空间 B 中的区域 $C(Y)$ 对应于 T 中的某一个区域 $K(Y)$, 并且有

$$P_\beta\{\beta: \beta\in C(Y)\} = P_\beta\{\beta: t(\beta)\in K(Y)\}, \tag{13}$$

而 $K(Y)$ 在 T 中的形状相对比较简单, 这样就可以通过 T 中的置信域 $K(Y)$ 以及变换关系式 $t = t(\beta)$ 来了解原参数空间 B 中的置信域 $C(Y)$ 的性质. 对正态非线性回归模型和指数族非线性模型, Hamilton 等[3] 和 Wei[7] 已经成功地给出了一个从参数空间 B 到切空间的 1-1 变换, 由此获得了置信域在切空间的投影, 该投影为由模型的曲率表示的椭球. 对正态非线性模型可使用类似的方法导出置信域. 由于切空间是一个线性子空间, 而且又是 π 的线性近似, 因而置信域在其中的投影往往比在原参数空间更加简单清楚,

我们可以作一个参数空间到切空间的变换. 在参数空间中, 点 β 和 $\hat{\beta}$ 分别映射到 π 上的 $\eta(\beta)$ 和 $\eta(\hat{\beta})$. $\eta(\beta)-\eta(\hat{\beta})$ 在 π 上的切空间的投影可记为 $\boldsymbol{t}(\beta)=\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}(\eta(\beta)-\eta(\hat{\beta}))$, 其中 $\boldsymbol{Q}$ 在 $\hat{\beta}$ 处记值(在这一节所有的量, 例如 $\boldsymbol{Q}, \boldsymbol{D}, \boldsymbol{R}$ 等都在 $\hat{\beta}$ 处记值). 如果取 $\boldsymbol{Q}$ 的列向量作为在 π 处切空间的正交基, 那么 $\boldsymbol{t}(\beta)$ 在切空间投影的坐标为

$$\xi(\beta)=\boldsymbol{Q}^{\mathrm{T}}(\eta(\beta)-\eta(\hat{\beta})). \tag{14}$$

作为一个新参数, $\xi=\xi(\beta)$ 有良好的性质, $\xi=\xi(\beta)$ 表示一个从参数空间到切空间的非线性映射, 它把 π 和切空间联系在一起, 坐标 ξ 为 π 及其渐近线提供了一个自然的参考系统. 使用二阶逼近, 我们能构造参数 β 由坐标 $\xi=\xi(\beta)$ 表示的置信域. 类似于(14)式的变换曾经被 Hamulton 等[3] 使用. 在 $\hat{\beta}$ 的某个邻域内, 变换式(14)是 1–1 映射. 当 $\beta=\hat{\beta}$ 时, $\xi=\boldsymbol{0}$ 我们用 $\beta=\beta(\xi)$ 表示 $\xi=\xi(\beta)$ 的逆映射.

为简单起见, 当考虑 $\beta=\beta(\xi)$ 时, 我们把对数似然 $l(\beta)$ 和似然比统计量 $\boldsymbol{LR}(\beta)$ 记作 $l(\xi)$ 及 $\boldsymbol{LR}(\xi)$. 类似地, 当 $\beta=\hat{\beta}$ 即 $\xi=\boldsymbol{0}$, 我们把 $l(\hat{\beta}), \dot{l}(\hat{\beta}), \ddot{l}(\hat{\beta})$, 记作 $l(\boldsymbol{0}), \dot{l}(\boldsymbol{0}), \ddot{l}(\boldsymbol{0})$, 利用参数 $\xi=\xi(\beta)$ 表示 β, 我们可获得 $\boldsymbol{LR}(\xi)$ 的一个二阶逼近. 为此, 需要如下引理.

引理 1 对模型(1), 且满足[7]中的正则条件, 则(14)式中的 $\xi(\beta)$ 及 $\beta(\xi)$ 在 $\hat{\beta}$ 处的导数可表示为:

$$\frac{\partial\xi}{\partial\beta^{\mathrm{T}}}=\boldsymbol{R};\quad \frac{\partial^2\xi}{\partial\beta\partial\beta^{\mathrm{T}}}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{A}^{P}\boldsymbol{R}, \tag{15}$$

$$\frac{\partial\beta}{\partial\xi^{\mathrm{T}}}=\boldsymbol{L};\quad \frac{\partial^2\beta}{\partial\xi\partial\xi^{\mathrm{T}}}=-[\boldsymbol{L}^{\mathrm{T}}][\boldsymbol{A}^{P}]. \tag{16}$$

引理证明参看[6], 由此可得如下定理.

定理 1 对模型(1), 且满足[7]中的正则条件, 则参数 β 的 $100(1-\alpha)\%$ 似然置信域在 $\hat{\beta}$ 处切空间映射的二阶近似可表示为

$$\boldsymbol{K}(\boldsymbol{Y})=\{\beta: \xi^{\mathrm{T}}(\beta)(\boldsymbol{I}_P-\boldsymbol{B})\xi(\beta)\leqslant\sigma^2\chi^2(p,\alpha)=\sigma^2\rho^2(\alpha)\}, \tag{17}$$

且 $\boldsymbol{I}_P-\boldsymbol{B}>\boldsymbol{0}$, 其中 $\boldsymbol{B}=[\boldsymbol{e}^{\mathrm{T}}(\beta)\boldsymbol{N}][\boldsymbol{A}^{I}]$, 且在 $\hat{\beta}$ 处记值, $\chi^2(p,\alpha)$ 是 χ^2 分布的分位数.

证明　使用变换(14)式, (12)式中的似然比统计量可表示为

$$\boldsymbol{LR}(\beta)=\boldsymbol{LR}(\xi)=-2\{l(\xi)-l(\boldsymbol{0})\}. \tag{18}$$

把(18)式在 $\xi=\boldsymbol{0}$ 处进行二阶 Taylor 展开可得

$$\boldsymbol{LR}(\beta)\approx-\xi^{\mathrm{T}}(\beta)\ddot{l}(\boldsymbol{0})\xi(\beta)-2(\dot{l}(\boldsymbol{0}))^{\mathrm{T}}\xi(\beta), \tag{19}$$

其中 $\dot{l}(\boldsymbol{0})=\dfrac{\partial l(\xi)}{\partial\xi}$, $\ddot{l}(\boldsymbol{0})=\dfrac{\partial^2 l(\xi)}{\partial\xi\partial\xi^{\mathrm{T}}}$ 在 $\xi=\boldsymbol{0}$(即 $\beta=\hat{\beta}$)处记值. 由于

$$\frac{\partial l(\xi)}{\partial\xi}=\left(\frac{\partial\beta}{\partial\xi^{\mathrm{T}}}\right)^{\mathrm{T}}\frac{\partial l(\beta)}{\partial\beta},$$

$$\frac{\partial^2 l(\xi)}{\partial\xi\partial\xi^{\mathrm{T}}}=\left(\frac{\partial\beta}{\partial\xi^{\mathrm{T}}}\right)^{\mathrm{T}}\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^{\mathrm{T}}}\right)\left(\frac{\partial\beta}{\partial\xi^{\mathrm{T}}}\right)+\left[\left(\frac{\partial l(\beta)}{\partial\beta}\right)^{\mathrm{T}}\right]\left[\frac{\partial^2\beta}{\partial\xi\partial\xi^{\mathrm{T}}}\right],$$

由(1)可得

$$\dot{l}(\boldsymbol{0})=\boldsymbol{L}^{\mathrm{T}}\dot{l}(\hat{\beta})=\boldsymbol{0},\quad \ddot{l}(\boldsymbol{0})=\boldsymbol{L}^{\mathrm{T}}\ddot{l}(\hat{\beta})\boldsymbol{L},$$

因此由(19), 我们有

$$\boldsymbol{LR}(\beta)\approx\xi^{\mathrm{T}}(\beta)\{-\ddot{l}(\boldsymbol{0})\}\xi(\beta)=\xi^{\mathrm{T}}(\beta)\boldsymbol{L}^{\mathrm{T}}\{-\ddot{l}(\hat{\beta})\}\boldsymbol{L}\xi(\beta). \tag{20}$$

由 $\boldsymbol{D}=\boldsymbol{Q}\boldsymbol{R}$, 则 $\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$. 由此可得

$$-\ddot{l}(\hat{\beta})=\frac{1}{\sigma^2}\{\boldsymbol{D}^{\mathrm{T}}(\hat{\beta})\boldsymbol{D}(\hat{\beta})-[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})][\boldsymbol{W}(\hat{\beta})]\}=\frac{1}{\sigma^2}\{\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_P-[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})][\boldsymbol{U}(\hat{\beta})])\boldsymbol{R}\}. \tag{21}$$

由于 $\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})\boldsymbol{D}(\hat{\beta})=\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})\boldsymbol{Q}\boldsymbol{R}=\boldsymbol{0}$, 则 $\boldsymbol{e}^{\mathrm{T}}(\beta)\boldsymbol{Q}=\boldsymbol{0}$, 由此可得

$$[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})][\boldsymbol{U}(\hat{\beta})]=[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})][(\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}+\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}})\boldsymbol{U}]=[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}][\boldsymbol{U}]=[\boldsymbol{e}^{\mathrm{T}}(\hat{\beta})\boldsymbol{N}][\boldsymbol{A}^{I}]=\boldsymbol{B},$$

因此有

$$-\ddot{l}(\hat{\beta})=\frac{1}{\sigma^2}\{\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_P-\boldsymbol{B})\boldsymbol{R}\}, \tag{22}$$

把(22)式带入(20)式即可得到(19)式. 由于 $\hat{\beta}$ 是极大似然估计, 因此 $-\ddot{l}(\hat{\beta})>\boldsymbol{0}$, 由此可得 $\boldsymbol{I}_P-\boldsymbol{B}>\boldsymbol{0}$

(17)式表示, 似然置信域在切空间投影的二阶近似是一个在坐标 ξ 下的椭球, 并且这个椭球只与固有曲率有关, 与参数效应曲率无关. 矩阵 $\boldsymbol{B}$ 称为有效残差曲率阵.

3.2 子集参数的置信域

给定模型(1), 今考虑模型的某个子集参数的置信域, Hamilton[4] 最早研究了子集参数的置信域, 不失一般性, 可设 $\beta^{\mathrm{T}}=(\beta_1^{\mathrm{T}},\beta_2^{\mathrm{T}})^{\mathrm{T}}$, 其中 β_1 为多余参数, β_2 为有兴趣参数, 维数分别为 k 和 q. 由于多余参数 β_1 的存在, 寻求参数 β_2 的置信域要比一般情况困难. 分两步: 第一、找出参数 β_2 的某些置信域; 第二、映射到切空间并进行二阶近似. 为此, 我们首先把一些量按 β 的维数进行划分:

$\xi^{\mathrm{T}}(\beta)=(\xi_1^{\mathrm{T}}(\beta),\xi_2^{\mathrm{T}}(\beta))^{\mathrm{T}}$, $\boldsymbol{D}(\beta)=(\boldsymbol{D}_1(\beta),\boldsymbol{D}_2(\beta))$, $\boldsymbol{R}=(\boldsymbol{R}_{ij})$, $\boldsymbol{B}=(\boldsymbol{B}_{ij})$, $i, j=1, 2$

下面我们也会用到其他一些量的类似分块划分.

与(12)式类似, β_2 的似然置信域可表示成

$$\boldsymbol{C}_S(\boldsymbol{Y})=\{\beta_2: \boldsymbol{LR}_S(\beta_2)\leqslant\rho_S^2(\alpha)\}, \tag{23}$$

其中

$$\boldsymbol{LR}_S=2\{\boldsymbol{l}(\hat{\beta})-\boldsymbol{l}(\tilde{\beta})\}, \tag{24}$$

这里 $\tilde{\beta}=(\tilde{\beta}_1^{\mathrm{T}}(\beta_2),\beta_2^{\mathrm{T}})^{\mathrm{T}}$; $\tilde{\beta}(\beta_2)$ 是固定 β_2 时, β_1 关于 $\boldsymbol{l}(\beta_1,\beta_2)$ 的极大似然估计; $\rho_S^2(\alpha)=\chi^2(q,\alpha)$ 是置信水平.

为获得子集参数 β_2 的近似似然置信域, 我们仍使用变换(14)式, 在这种情况下, (14)式可表示成:

$$\tilde{\xi}=\xi(\tilde{\beta})=\boldsymbol{Q}^{\mathrm{T}}\{\eta(\tilde{\beta})-\eta(\hat{\beta})\}, \tag{25}$$

其中 $\tilde{\xi}=\xi(\tilde{\beta})$ 是 β_2 的函数, 利用(24)和(25)两式, 我们有如下定理.

定理 2 条件同定理 1, 则子集参数 β_2 的似然置信域在切空间的二阶近似可表示为

$$\boldsymbol{K}_S(\boldsymbol{Y})=\{\beta_2: \tilde{\xi}_2^{\mathrm{T}}(\beta_2)(\boldsymbol{I}_q-\boldsymbol{T})\tilde{\xi}(\beta_2)\leqslant\sigma^2\rho_S^2(\alpha)\}, \tag{26}$$

其中 $\boldsymbol{T}=\boldsymbol{B}_{22}+\boldsymbol{B}_{21}(\boldsymbol{I}_k-\boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}$, $k=p-q$, $\tilde{\xi}^{\mathrm{T}}=(\tilde{\xi}_1^{\mathrm{T}},\tilde{\xi}_2^{\mathrm{T}})^{\mathrm{T}}$.

证明 由(20)、(22)及(25)式可知, 以下表达式在 $\beta=\tilde{\beta}$ 处仍然成立

$$\boldsymbol{LR}_S(\beta_2)\approx\frac{1}{\sigma^2}\{\xi^{\mathrm{T}}(\tilde{\beta})(\boldsymbol{I}_P-\boldsymbol{B})\xi(\tilde{\beta})\}. \tag{27}$$

为了化简该等式, 我们先求 $\tilde{\xi}_1$ 和 $\tilde{\xi}_2$ 的近似关系式. 事实上, 由(14)式和(25)式可得

$$\xi=\boldsymbol{Q}^{\mathrm{T}}(\eta(\beta)-\eta(\hat{\beta}))\approx\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}(\beta-\hat{\beta})=\boldsymbol{R}(\beta-\hat{\beta}),$$
$$\tilde{\xi}=\boldsymbol{Q}^{\mathrm{T}}(\eta(\tilde{\beta})-\eta(\hat{\beta}))\approx\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}(\tilde{\beta}-\hat{\beta})=\boldsymbol{R}(\tilde{\beta}-\hat{\beta}).$$

由于 $\boldsymbol{R}$ 是上三角矩阵, 把上述两等式分块

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}\approx\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{pmatrix}\begin{pmatrix}\beta_1-\hat{\beta}_1\\ \beta_2-\hat{\beta}_2\end{pmatrix},$$

$$\begin{pmatrix}\tilde{\xi}_1\\ \tilde{\xi}_2\end{pmatrix}\approx\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{pmatrix}\begin{pmatrix}\tilde{\beta}_1-\hat{\beta}_1\\ \tilde{\beta}_2-\hat{\beta}_2\end{pmatrix},$$

由此可得

$$\tilde{\xi}_2\approx\xi_2\approx\boldsymbol{R}_{22}(\beta_2-\hat{\beta}_2).$$

另一方面, 把 $\dot{\boldsymbol{l}}(\tilde{\beta})$ 在 $\hat{\beta}$ 处进行一阶 Tarylor展开并利用(22)式可得

$$\dot{\boldsymbol{l}}(\tilde{\beta})\approx\ddot{\boldsymbol{l}}(\hat{\beta})(\tilde{\beta}-\hat{\beta})=-\frac{1}{\sigma^2}\{\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_P-\boldsymbol{B})\boldsymbol{R}\}(\tilde{\beta}-\hat{\beta})\approx-\frac{1}{\sigma^2}\{\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_P-\boldsymbol{B})\}\tilde{\xi}$$

因此

$$-\boldsymbol{L}^{\mathrm{T}}\dot{\boldsymbol{l}}(\tilde{\beta})\approx\frac{1}{\sigma^2}(\boldsymbol{I}_P-\boldsymbol{B})\tilde{\xi}. \tag{28}$$

由于 $\tilde{\beta}$ 是极大子集参数似然估计, 即对任意 β_2 有

$$\left.\frac{\partial\boldsymbol{l}(\beta)}{\partial\beta}\right|_{\tilde{\beta}}=\begin{pmatrix}\dfrac{\partial\boldsymbol{l}(\beta)}{\partial\beta_1}\\[2mm] \dfrac{\partial\boldsymbol{l}(\beta)}{\partial\beta_2}\end{pmatrix}_{\tilde{\beta}}=\begin{pmatrix}\boldsymbol{0}\\[2mm] \dfrac{\partial\boldsymbol{l}(\beta)}{\partial\beta_2}\end{pmatrix}_{\tilde{\beta}},$$

把上式代入(28)式,可得

$$\begin{pmatrix} \boldsymbol{L}_{11}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{L}_{12}^{\mathrm{T}} & \boldsymbol{L}_{22}^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{0} \\ \dfrac{\partial \boldsymbol{l}(\beta)}{\partial \beta_2} \end{pmatrix} \approx -\frac{1}{\sigma^2}\left\{ \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_q \end{pmatrix} - \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix} \right\} \begin{pmatrix} \xi_1 \\ \tilde{\xi}_2 \end{pmatrix},$$

由此可得$(\boldsymbol{I}_k - \boldsymbol{B}_{11})\tilde{\xi}_1 - \boldsymbol{B}_{12}\tilde{\xi}_2 \approx \boldsymbol{0}$,因而有

$$\tilde{\xi}_1 = (\boldsymbol{I}_k - \boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}\tilde{\xi}_2, \quad \tilde{\xi}_2 \approx \xi_2, \tag{29}$$

把上式代入(27)式,可得

$$\boldsymbol{LR}_S(\beta) \approx \frac{1}{\sigma^2}\begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{I}_k - \boldsymbol{B}_{11} & -\boldsymbol{B}_{12} \\ -\boldsymbol{B}_{21} & \boldsymbol{I}_q - \boldsymbol{B}_{22} \end{pmatrix} \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} = \frac{1}{\sigma^2}\xi_2^{\mathrm{T}}(\beta_2)(\boldsymbol{I}_q - \boldsymbol{T})\xi_2(\beta_2).$$

由此可知(26)式成立.

定理2的表达式和几何意义是Hamilton[4]关于非线性回归模型和Wei[7]关于指数族非线性模型中的相应结果在指数非线性随机效应模型中的进一步推广.

根据韦博成[8]的结果,子集参数β_2的置信域可表示为

$$\boldsymbol{SC} = \left\{\left(\frac{\partial \boldsymbol{l}}{\partial \beta_2}\right)^{\mathrm{T}} \boldsymbol{J}^{22}\left(\frac{\partial \boldsymbol{l}}{\partial \beta_2}\right)\right\}_{\beta = \tilde{\beta}}, \tag{30}$$

其中$\boldsymbol{J}^{22}$是矩阵$\boldsymbol{J}$的逆矩阵$\boldsymbol{J}^{-1} = (\boldsymbol{J}^{ij})$的左下角分块阵,$i, j = 1, 2$, $\boldsymbol{J} = \dfrac{1}{\sigma^2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}$,对每一个$\beta_2$, $\boldsymbol{SC}$有渐近$\chi^2(q)$分布.为获得$\boldsymbol{SC}$的二次逼近,我们首先证明几个有用的引理.

引理2　条件同定理1,令

$$\boldsymbol{P} = \boldsymbol{D}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})^{-1}\boldsymbol{D}^{\mathrm{T}}, \quad \boldsymbol{P}_1 = \boldsymbol{D}_1(\boldsymbol{D}_1^{\mathrm{T}}\boldsymbol{D}_1)^{-1}\boldsymbol{D}_1^{\mathrm{T}},$$

则对Bayes条件下非线性模型(1), (30)式的$\boldsymbol{SC}$能表示成

$$\boldsymbol{SC} = \frac{1}{\sigma^2}\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{e}(\tilde{\beta}), \tag{31}$$

其中, $\boldsymbol{P}, \boldsymbol{P}_1$在$\tilde{\beta}$处记值.

证明参看[6],由此可得如下定理.

定理3　条件同定理2,则参数β_2的似然置信域在切空间的二阶近似为

$$K_{SC}(\boldsymbol{Y}) = \{\beta_2: \xi_2^{\mathrm{T}}(\beta_2)(\boldsymbol{I}_q - \boldsymbol{T})^2\xi_2(\beta_2) \leqslant \sigma^2\chi_q^2(\alpha)\}, \tag{32}$$

其中$\boldsymbol{T}$与ξ_2的意义与定理2相同.

证明　利用引理1和2,我们首先计算$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{P}\boldsymbol{e}(\tilde{\beta})$,由上述分析,容易得到

$$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{P}\boldsymbol{e}(\tilde{\beta}) = \boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{D}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})^{-1}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}(\tilde{\beta}), \tag{33}$$

其中$\boldsymbol{D}$在ξ(即$\beta = \tilde{\beta}$)处记值,为计算$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{D}$和$(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})$, (33)经化简可得

$$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{P}\boldsymbol{e}(\tilde{\beta}) = (\boldsymbol{L}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}(\tilde{\beta}))^{\mathrm{T}}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}(\tilde{\beta}).$$

由(28)式得

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{e}(\tilde{\beta}) = -(\boldsymbol{I}_p - \boldsymbol{B})\tilde{\xi}$$

$$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{P}\boldsymbol{e}(\tilde{\beta}) \approx \tilde{\xi}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{B})\tilde{\xi}$$

令$\boldsymbol{E}_1 = \mathrm{diag}(\boldsymbol{I}_k, 0)$是一个$p \times p$矩阵,则$\boldsymbol{D}_1 = (\boldsymbol{D}_1, \boldsymbol{D}_2)\boldsymbol{E}_1$,因而容易得到

$$\boldsymbol{e}^{\mathrm{T}}(\tilde{\beta})\boldsymbol{P}_1\boldsymbol{e}(\tilde{\beta}) \approx \tilde{\xi}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{B})\boldsymbol{E}_1(\boldsymbol{I}_p - \boldsymbol{B})\tilde{\xi}$$

因此由引理2,可得

$$\boldsymbol{SC} \approx \frac{1}{\sigma^2}\tilde{\xi}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{B})(\boldsymbol{I}_p - \boldsymbol{E}_1)(\boldsymbol{I}_p - \boldsymbol{B})\tilde{\xi}$$

把(29)式代入上式即可得到(32)式.

以上得到的结果以及使用的方法与Hamilton等[3]关于正态非线性回归模型的相应工作有一定的相似之处,但讨论的模型不同.

[参考文献]

[1] Efron B. The geometry of exponential families[J]. Ann Staist, 1975, 3: 1189-1242.
[2] Bates D M, Watts D G. Relative curvature measures of nonlinearity[J]. J Roy Statist Soc Ser B, 1980, 42: 1-25.
[3] Hamilton D C, Watts D G, Bates D M. Accounting for intrinstic nonlinearity in nonlinear regression parameter inference regions[J]. Ann Statist, 1982, 10(2): 386-393.
[4] Hamilton D C. Confidence regions for parameter subset in nonlinear regression[J]. Biometrika, 1986, 73: 57-64.
[5] 唐年胜, 韦博成, 王学仁. 非线性再生散度模型参数置信域的曲率表示[J]. 高校应用数学学报, 1999, 14(3): 293-300.
[6] Zhong Xuping, Meng Guoming, Wei Bocheng. On confidence region of nonlinear models with random effects[J]. Mathematica Applicata, 2000, 13(4): 100-105.
[7] Wei B C. Exponential Family Nonlinear Models[M]. Singapore: Springer-Verlag, 1998.
[8] 韦博成. 参数统计教程[M]. 北京: 高等教育出版社, 2006: 276.

[责任编辑: 丁　蓉]