Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

DIFFERENTIAL GEOMETRIC METHOD FOR CURVED EXPONENTIAL FAMILIES*

BY BO CHENG WEI AND PETER CHE BOR LAM
Southeast University and Hong Kong Baptist College

Efron (1975) first proposed a differential geometric method to study curved exponential families (CEF) and Amari (1982, 1985) presented a Riemannian geometric framework. In this paper we modify BW's (Bates and Watts, 1980) framework for nonlinear regression models (NRM) by introducing a Fisher information inner product so that it can be applied to CEF. Based on this modified BW framework (MBW), we study the parameter effect and confidence regions for CEF.

1. Modified BW Geometric Framework. Suppose that observations $\chi_1, \cdots, \chi_n$ are independently and identically distributed and χ_1 has the density

$$p(\chi_1; \vartheta) = \exp\{\chi_1^T \vartheta - \psi(\vartheta)\}, \tag{1.1}$$

where $\vartheta = (\vartheta_1, \cdots, \vartheta_m)^T$ is the natural parameter belonging to a convex set $\Theta \subseteq I\!R^m$. We assume that (1.1) is a full, regular, and minimally represented exponential family. It is well known that

$$E(\chi_1) \equiv \mu(\vartheta) = \psi'(\vartheta), \qquad \mathrm{Var}(\chi_1) \equiv g(\vartheta) = \psi''(\vartheta), \tag{1.2}$$

where $\psi'(\vartheta)$ and $\psi''(\vartheta)$ are the first two derivatives of $\psi(\vartheta)$. Let $\overline{\chi} = n^{-1}\Sigma_i \chi_i$, the log likelihood of $\boldsymbol{\chi} = (\chi_1, \cdots, \chi_m)^T$ is

$$l(\vartheta; \chi) = n\big[(\overline{\chi}^T \vartheta - \psi(\vartheta)\big]. \tag{1.3}$$

The Fisher information of $\boldsymbol{\chi}$ for ϑ and μ are $g(\vartheta)$ and $g^{-1}(\vartheta)$ respectively. Let ϑ in (1.1) be defined for $\beta \in \boldsymbol{B} \subseteq I\!R^p$, so (1.1) becomes a CEF. Suppose the first three derivatives of $\vartheta(\beta)$ are finite in $\boldsymbol{B}$. Let $V_\vartheta = \partial\vartheta/\partial\beta^T, W_\vartheta = \partial^2\vartheta/\partial\beta\partial\beta^T$, $V = \partial\mu/\partial\beta^T$ and $W = \partial^2\mu/\partial\beta\partial\beta^T$, then the score function

* Partially supported by Hong Kong UPGC Grant.

AMS 1980 Subject Classifications: 62J02, 62F25.

Key words and phrases: Curved exponential Families, fisher information, statistical curvature, confidence region, score statistic.

$l'(\beta)$, observed information $-l''(\beta)$ and Fisher information $J(\chi)$ with respect to β are respectively:

$$l'(\beta) = \sqrt{n}\, V^T(\beta)\, g^{-1}(\beta)\, e(\beta), \qquad e(\beta) = \sqrt{n}\,\{\overline{\chi} - \mu(\vartheta(\beta))\} \tag{1.4}$$

$$-l''(\beta) = nV^T g^{-1} V - \sqrt{n}[e^T g^{-1}][W - \Gamma], \tag{1.5}$$

$$J(\chi) = nV^T(\beta)\, g^{-1}(\beta) V(\beta), \tag{1.6}$$

where $\Gamma = V^T g^{-1} S g^{-1} V$ and $S = \psi^{(3)}(\vartheta)$ are $m \times p \times p$ arrays. The maximum likelihood estimate $\widehat{\beta}$ of β satisfies

$$V^T(\widehat{\beta}) g^{-1}(\widehat{\beta}) e(\widehat{\beta}) = 0, \tag{1.7}$$

i.e. the residual vector $\widehat{e} = e(\widehat{\beta})$ is orthogonal to all column vectors of $V(\widehat{\beta})$ under the Fisher information inner product induced by g^{-1} on $I\!R^m$. Let the QR decomposition of $V(\beta)$ be $V = (Q, N)(R^T, 0)^T = QR$, where R and $L = R^{-1}$ are $p \times p$ nonsingular upper triangular matrices and columns of Q and N are the orthonormal basis for the tangent space T_β and the normal space T'_β of solution locus π at β. The intrinsic curvature array A^I and parameter-effects curvature array A^p are defined as follows:

$$A^I = [N^T g^{-1}][U], \qquad A^p = [Q^T g^{-1}][U] \qquad \text{and} \qquad U = L^T W L, \tag{1.8}$$

where $[\cdot][\cdot]$ indicates array multiplication as in BW (1980). We can also introduce a dual geometry based on the dual parameter ϑ of μ (Efron (1975), Amari (1985) and Kass (1983). We have a very important relation between A^I and the dual curvature arrays A^I_ϑ as follows:

$$A^I_\vartheta = A^I - \Gamma^I \qquad \text{and} \qquad \Gamma^I = [N^T g^{-1}][L^T \Gamma L], \tag{1.9}$$

where Γ is given by (1.5).

2. Curvature Representation for Parameter Confidence Regions. Hamilton et al. (1982) studied improved approximate confidence regions for normal NRM using BW geometric framework. Now we study similar problems for CEF using MBW geometric framework. A usual approximate confidence region is based on the likelihood ratio statistic:

$$LR(\beta) = -2[l(\beta) - l(\widehat{\beta})] \xrightarrow{L} x^2(p).$$

To derive improved approximate projections of solution locus inference regions onto the tangent space, we introduce a nonlinear transformation as follows:

$$u = u(\beta) = \sqrt{n} Q^T g^{-1}\{\mu(\beta) - \mu(\widehat{\beta})\}. \tag{2.1}$$

THEOREM 1. *Under the preceding assumptions, the approximate tangent space projection of the solution locus likelihood region of β with $1-\alpha$ level may be expressed as:*

$$u^T(\beta)\{I_p - B_\vartheta/\sqrt{n}\}u(\beta) \leq \chi^2(p,\alpha),$$
$$B_\vartheta = B - B^\Gamma = \left[\hat{e}^T g^{-1} N\right]\left[A_\vartheta\right],$$

where $B = \left[\hat{e}^T g^{-1} N\right]\left[A^I\right]$, $B^\Gamma = \left[\hat{e}^T g^{-1} N\right]\left[\Gamma^I\right], \chi^2(p,\alpha)$ is the upper α percentage point of chi-square distribution with p degree of freedom, and I_p a $p \times p$ identity matrix. ▮

This result is very similar to that of Hamilton et al. (1982), but our CEF is completely different from normal NRM.

If a subset of parameters is of primary interest as studied by Hamilton (1986), the parameter vector β can often be partitioned as $\beta^T = (\beta_1^T, \beta_2^T)$, where β_1 is nuisance parameter in $\mathbb{R}^k$ and β_2 is parameter of interest in $\mathbb{R}^{p-k}$. The likelihood ratio statistic associated with β_2 is:

$$LR_s(\tilde{\beta}) = -2\{l(\tilde{\beta}_1(\beta_2), \beta_2) - l(\hat{\beta})\},$$

where $\tilde{\beta} = (\tilde{\beta}_1^T(\beta_2), \tilde{\beta}_2^T)^T$ and $\tilde{\beta}_1(\beta_2)$ is MLE of β_1 with β_2 fixed. The transformation (2.1) has the form

$$\tilde{u} = u(\tilde{\beta}) = \sqrt{n}\, Q^T g^{-1}\{u(\tilde{\beta}) - u(\hat{\beta})\}.$$

We have the following theorems.

THEOREM 2. *The improved $1-\alpha$ confidence region for β_2 may be expressed as:*

$$\tilde{u}_2^T(\beta_2)(I_q - T/\sqrt{n})\tilde{u}_2(\beta) \leq \chi^2(q,\alpha),$$
$$T = (B_\vartheta)_{22} + (B_\vartheta)_{21}(\sqrt{n}I_k - (B_\vartheta)_{11})^{-1}(B_\vartheta)_{12},$$

$(B_\vartheta)_{ij}$ is the partitioned matrix of $B_\vartheta = B - B^\Gamma$. ▮

The score statistic can be used to construct confidence region for parameter subsets as studied by Hamilton (1986) for normal NRM. For our CEF, the score statistic associated with the parameter subset β_2 is

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

where J^{22} is the lower right corner of the partition of $J^{-1}(\chi) = (J^{ij}), \quad i,j = 1,2$, and $J(\chi)$ is given in (1.6). Then we have:

LEMMA 1. *SC can be written as* $SC = \tilde{e}^T g^{-1}(P - P_1)\tilde{e}$, *where* $P = V(V^T g^{-1} V)^{-1} V^T g^{-1}$, $P = V_1(V_1^T g^{-1} V_1)^{-1} V_1^T g^{-1}$, g^{-1}, P, P_1 *and* $\tilde{e}$ *are all evaluated at* $\tilde{\beta}$.

THEOREM 3. *The* $1-\alpha$ *improved approximate tangent space projection of solution locus inference region for parameter subset* β_2 *based on score statistic may be represented as*

$$\tilde{u}_2^T (I_q - Z/\sqrt{n})^T (I_q - Z/\sqrt{n}) \tilde{u}_2 \leq \chi^2(q, \alpha)$$

where $Z = B_{22} + B_{21}\left[\sqrt{n}\, I_k - (B_\vartheta)_{11}\right]^{-1}(B_\vartheta)_{12}$. ∎

These results are similar to Hamilton et al (1982) and Hamilton (1986) with different CEF and geometric framework. Kumon and Amari (1983) also studied some problems related to confidence regions for CEF, but the methods are different from ours.

REFERENCES

AMARI, S. (1982). Differential geometry of curved exponential family - curvatures and information loss, *Ann. Statist.* **10**, 357–385.

AMARI, S. (1982). Differential geometrical methods in Statistics, *Lecture Notes in Statistics* **28**, Springer, Berlin.

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity, *J.R. Statist. Soc.* **B42**, 1–25.

EFRON, B. (1975). Defining the curvature of a statistical problem (with application to second order efficiency), *Ann. Statist.* **3**, 1189–1242.

EFRON, B. and HINKLEY, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika* **65**, 457–487.

HAMILTON, D.C. (1986). Confidence regions for parameter subsets in nonlinear regression, *Biometrika* **73**, 57–64.

HAMILTON, D.C., WATTS, D.G. and BATES, D.C. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions, *Ann. Statist.* **10**, 386–393.

KASS, R.E. (1984). Canonical parameterization and zero parameter effects curvature, *J.R. Statist. Soc.* **B46**, 86–92.

KUMON, M. and AMARI, S. (1983). Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference, *Proc. Royal Society*, London **A387**, 429–458.

WEI, B.C. (1991). Some second order asymptotics in nonlinear regression, *Aust. J. Statist.* **33**, 75–84.

DEPARTMENT OF MATHEMATICS
SOUTHEAST UNIVERSITY
210018 NANJING, JIANGSU
PEOPLE'S REPUBLIC OF CHINA

DEPARTMENT OF MATHEMATICS
HONG KONG BAPTIST COLLEGE
KOWLOON, HONG KONG