

## Statistical Analysis of Curved Probability Densities

MASANOBU TANIGUCHI

*Osaka University, Toyonaka 560, Japan*

AND

YOSHIHIDE WATANABE

*Hiroshima University, Higashi-Hiroshima 724, Japan*

Suppose that  $p_n(\cdot; \theta)$  is the joint probability density of  $n$  observations which are not necessarily i.i.d. In this paper we discuss the estimation of an unknown parameter  $u$  of a family of "curved probability densities" defined by  $\mathbf{M} = \{p_n(\cdot; \theta(u)), \dim u < \dim \theta\}$  embedded in  $\mathbf{S} = \{p_n(\cdot; \theta), \theta \in \Theta\}$ , and develop the higher order asymptotic theory. The third-order Edgeworth expansion for a class of estimators is derived. It is shown that the maximum likelihood estimator is still third-order asymptotically optimal in our general situation. However, the Edgeworth expansion contains two terms which vanish in the case of curved exponential family. Regarding this point we elucidate some results which did not appear in Amari's framework. Our results are applicable to time series analysis and multivariate analysis. We give a few examples (e.g., a family of curved ARMA models, a family of curved regression models). © 1994 Academic Press, Inc.

### 1. INTRODUCTION

Amari (1982, 1985) developed differential geometry of statistical inference for a curved exponential family  $\mathbf{M} = [q(x; u) | q(x; u) = \exp\{\theta'(u)x_i - \psi(\theta(u))\}]$  embedded in the exponential family  $\mathbf{S} = [p(x; \theta) | p(x; \theta) = \exp\{\theta'x_i - \psi(\theta)\}]$ . In this paper we extend his analysis to the case where the  $n$ -consecutive observations are (i) not necessarily identically distributed, and (ii) not necessarily independent. Suppose that the joint probability density function is given by  $p_n(\cdot; \theta)$ . Our main purpose is to estimate an unknown parameter  $u$  of a family of curved probability densities defined by

Received August 26, 1992; revised March 16, 1993.

AMS 1980 subject classifications: primary 62F12, 62M99; secondary 62E20.

Key words and phrases: curved probability density, curved ARMA models, curved regression model, higher order asymptotic theory, Edgeworth expansion, differential-geometrical method.

$\mathbf{M} = \{p_n(\cdot; \theta(u)), \dim u < \dim \theta\}$  embedded in  $\mathbf{S} = \{p_n(\cdot; \theta), \theta \in \Theta\}$ , and to develop the higher order asymptotic theory. Here we assumed that the observations come from  $\mathbf{M}$ .

In Section 2 the third-order Edgeworth expansion for a class of estimators of  $u$  is derived. We investigate higher order asymptotic properties of the maximum likelihood estimator in Section 3. Amari (1982, 1985) showed that the bias-adjusted maximum likelihood estimator is third-order asymptotically efficient for his i.i.d. curved exponential family. It can be shown that this result is still valid for our generalized case. However, our Edgeworth expansion contains two terms  $A_1^{ab}$  and  $A_3^{ab}$  which vanish in the case of a curved exponential family. The term  $A_3^{ab}$  elucidates the following. Suppose that a likelihood function  $l$  depends on a vector parameter  $(u, s)$ ,  $u \in \Theta_1$ ,  $s \in \Theta_2$ . When  $s$  is known we can construct the following two estimators  $\hat{u}$  and  $\bar{u}$  of  $u$ , which are defined by

$$\max_{(u, s) \in \Theta_1 \times \Theta_2} l(u, s) = l(\hat{u}, \hat{s}),$$

$$\max_{u \in \Theta_1} l(u, s) = l(\bar{u}, s),$$

respectively. If  $A_3^{ab} = 0$ , it is shown that the third-order asymptotics of  $\hat{u}$  and  $\bar{u}$  are identical, otherwise this result does not hold. Our results are general enough to be applicable to various fields (e.g., regression analysis, multivariate analysis, and time series analysis). In Section 4, we give a few examples (e.g., a family of curved ARMA models, a family of curved regression models). The family of curved ARMA models seems especially interesting because  $A_3^{ab} \neq 0$  for this family. Finally, we briefly mention a development of differential-geometrical methods in time series analysis. Amari (1983) discussed differential geometry of spectrum estimation and showed that the AR model is 1-flat and the MA model is -1-flat. Amari (1987) developed the  $\alpha$ -geometry for a parametric family of invertible linear systems. Ravishanker *et al.* (1990) characterized ARMA models as members of the curved exponential family. They gave a framework for the determination of parameter transformations for which (i) the asymptotic bias vanishes, (ii) the Fisher information matrix is a constant matrix, etc.

## 2. CURVED PROBABILITY DENSITIES AND EDGEWORTH EXPANSIONS

In this section we give a rigorous definition of a parametrized family of curved probability densities, which is a generalization of Amari's curved exponential family, and derive the Edgeworth expansions of estimators for the parameter.

Let  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)'$  be a collection of  $m$ -dimensional random vectors  $X_i$  which are not necessarily i.i.d. (i.e., our results can be applied to

regression analysis, multivariate analysis, and time series analysis). Let  $p_n(x_n; \theta)$  denote the probability density function of  $\mathbf{X}_n$  with respect to a carrier measure, where  $x_n \in \mathbb{R}^m$  and  $\theta = (\theta^1, \dots, \theta^p)' \in \Theta \subset \mathbb{R}^p$ . Now we are interested in a family of curved probability densities  $\mathbf{M} = \{p_n(x_n; \theta(u)) | u = (u^1, \dots, u^q), q < p\}$  embedded in  $\mathbf{S} = \{p_n(x_n; \theta) | \theta \in \Theta\}$ ; that is, the embedding map  $\theta = \theta(u)$  is smooth and its Jacobian matrix

$$\left\{ \frac{\partial \theta^i}{\partial u^a} \mid i = 1, \dots, p, a = 1, \dots, q \right\}$$

has full rank ( $=q$ ) everywhere in  $\mathbf{M}$ . The main purpose is to estimate the unknown parameter  $u$  under the assumption that the true probability density function of  $\mathbf{X}_n$  belongs to  $\mathbf{M}$ .

First, it is natural to estimate  $\theta$  in the ambient large class  $\mathbf{S}$ . Amari (1985) used a sufficient statistic  $\hat{\theta}_S$  for  $\theta$  in his analysis of the curved exponential family. In time series analysis Ravishanker *et al.* (1990) gave a characterization of ARMA models as curved exponential families embedded in the ordinary exponential family  $\mathbf{S}$ . However, for ARMA models, the dimension of sufficient statistics in  $\mathbf{S}$  is infinite (sample size) in general, so that the higher order asymptotic estimation theory cannot be developed by the differential-geometric approach. In this paper we use the maximum likelihood estimator (MLE)  $\hat{\theta}$  of  $\theta$  in  $\mathbf{S}$ . Since the MLE is sufficient for ordinary exponential family  $\mathbf{S}$ , our results contain Amari's ones. Also the MLE in  $\mathbf{S}$  is third-order asymptotically efficient (e.g., Amari (1985) for the i.i.d. case, Taniguchi (1986) for the dependent case). In general, it may be noted that the MLE is first-order asymptotically sufficient although it is not higher order asymptotically sufficient (e.g., Suzuki (1978) for the i.i.d. case, Taniguchi (1991) for the dependent case).

Turning to the main problem, consider the estimation of  $u$  in  $\mathbf{M}$  in terms of  $\hat{\theta}$  in  $\mathbf{S}$ . This leads us to solve the equation

$$\hat{\theta} = \theta(\hat{u})$$

with respect to  $\hat{u}$ , but this cannot be executed because of the fact that  $\dim \theta = p > \dim u = q$ . Therefore we introduce new extra coordinates  $v = (v^1, \dots, v^{p-q})$  such that  $w = (w^1, \dots, w^p) = (u, v) = (u^1, \dots, u^q, v^1, \dots, v^{p-q})$  becomes a coordinate system in  $\mathbf{S}$ . Then the equation

$$\hat{\theta} = \theta(\hat{u}, \hat{v}) \quad (2.1)$$

can uniquely solved with respect to  $\hat{u}$  and  $\hat{v}$ , where we note that the condition  $\theta(u, 0) = \theta(u)$  is assumed. Fixing  $u$ , we locally define the ancillary space  $A(u) = \{(u, v) | (u, v) \in \mathbf{S}\}$  so that the family  $\{A(u)\}$  defines a local foliation of  $\mathbf{S}$ . We can see that the determination of the estimator  $\hat{u}$  of  $u$  is in one to one correspondence with the introduction of the local foliation  $\{A(u)\}$ , which is called the ancillary family associated with the estimator  $\hat{u}$ .

Henceforth, we use  $\alpha, \beta, \gamma, \dots$  for indices of  $w$ -coordinates and related quantities; use  $a, b, c, \dots$  for indices of  $u$ -coordinates and related quantities; use  $\kappa, \mu, \lambda, \dots$  for indices of  $v$ -coordinates and related quantities. Also  $i, j, k, \dots$  are used for indices of  $\theta$ -coordinates and related quantities.

In order to discuss higher order asymptotic properties of  $\hat{u}$  defined by (2.1) we give the Edgeworth expansion of the distribution of  $\hat{u}$  in the following five steps:

- (1) Give the stochastic expansion of the MLE  $\hat{\theta}$  in  $S$ .
- (2) Derive the stochastic expansion of  $\hat{w} = (\hat{u}, \hat{v})$  defined in (2.1).
- (3) Evaluate the asymptotic moments (cumulants) of  $\hat{w}$ .
- (4) Using the general formula, give the Edgeworth expansion for  $\hat{w}$  in terms of the asymptotic moments (cumulants).
- (5) Derive the Edgeworth expansion of  $\hat{u}$ .

We now follow the above steps. Henceforth it is assumed that  $p_n(\cdot; \theta)$  is differentiable with respect to  $\theta$  up to necessary order. Define

$$\begin{aligned} Z_i &= n^{-1/2} \partial_i l_n(\theta), \\ Z_{ij} &= n^{-1/2} \{ \partial_i \partial_j l_n(\theta) - E_\theta [\partial_i \partial_j l_n(\theta)] \}, \\ Z_{ijk} &= n^{-1/2} \{ \partial_i \partial_j \partial_k l_n(\theta) - E_\theta [\partial_i \partial_j \partial_k l_n(\theta)] \}, \end{aligned}$$

where  $i, j, k = 1, \dots, p$ ,  $l_n(\theta) = \log p_n(\mathbf{X}_n; \theta)$ , and  $\partial/\partial\theta^i$  is abbreviated to  $\partial_i$ . We make the following assumption, which is very reasonable even in the non-i.i.d. case because it is satisfied by many regular statistical models (see Taniguchi (1986, 1991) for the dependent or non-identical case).

*Assumption 1.* The asymptotic moments (cumulants) of  $Z_i$ ,  $Z_{ij}$ , and  $Z_{ijk}$  are evaluated as follows:

$$\begin{aligned} E(Z_i Z_j) &= g_{ij} + O(n^{-1}), \\ E(Z_i Z_{jk}) &= J_{ijk} + O(n^{-1}), \\ E(Z_i Z_j Z_k) &= \frac{1}{\sqrt{n}} K_{ijk} + O(n^{-3/2}), \\ E(Z_i Z_{jkm}) &= L_{ijkm} + O(n^{-1}), \\ \text{cum}(Z_{ij}, Z_{km}) &= M_{ijkm} + O(n^{-1}), \\ E(Z_i Z_j Z_{km}) &= \frac{1}{\sqrt{n}} N_{ijkm} + O(n^{-3/2}), \\ \text{cum}(Z_i, Z_j, Z_k, Z_m) &= \frac{1}{n} H_{ijkm} + O(n^{-2}) \end{aligned}$$

and  $J$ th-order ( $J \geq 3$ ) cumulants of  $Z_i$ ,  $Z_{ij}$ , and  $Z_{ijk}$  are all  $O(n^{-J/2+1})$ .

We can define a connection in  $S$  by the coefficients

$$\Gamma_{ijk} = J_{ijk}, \quad i, j, k = 1, \dots, p,$$

in terms of which various geometric quantities such as the curvature of  $S$  can be calculated. We also note that this connection corresponds to the 1-connection in Amari's sense. To avoid many regularity conditions for the stochastic expansion of MLE, we also make the following assumption.

*Assumption 2.* The MLE of  $\hat{\theta}$  of  $\theta$  is defined by the equation

$$\frac{\partial}{\partial \theta} l_n(\hat{\theta}) = 0,$$

and the  $i$ th component of  $\hat{\theta}$  has the stochastic expansion

$$\begin{aligned} \hat{\theta}^i &= \theta^i + \frac{1}{\sqrt{n}} g^{ij} Z_j + \frac{1}{n} g^{ij} g^{km} Z_{jk} Z_m \\ &\quad + \frac{1}{2n} R_{jkm} g^{ij} g^{kk'} g^{mm'} Z_{k'} Z_{m'} + o_p(n^{-1}), \end{aligned} \quad (2.2)$$

where

$$R_{jkm} = -K_{jkm} - J_{jkm} - J_{kmj} - J_{mjk}$$

and  $g^{ij}$  denotes the  $(i, j)$ -component of the inverse matrix of  $g_{ij}$ . Here we adopt the Einstein summation convention.

Many regular statistical models satisfy Assumption 2 even if  $\mathbf{X}_n$  consists of non-i.i.d. random variables (e.g., Akahira and Takeuchi (1981) and Taniguchi (1986)).

Setting  $\tilde{u} = \sqrt{n}(\hat{u} - u)$ ,  $\tilde{v} = \sqrt{n}(\hat{v} - v)$ ,  $\tilde{w} = (\tilde{u}, \tilde{v})$ , and  $w = (u, 0)$ , we now proceed to derive the stochastic expansion of  $\tilde{w}$ , which is given by the following proposition. We put the proofs of the propositions and theorems in the Appendix.

**PROPOSITION 1.** Abbreviate  $\partial/\partial w^\alpha$  as  $\partial_\alpha$  and set  $B_i^\alpha = \partial_\alpha \theta^i$  and  $\bar{B}_i^\alpha = \partial_i w^\alpha$ . Then the  $\alpha$ th component of  $\tilde{w}^\alpha$  of  $\tilde{w}$  has the stochastic expansion.

$$\tilde{w}^\alpha = \tilde{e}^\alpha + \frac{1}{\sqrt{n}} \tilde{h}^\alpha + o_p(n^{-1/2}), \quad (2.3)$$

where

$$\tilde{e}^\alpha = \bar{B}_i^\alpha g^{ij} Z_j, \quad (2.4)$$

$$\begin{aligned} \tilde{h}^\alpha &= \bar{B}_i^\alpha g^{ij} g^{km} Z_m \{Z_{jk} - J_{ijk} g^{ll'} Z_{l'}\} \\ &\quad - \frac{1}{2} \{g^{\alpha\gamma} \bar{B}_k^\delta \bar{B}_m^\psi C_{\gamma\delta\psi}\} g^{kk'} g^{mm'} Z_{k'} Z_{m'}, \end{aligned} \quad (2.5)$$

and

$$C_{\gamma\delta\psi} = \lim_{n \rightarrow \infty} \frac{1}{n} E[\partial_\gamma l_n \partial_\delta l_n \partial_\psi l_n + \partial_\gamma l_n \partial_\delta \partial_\psi l_n],$$

with  $l_n = \log p_n(\mathbf{X}_n; \theta(w))$ .

*Remark.* Henceforth we use quantities with Greek indices  $g_{\alpha\beta}$ ,  $J_{\alpha\beta\gamma}$ , etc., for the corresponding quantities obtained by replacing the derivatives with respect to  $\theta$  ( $\partial_i$ ,  $i = 1, \dots, p$ ) of the quantities defined in Assumption 1 with those with respect to  $w$  ( $\partial_\alpha$ ,  $\alpha = 1, \dots, p$ ). For example,

$$g_{\alpha\beta} = \lim_{n \rightarrow \infty} \frac{1}{n} E[\partial_\alpha l_n \partial_\beta l_n].$$

Next we evaluate the asymptotic moments of  $\tilde{w}^\alpha$ . The first one is the second-order bias evaluation.

PROPOSITION 2. For  $\alpha = 1, \dots, p$ ,

$$E[\tilde{w}^\alpha] = n^{-1/2} \mu^\alpha + o(n^{-1}), \quad (2.6)$$

where

$$\mu^\alpha = -\frac{1}{2} g^{\alpha\gamma} g^{\delta\psi} C_{\gamma\delta\psi}.$$

Let  $O(n^{-1})$  terms of  $E[Z_i Z_j]$  be  $\eta_{ij}$ ; that is,

$$E(Z_i Z_j) = g_{ij} + \frac{1}{n} \eta_{ij} + o(n^{-1}). \quad (2.7)$$

Define  $S^\alpha = \sqrt{n} (\tilde{w}^\alpha - E\tilde{w}^\alpha)$  so that we have  $E[S^\alpha] = 0$ . Then we have the following proposition.

PROPOSITION 3. For  $\alpha, \beta = 1, \dots, p$ ,

$$E[S^\alpha S^\beta] = g^{\alpha\beta} + \frac{c^{\alpha\beta}}{n} + o(n^{-1}), \quad (2.8)$$

where

$$\begin{aligned} c^{\alpha\beta} = & -g^{\alpha\alpha'} g^{\beta\beta'} B_{\alpha'}^j B_{\beta'}^{j'} \eta_{jj'} + g^{\alpha\gamma} \partial_\gamma \mu^\beta + g^{\beta\gamma} \partial_\gamma \mu^\alpha \\ & + g^{\alpha\alpha'} g^{\beta\beta'} g^{\psi\psi'} \{M_{\alpha'\psi\beta'\psi'} - J_{\gamma\alpha'\psi} J_{\mu\beta'\psi'} g^{\gamma\mu}\} \\ & + \frac{1}{2} g^{\alpha\alpha'} g^{\beta\beta'} g^{\delta\delta'} g^{\psi\psi'} C_{\alpha'\delta\psi} C_{\beta'\delta'\psi'}. \end{aligned}$$

We state the following proposition without proof because of the fact that the idea and the method in the proof are quite similar to those of Propositions 2 and 3; further, the proof takes up far too much space.

PROPOSITION 4. For  $\alpha, \beta, \gamma, \delta = 1, \dots, p$ , we have

(i)

$$\begin{aligned} E[S^\alpha S^\beta S^\gamma] \\ = -n^{-1/2} g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} \{2K_{\alpha'\beta'\gamma'} + J_{\beta'\alpha'\gamma'} + J_{\alpha'\beta'\gamma'} + J_{\gamma'\alpha'\beta'}\} + o(n^{-1}) \\ = n^{-1/2} c^{\alpha\beta\gamma} + o(n^{-1}), \quad (\text{say}) \end{aligned}$$

(ii)

$$\begin{aligned} \text{cum}\{S^\alpha, S^\beta, S^\gamma, S^\delta\} \\ = -n^{-1} \left\{ 3H^{\alpha\beta\gamma\delta} + 2 \sum_{(\alpha\beta)(\gamma\delta)} N^{\alpha\beta\gamma\delta} \right. \\ \left. + \sum_{(\alpha)(\beta\gamma\delta)} L^{\alpha\beta\gamma\delta} - \frac{1}{2} \sum_{(\alpha)(\beta)(\gamma\delta)} \Gamma^{\alpha\beta\gamma\delta} \right\} + o(n^{-1}) \\ = n^{-1} c^{\alpha\beta\gamma\delta} + o(n^{-1}), \quad (\text{say}) \end{aligned}$$

where

$$\begin{aligned} \Gamma^{\alpha\beta\gamma\delta} &= g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} g^{\delta\delta'} g^{\zeta\zeta'} \\ &\quad \times (K_{\beta'\zeta'\alpha'} + J_{\zeta'\alpha'\beta'} + J_{\beta'\alpha'\zeta'}) (2K_{\zeta'\gamma'\delta'} + J_{\gamma'\zeta'\delta'} + J_{\delta'\zeta'\gamma'}) \\ H^{\alpha\beta\gamma\delta} &= g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} g^{\delta\delta'} H_{\alpha'\beta'\gamma'\delta'}, \\ N^{\alpha\beta\gamma\delta} &= g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} g^{\delta\delta'} N_{\alpha'\beta'\gamma'\delta'}, \\ M^{\alpha\beta\gamma\delta} &= g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} g^{\delta\delta'} M_{\alpha'\beta'\gamma'\delta'}, \\ L^{\alpha\beta\gamma\delta} &= g^{\alpha\alpha'} g^{\beta\beta'} g^{\gamma\gamma'} g^{\delta\delta'} L_{\alpha'\beta'\gamma'\delta'}, \end{aligned}$$

and  $\sum_{(\alpha)(\beta)(\gamma)(\delta)}$  means the sum over all the combinations  $(\alpha)(\beta)(\gamma)(\delta)$ ,

(iii)

$$\text{cum}\{S^{\alpha_1}, S^{\alpha_2}, \dots, S^{\alpha_J}\} = o(n^{-J/2+1}), \quad \text{for } J \geq 5.$$

Next, we turn to seek the Edgeworth expansion for the estimators. To this end we make the following assumption.

*Assumption 3.* The validity of the Edgeworth expansions for the estimators  $\hat{\theta}$ ,  $\hat{w}$ , and  $\hat{u}$  holds.

This assumption is reasonable because for the i.i.d. case Bhattacharya and Ghosh (1978) proved the validity of the Edgeworth expansion for a wide class of estimators, and Taniguchi (1987) also proved the validity of the Edgeworth expansion of minimum contrast estimators, which contain the MLE, for Gaussian ARMA processes.

Now we may state

THEOREM 5. For  $\tilde{u} = \sqrt{n}(\hat{u} - u)$ ,

$$\begin{aligned} P\{\tilde{u}^1 < x^1, \dots, \tilde{u}^q < x^q\} \\ = \int_{-\infty}^{x^1} \dots \int_{-\infty}^{x^q} N(x; \mathcal{F}) \left[ 1 + n^{-1/2} \mu^a H_a(x) \right. \\ \left. + \frac{1}{6\sqrt{n}} c^{abc} H_{abc}(x) + \frac{1}{2n} (c^{ab} + \mu^a \mu^b) H_{ab}(x) \right. \\ \left. + \left\{ \frac{1}{24n} c^{abcd} + \frac{1}{6n} \mu^a c^{bcd} \right\} H_{abcd}(x) \right. \\ \left. + \frac{1}{72n} c^{abc} c^{a'b'c'} H_{abca'b'c'}(x) \right] dx + o\left(\frac{1}{n}\right), \end{aligned}$$

where  $x = (x^1, \dots, x^q)'$ ,

$$N(x; \mathcal{F}) = (2\pi)^{-q/2} |\mathcal{F}|^{-1/2} \exp\left(-\frac{1}{2} x' \mathcal{F}^{-1} x\right),$$

$$H_{j_1, \dots, j_s}(x) = \frac{(-1)^s}{N(x; \mathcal{F})} \frac{\partial^s}{\partial x_{j_1} \dots \partial x_{j_s}} N(x; \mathcal{F}),$$

and indices  $a, b, c, d, a', b', c'$  run from 1 to  $q$ . Here

$$\mathcal{F} = \{g_{ab} - g_{a\kappa} g_{b\mu} g^{\kappa\mu}\}^{-1}, \quad (2.10)$$

where indices  $\kappa$  and  $\mu$  run from  $q+1$  to  $p$ .

### 3. HIGHER ORDER OPTIMALITY OF THE MLE

In this section we discuss the higher order asymptotic optimality of the MLE. Recall Theorem 5. Then the matrix  $\mathcal{F}$  defined by (2.10) is minimized if and only if

$$g_{a\kappa} \equiv 0, \quad \text{for } a = 1, \dots, q, \quad \kappa = q+1, \dots, p. \quad (3.1)$$

That is, the estimator  $\hat{u}$  is first-order asymptotically efficient if and only if the ancillary space of  $\hat{u}$  is orthogonal to  $M$ . Henceforth we restrict ourselves to the class  $\mathcal{A}_1$  of the first-order asymptotically efficient estimators for  $u$ . To discuss the higher order asymptotics of  $\hat{u} \in \mathcal{A}_1$ , we make a modification of the second-order bias of  $\hat{u}$ . In view of Proposition 2, we modify  $\hat{u} = (\hat{u}^\alpha)$  as

$$\hat{u}^{*\alpha} = \hat{u}^\alpha - \frac{1}{n} \mu^\alpha(\hat{u}), \quad \alpha = 1, \dots, q. \quad (3.2)$$

For this modified estimator, we have the following theorems.

**THEOREM 6.** *Let  $\tilde{u}^{*\alpha} = \sqrt{n} (\hat{u}^{*\alpha} - u^\alpha)$ ,  $\alpha = 1, \dots, q$ . Then*

$$\begin{aligned} & P\{\tilde{u}^{*1} < x^1, \dots, \tilde{u}^{*q} < x^q\} \\ &= \int_{-\infty}^{x^1} \cdots \int_{-\infty}^{x^q} N(x; \mathcal{F}_0) \left[ 1 + \frac{1}{6\sqrt{n}} c^{abc} H_{abc}(x) \right. \\ &\quad + \frac{1}{2n} (A_1^{ab} + A_2^{ab} + A_3^{ab} + A_4^{ab} + A_5^{ab}) H_{ab}(x) \\ &\quad + \frac{1}{24n} c^{abcd} H_{abcd}(x) \\ &\quad \left. + \frac{1}{72n} c^{abc} c^{a'b'c'} H_{abca'b'c'}(x) \right] dx + o\left(\frac{1}{n}\right), \end{aligned} \quad (3.3)$$

where  $\mathcal{F}_0 = \{g^{ab}\}$ ,

$$A_1^{ab} = -g^{aa'} g^{bb'} B_{a'}^j B_{b'}^{j'} \eta_{jj'}, \quad (3.4)$$

$$A_2^{ab} = g^{aa'} g^{bb'} g^{cc'} \{M_{a'cb'c'} - J_{da'c} J_{d'b'c'} g^{dd'}\} \quad (3.5)$$

$$A_3^{ab} = g^{aa'} g^{bb'} g^{\kappa\kappa'} \{M_{a'\kappa b'\kappa'} - J_{\alpha a'\kappa} J_{\alpha' b'\kappa'} g^{\alpha\alpha'}\}, \quad (3.6)$$

$$A_4^{ab} = \frac{1}{2} g^{aa'} g^{bb'} g^{\kappa\kappa'} g^{\mu\mu'} C_{a'\kappa\mu} C_{b'\kappa'\mu'}, \quad (3.7)$$

$$A_5^{ab} = \frac{1}{2} g^{aa'} g^{bb'} g^{c'c'} g^{d'd'} C_{a'cd} C_{b'c'd'}. \quad (3.8)$$

Here indices  $a, a', b, b', c, c', d, d'$  run from 1 to  $q$  and  $\kappa, \kappa', \mu, \mu'$  run from  $q+1$  to  $p$ , and  $\alpha, \alpha'$  run from 1 to  $p$ .

**THEOREM 7.**

$$\begin{aligned} E\{\tilde{u}^{*e} \tilde{u}^{*f}\} &= g^{ef} + \frac{1}{n} (A_1^{ef} + A_2^{ef} + A_3^{ef} + A_4^{ef} + A_5^{ef}) \\ &\quad + o\left(\frac{1}{n}\right), \quad e, f = 1, 2, \dots, q. \end{aligned}$$

It can be shown that the quantities  $c^{abc}$  and  $c^{abcd}$  in Theorem 6 are independent of the choice of an estimator (or the ancillary space associated with it). Thus, in view of Theorems 6 and 7, we can see that the terms  $A_i^{ef}$  ( $i=1, 2, \dots, 5$ ) are the key quantities which describe the third-order asymptotic optimality of the estimator. We call  $\hat{u} \in \mathcal{A}_1$  the third-order asymptotically efficient estimator if the matrix  $\{A_1^{ef} + A_2^{ef} + A_3^{ef} + A_4^{ef} + A_5^{ef}\}$  is minimized. We must investigate the quantities  $A_i^{ef}$  in detail. To begin with, we give the table of the corresponding quantities in the case for the i.i.d. curved exponential family investigated by Amari (1985).

Here we explain the meaning of the five terms  $A_i^{ab}$  ( $i=1, \dots, 5$ ) in more detail. If  $X_i$  are i.i.d. then it follows from the definition (2.7) of  $\eta_{ij}$  that  $A_1^{ab}=0$ , whence  $A_1^{ab}$  represents an effect of non-i.i.d. case. To explain the meaning of  $A_2^{ab}$ , we rewrite it as

$$A_2^{ab} = g^{aa'} g^{bb'} g^{cc'} \{M_{a'cb'c'} - J_{\alpha a'c} J_{\alpha' b'c'} g^{\alpha\alpha'}\} \\ + g^{aa'} g^{bb'} g^{cc'} \{J_{\kappa a'c} J_{\kappa' b'c'} g^{\kappa\kappa'}\}, \quad (3.9)$$

where  $a, a', b, b', c, c'$  run from 1 to  $q$  and  $\kappa, \kappa'$  run from  $q+1$  to  $p$ , and  $\alpha, \alpha'$  run from 1 to  $p$ . Applying  $\partial_c$  to  $g_{\kappa a'} = 0$  we have  $J_{\kappa a'c} + C_{a'c\kappa} = 0$ . Thus we obtain

$$A_2^{ab} = g^{aa'} g^{bb'} g^{cc'} \{M_{a'cb'c'} - J_{\alpha a'c} J_{\alpha' b'c'} g^{\alpha\alpha'}\} \\ + g^{aa'} g^{bb'} g^{cc'} g^{\kappa\kappa'} C_{a'c\kappa} C_{b'c'\kappa'} \\ = H_{ab}^{eM} + H_{ab}^{eM}, \quad (\text{say}).$$

Here  $H_{ab}^{eM}$  corresponds to the exponential curvature of  $\mathbf{M}$  in the case of curved exponential family. The term  $H_{ab}^{SM}$  is the  $a$ -direction exponential curvature of  $\mathbf{S}$  (i.e., the residual of  $Z_{ab}$  after linear regression on  $Z_i$ ), which represents a loss of information by use of the MLE  $\hat{w}$  of  $\mathbf{S}$ . We can also see that  $A_3^{ab}$  is the  $(a, \kappa)$ -direction exponential curvature of  $\mathbf{S}$  (i.e., the residual of  $Z_{a\kappa}$  after linear regression on  $Z_i$ ), which represents a loss of information of  $\hat{w}$ . Table I shows that the  $A_1^{ab}$  and  $A_3^{ab}$  terms present a sharp contrast

TABLE I

In the Case of i.i.d. Curved Exponential Family Investigated by Amari	
$A_1^{ab}$	Vanish (=0)
$A_2^{ab}$	{Exponential curvature of $\mathbf{M}$ } <sup>2</sup> , which is independent of the choice of an estimator
$A_3^{ab}$	Vanish (=0)
$A_4^{ab}$	{Mixture curvature of the ancillary space} <sup>2</sup> , which depends on the choice of an estimator, and is minimized by the MLE
$A_5^{ab}$	{Mixture connection of $\mathbf{M}$ } <sup>2</sup> , which is independent of the choice of an estimator

between Amari's results and ours. Returning to our generalized situation, we can state whether  $A_i^{cf}$  are independent of the choice of estimators or not.

**THEOREM 8.** *In the notations of Theorem 6,*

(i)  $A_1^{ab}$ ,  $A_2^{ab}$ ,  $A_3^{ab}$ , and  $A_5^{ab}$  are independent of the choice of an estimator (the ancillary space associated with it), and  $A_4^{ab}$  depends on the choice of an estimator.

(ii) *If*

$$M_{ij i' j'} - J_{kij} J_{k' i' j'} g^{kk'} = 0, \quad (3.10)$$

for all  $i, j, i' j' = 1, \dots, p$  (indices of the  $\theta$  coordinates), then  $A_3^{ab} = 0$ .

Here it should be noted that the left-hand side of (3.10) is the residual of  $Z_{ij}$  after linear regression on  $Z_k$ . It is not difficult to show that if (3.10) holds, then  $S$  is flat (1-flat); that is, the Riemann-Christoffel curvature calculated in terms of the connection coefficients  $J_{ijk}$  vanishes.

Let  $\hat{u}$  be the MLE of  $u$  defined by

$$\frac{\partial}{\partial u^a} \log p_n(\mathbf{X}_n; \theta(u, 0))|_{u=\hat{u}} = 0, \quad a = 1, \dots, q. \quad (3.11)$$

Then the ancillary space  $A(u)$  associated with  $\hat{u}$  is defined by

$$\begin{aligned} A(u) = & \left\{ \eta = \theta(u, v) \left| \lim_{n \rightarrow \infty} \frac{1}{n} \int \frac{\partial}{\partial u^a} \log p_n(x_n; \theta(u, 0)) p_n(x_n; \theta(u, v)) dx_n = 0 \right. \right\} \\ & = \left\{ \eta = \theta(u, v) \left| \lim_{n \rightarrow \infty} \frac{1}{n} E_\eta \left[ \frac{\partial}{\partial u^a} \log p_n(\mathbf{X}_n; \theta(u, 0)) \right] = 0 \right. \right\}. \end{aligned} \quad (3.12)$$

The third-order optimality of  $\hat{u}$  is extended to our case beyond the curved exponential family.

**THEOREM 9.** *The bias-adjusted MLE  $\hat{u}^* = (\hat{u}^{**})$  defined by (3.2) is third-order asymptotically efficient.*

*Remark.* This theorem states that  $\hat{u}^*$  is third-order asymptotically efficient in the restricted framework of functions of the MLE, and it includes the loss of information due to restricting the estimators as functions of the MLE of the larger model  $S$ .

Our results elucidate the following fact—which is not involved in Amari's framework. Suppose that the true distribution of  $\mathbf{X}_n$  is specified by

an unknown parameter  $u = (u^1, \dots, u^q)'$  completely, then we can calculate the MLE  $\bar{u}$  by the usual method. On the other hand, we reparametrize the true model by a higher dimensional unknown parameter  $\theta = (\theta^1, \dots, \theta^p)'$  ( $p > q$ ), and construct the MLE  $\hat{u}$  of  $u$  based on the MLE  $\hat{\theta}$  of  $\theta$  by solving Eq. (2.1). Then we have

**THEOREM 10.** *If  $A_3^{ab} = 0$ ,  $a, b = 1, \dots, q$ , the third-order Edgeworth expansion for the bias-adjusted estimator  $\hat{u}^*$  of  $\hat{u}$  and that for the bias-adjusted estimator  $\bar{u}^*$  of  $\bar{u}$  are identical. However, if  $A_3^{ab} \neq 0$ , this is not true.*

*Remark.* This theorem means that if the true model with the parameter  $u$  is embedded in a class of models with parameter  $\theta$  ( $\dim \theta > \dim u$ ) satisfying (3.10), then the third-order asymptotics of  $\hat{u}^*$  calculated from the MLE of the higher dimensional model is equal to that of  $\bar{u}^*$ . But this is not true if  $A_3^{ab} \neq 0$ . Thus  $A_3^{ab}$  becomes a key quantity in our family of curved probability densities; it seems to be a measure of higher order asymptotic insufficiency of the MLE  $\hat{\theta}$ .

#### 4. EXAMPLES

In this section we give three examples: the first two are taken from time series analysis and the last is taken from regression analysis. The calculation in the former two examples is so involved that we use the well known computer algebra system REDUCE.

**EXAMPLE 1:** Curved AR(1)-Model. Let  $\mathbf{S}$  be a class of AR(1)-models defined by

$$\mathbf{S} = \{ \{X_t\} \mid X_t = (\theta^1) X_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d. } N(0, (\theta^2)) \}.$$

In terms of the spectral density, the family  $\mathbf{S}$  can be identified with the set

$$\mathbf{S} = \left\{ f_{\theta}(\lambda) = \frac{(\theta^2)}{2\pi} \frac{1}{|1 - (\theta^1) e^{i\lambda}|^2}, \theta = ((\theta^1), (\theta^2))' \right\}.$$

Let  $\mathbf{M}$  be a class of curved AR(1)-models defined by

$$\mathbf{M} = \{ \{X_t\} \mid X_t = u X_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d. } N(0, 1) \}.$$

In terms of the spectral density,  $\mathbf{M}$  can be identified with

$$\mathbf{M} = \left\{ f_{\theta(u)}(\lambda) = \frac{1}{2\pi} \frac{1}{|1 - u e^{i\lambda}|^2} \right\}.$$

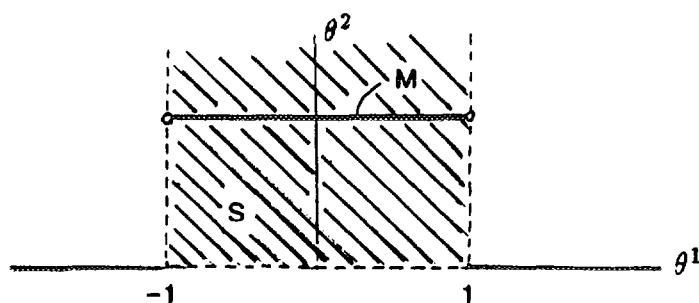


FIGURE 1

We assume that the true model belongs to **M**. Figure 1 illustrates **M** embedded in **S**.

The ancillary space associated with the MLE  $\hat{u}$  is given by

$$\begin{aligned} A(u) &= \left\{ ((\theta^1), (\theta^2)) \left| \frac{\partial}{\partial u} \int_{-\pi}^{\pi} f_{\theta}(\lambda) f_{\theta(u)}(\lambda)^{-1} d\lambda = 0 \right. \right\} \\ &= \{ ((\theta^1), (\theta^2)) | 2(\theta^2)((\theta^1) - u) = 0 \} \\ &= \{ ((\theta^1), (\theta^2)) | (\theta^1) = u, (\theta^2) = 1 + dv, d: \text{an arbitrary real number} \}. \end{aligned}$$

Hence,

$$\hat{u} = (\hat{\theta}^1), \quad (4.1)$$

where  $\hat{\theta}^1$  is the MLE of  $(\theta^1)$ . According to Anderson (1971, p. 354),  $(\hat{\theta}^1)$  is given as the solution of two simultaneous equations with respect to  $(\theta^1)$  and  $(\theta^2)$ . It can be shown that (3.10) is valid in **S**; so **S** is flat and  $A_3^{ab} = 0$ . This together with Theorem 10 implies that the third-order Edgeworth expansion for  $\hat{u}^* = (\hat{\theta}^{1*})$  is equal to that for  $\bar{u}^*$ . That is, we can use the MLE  $(\hat{\theta}^1)$  on **S** in place of the MLE  $\bar{u}$  on **M** up to third-order.

**EXAMPLE 2: Curved ARMA Model.** Let **S** be a class of ARMA(1, 1)-models defined by

$$\mathbf{S} = \{ \{X_t\} | X_t - (\theta^1) X_{t-1} = \varepsilon_t - (\theta^2) \varepsilon_{t-1}, \varepsilon_t \sim \text{i.i.d. } N(0, 1) \}.$$

The spectral density of  $\{X_t\} \in \mathbf{S}$  is given by

$$f_{\theta}(\lambda) = \frac{1}{2\pi} \frac{|1 - (\theta^2)e^{i\lambda}|^2}{|1 - (\theta^1)e^{i\lambda}|^2}.$$

Let  $\mathbf{M}$  be a class of ARMA(1, 1)-models defined by

$$\mathbf{M} = \{ \{X_t\} \mid X_t - uX_{t-1} = \varepsilon_t + u\varepsilon_{t-1}, \varepsilon_t \sim \text{i.i.d. } N(0, 1) \},$$

which is identified with spectral form

$$\mathbf{M} \left\{ f_{\theta(u)}(\lambda) = \frac{1}{2\pi} \frac{|1 + ue^{i\lambda}|^2}{|1 - ue^{i\lambda}|^2} \right\}.$$

Figure 2 illustrates  $\mathbf{M}$  embedded in  $\mathbf{S}$ .

Assume that the true model belongs to  $\mathbf{M}$ . Then the ancillary space associated with the MLE  $\hat{u}$  is given by

$$\begin{aligned} A(u) = \{ & ((\theta^1), (\theta^2)) \mid (\theta^1) = u + v, \\ & (\theta^2) = a \text{ root of the equation with respect to } b: \\ & (u^5 - u + 2u^3v^2 + 3u^4v + v)b^2 \\ & + (u^6 + u^4 - u^2 - 1 + u^4v^2 + 4u^2v^2 \\ & - v^2 + 2u^5v + 4u^3v + 2uv)b \\ & + u^5 - u + 2u^3v^2 + 3u^4v + v = 0 \}. \end{aligned}$$

By a straightforward calculation (using REDUCE) we have

$$A_3^{ab} = \frac{1 + u^2}{4} \neq 0.$$

Therefore, in view of Theorem 10, the third-order asymptotics of  $\hat{u}^*$  is not equivalent to that of  $\bar{u}^*$ . So, we have to be careful of the third-order difference between  $\hat{u}^*$  and  $\bar{u}^*$ .

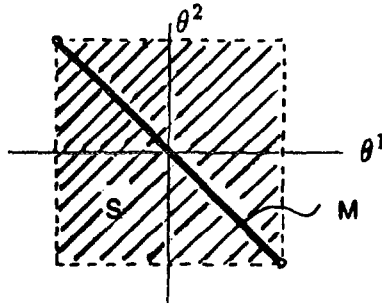


FIGURE 2

We can construct infinitely many third-order asymptotically efficient estimators. Let  $\hat{u}$  be an estimator whose ancillary space is given by

$$\begin{aligned}(\theta^1) &= u + (1 + qu)v, \\ (\theta^2) &= -u + (1 + qu)v + \frac{2q^2u^5 + 4qu^4 + 2u^3}{1 - u^4}v^2,\end{aligned}$$

where  $q$  is an arbitrary real number. It is not difficult to show that  $\hat{u}$  is third-order asymptotically efficient.

**EXAMPLE 3: Curved Regression Model.** As an example of the non-identically distributed case, we consider the following class of regression models:

$$\begin{aligned}\mathbf{S} = \left\{ \{X_t\} \mid X_t &= (\theta^1) + (\theta^2)Z_t + \varepsilon_t, t = 1, \dots, n, \right. \\ \left. Z_t = \cos\left(\frac{2\pi jt}{n}\right), j \text{ a known integer, } \varepsilon_t \sim \text{i.i.d. } N(0, 1) \right\}.\end{aligned}$$

Let  $\mathbf{M}$  be a class of curved regression models defined by

$$\mathbf{M} = \{ \{X_t\} \mid X_t = u + u^2Z_t + \varepsilon_t, t = 1, \dots, n \},$$

(See Fig. 3.)

Assume that the true model belongs to  $\mathbf{M}$ . The ancillary space associated with  $\hat{u}$  is given by

$$A(u) = \{((\theta^1), (\theta^2)) \mid (\theta^1) = u - 2uv, (\theta^2) = u^2 + v\}.$$

Since (3.10) holds in  $\mathbf{S}$ , it is flat and  $A_3^{ab} = 0$ . Hence, from Theorem 10, we see that the third-order asymptotics of  $\hat{u}^*$  is identical with that of  $\bar{u}^*$ .

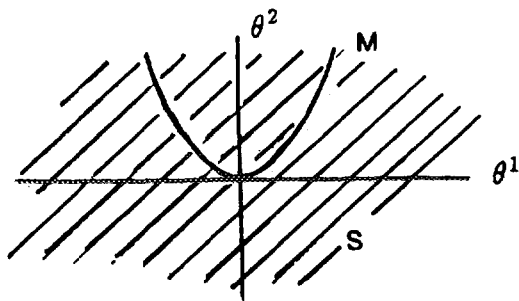


FIGURE 3

## APPENDIX

*Proof of Proposition 1.* Recalling (2.2), we can rewrite the  $i$ th component of (2.1) as

$$\theta^i + \frac{1}{\sqrt{n}} g^{ij} Z_j + \frac{1}{n} g^{ij} g^{km} Z_{jk} Z_m + \frac{1}{2n} R_{jkm} g^{ij} g^{kk'} g^{mm'} Z_{k'} Z_{m'} + o_p\left(\frac{1}{n}\right) \\ = \theta^i(\hat{u}, \hat{v}), \quad (5.1)$$

where all the quantities are evaluated at  $w = (u, 0)$ . Expansion of  $\theta^i(\hat{u}, \hat{v}) = \theta^i(w + (1/\sqrt{n})\tilde{w})$  at  $w$  in (5.1) yields the following stochastic equation

$$\frac{1}{\sqrt{n}} g^{ij} Z_j + \frac{1}{n} g^{ij} g^{km} Z_{jk} Z_m + \frac{1}{2n} R_{jkm} g^{ij} g^{kk'} g^{mm'} Z_{k'} Z_{m'} \\ = \frac{1}{\sqrt{n}} \{ \partial_\alpha \theta^i \} \tilde{w}^\alpha + \frac{1}{2n} \{ \partial_\alpha \partial_\beta \theta^i \} \tilde{w}^\alpha \tilde{w}^\beta + \text{higher order terms.} \quad (5.2)$$

Substituting

$$\tilde{w}^\alpha = \tilde{e}^\alpha + \frac{1}{\sqrt{n}} \tilde{h}^\alpha + \text{higher order terms}$$

to (5.2), and comparing the coefficients of the powers of  $n^{-1/2}$  and  $n^{-1}$ , respectively, we have

$$B_\alpha^i \tilde{e}^\alpha = g^{ij} Z_j, \quad (5.3)$$

$$B_\alpha^i \tilde{h}^\alpha = g^{ij} g^{km} Z_{jk} Z_m + \frac{1}{2} R_{jkm} g^{ij} g^{kk'} g^{mm'} Z_{k'} Z_{m'} \\ - \frac{1}{2} (\partial_\beta B_\alpha^i) \tilde{e}^\alpha \tilde{e}^\beta. \quad (5.4)$$

The following fundamental relations in the coordinate transformations can be easily verified:

$$\bar{B}_\alpha^i B_\beta^j = \delta(\alpha, \beta), \quad (\text{Kronecker's } \delta) \quad (5.5)$$

$$g^{ij} = B_\alpha^i B_\beta^j g^{\alpha\beta}, \quad (5.6)$$

$$K_{jkm} = \bar{B}_j^\gamma \bar{B}_k^\delta \bar{B}_m^\psi K_{\gamma\delta\psi}, \quad (5.7)$$

$$J_{jkm} = \bar{B}_j^\gamma \bar{B}_k^\delta (\partial_\delta \bar{B}_m^\psi) g_{\psi\gamma} + \bar{B}_j^\gamma \bar{B}_m^\psi \bar{B}_k^\delta J_{\gamma\delta\psi}, \quad (5.8)$$

$$(\partial_\delta B_\alpha^i) \bar{B}_k^{\alpha'} + B_\alpha^i (\partial_\delta \bar{B}_k^{\alpha'}) = 0. \quad (5.9)$$

By using (5.5)–(5.9), we can solve Eqs.(5.3) and (5.4) with respect to  $\tilde{e}^\alpha$  and  $\tilde{h}^\alpha$ , obtaining expressions (2.4) and (2.5), respectively.

*Proof of Proposition 2.* If we proceed to evaluate, in the same way as in Proposition 1, the stochastic expansion of  $\tilde{w}^x$  up to third-order (i.e.,  $\tilde{e}^x + n^{-1/2}\tilde{h}^x + n^{-1}\tilde{\xi}^x + \dots$ ), we can observe that  $\tilde{\xi}^x$  is a third-order polynomial of  $Z_i$ ,  $Z_{ij}$ , and  $Z_{ijk}$ . Hence  $E[\tilde{\xi}^x] = O(n^{-1/2})$ . It follows from (2.5) that

$$\begin{aligned} E[\tilde{h}^x] &= -\frac{1}{2} g^{xy} \bar{B}_k^\delta \bar{B}_m^\psi C_{\gamma\delta\psi} g^{kk'} g^{mm'} g_{k'm'} \\ &= -\frac{1}{2} g^{xy} \bar{B}_k^\delta \bar{B}_m^\psi C_{\gamma\delta\psi} B_{\alpha'}^k B_{\beta'}^m g^{\alpha'\beta'} \quad (\text{by (5.5)}) \\ &= -\frac{1}{2} g^{xy} g^{\delta\psi} C_{\gamma\delta\psi}. \end{aligned}$$

Noting  $E[\tilde{e}^x] = 0$ , we get (2.6).

*Proof of Proposition 3.* Let  $W_\theta$  be a measurable function of  $\mathbf{X}_n$  depending on  $\theta$ . Suppose that  $W_\theta$  is differentiable with respect to  $\theta$ , and that the expectation  $E_\theta$  with respect to  $p_n(x_n; \theta)$  commutes with  $\partial_i$ . Then the calculation of  $\partial_i\{E_\theta[W_\theta]\}$  yields the following formula.

$$E_\theta[W_\theta Z_i] = \frac{1}{\sqrt{n}} \partial_i\{E_\theta[W_\theta]\} - \frac{1}{\sqrt{n}} E_\theta[\partial_i W_\theta]. \quad (5.10)$$

It follows that

$$\begin{aligned} \partial_i S^x &= \sqrt{n} \{\partial_i \hat{w}^x - \partial_i E(\hat{w}^x)\} \\ &= \sqrt{n} \{\partial_i w^x\} - \frac{1}{\sqrt{n}} \partial_i \mu^x + o\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

since  $\hat{w}^x$  is independent of  $\theta$ . Applying (5.10) to  $S^x$ , we have

$$E[S^x Z_i] = \bar{B}_i^x + \frac{1}{n} \bar{B}_i^y \partial_y \mu^x + o\left(\frac{1}{n}\right). \quad (5.11)$$

Then

$$\begin{aligned} E\{S^\alpha S^\beta\} &= E\{\tilde{e}^\alpha + (S^\alpha - \tilde{e}^\alpha)\}\{\tilde{e}^\beta + (S^\beta - \tilde{e}^\beta)\} \\ &= E\{S^\alpha \tilde{e}^\beta\} + E\{S^\beta \tilde{e}^\alpha\} - E\{\tilde{e}^\alpha \tilde{e}^\beta\} \\ &\quad + E\{(S^\alpha - \tilde{e}^\alpha)(S^\beta - \tilde{e}^\beta)\}. \end{aligned} \quad (5.12)$$

From (2.4) and (2.7) we have

$$E\{\tilde{e}^\alpha \tilde{e}^\beta\} = g^{\alpha\beta} + \frac{1}{n} g^{\alpha\alpha'} g^{\beta\beta'} B_{\alpha'}^j B_{\beta'}^{j'} \eta_{jj'} + o\left(\frac{1}{n}\right). \quad (5.13)$$

By (2.4), we obtain

$$\begin{aligned}
E\{S^\alpha \tilde{e}^\beta\} &= E\{S^\alpha \bar{B}_i^\beta g^{ij} Z_j\} \\
&= \bar{B}_i^\beta g^{ij} E\{S^\alpha Z_j\} \\
&= \bar{B}_i^\beta g^{ij} \{\bar{B}_j^\alpha + n^{-1} \bar{B}_j^\gamma \partial_\gamma \mu^\alpha\} + o\left(\frac{1}{n}\right) \quad (\text{by (5.11)}) \\
&= g^{\alpha\beta} + \frac{1}{n} g^{\beta\gamma} \partial_\gamma \mu^\alpha + o\left(\frac{1}{n}\right). \tag{5.14}
\end{aligned}$$

The fourth term of (5.12) is written as

$$\frac{1}{n} \text{Cov}(\tilde{h}^\alpha, \tilde{h}^\beta) + o\left(\frac{1}{n}\right). \tag{5.15}$$

For general random variables  $U_1, \dots, U_4$  it is known that

$$\begin{aligned}
&\text{Cov}(U_1 U_2, U_3 U_4) \\
&= \text{Cov}(U_1, U_3) \text{Cov}(U_2, U_4) + \text{Cov}(U_1, U_4) \text{Cov}(U_2, U_3) \\
&\quad + \text{cum}\{U_1, U_2, U_3, U_4\}, \tag{5.16}
\end{aligned}$$

(see Brillinger (1975)). We can see that  $Z_{jk} - J_{ljk} g^{ll'} Z_{l'}$  are asymptotically orthogonal to  $Z_r$  because

$$\begin{aligned}
\text{Cov}\{Z_{jk} - J_{ljk} g^{ll'} Z_{l'}, Z_r\} &= J_{rjk} - J_{ljk} g^{ll'} g_{l'r} + O\left(\frac{1}{n}\right) \\
&= O\left(\frac{1}{n}\right). \tag{5.17}
\end{aligned}$$

In view of (2.5), (5.16), and (5.17) we can show that

$$\begin{aligned}
\text{Cov}(\tilde{h}^\alpha, \tilde{h}^\beta) &= \text{Cov}[\bar{B}_i^\alpha g^{ij} g^{km} Z_m \{Z_{jk} - J_{ljk} g^{ll'} Z_{l'}\}, \\
&\quad \bar{B}_{i'}^\beta g^{i'j'} g^{k'm'} Z_{m'} \{Z_{j'k'} - J_{sj'k'} g^{ss'} Z_{s'}\}] \\
&\quad + \frac{1}{4} \text{Cov}[\{g^{\alpha\gamma} \bar{B}_k^\delta \bar{B}_m^\psi C_{\gamma\delta\psi}\} g^{kk'} g^{mm'} Z_{k'} Z_{m'}, \\
&\quad \{g^{\beta\gamma'} \bar{B}_{s'}^{\delta'} \bar{B}_{t'}^{\psi'} C_{\gamma'\delta'\psi'}\} g^{ss'} g^{tt'} Z_{s'} Z_{t'}] + o(1) \\
&= g^{\alpha\alpha'} g^{\beta\beta'} g^{\psi\psi'} \{M_{\alpha'\psi\beta'\psi'} - J_{\gamma\alpha'\psi} J_{\mu\beta'\psi'} g^{\gamma\mu}\} \\
&\quad + \frac{1}{2} g^{\alpha\alpha'} g^{\beta\beta'} g^{\delta\delta'} g^{\psi\psi'} C_{\alpha'\delta'\psi} C_{\beta'\delta'\psi'} + o(1). \tag{5.18}
\end{aligned}$$

The proof follows from (5.12), (5.13), (5.14), (5.15), and (5.18).

*Proof of Theorem 5.* First, we derive the Edgeworth expansion for  $\tilde{w}$ . Applying a general formula (see Taniguchi (1986, p. 6)) to Propositions 2–4, we get

$$\begin{aligned}
 & P(\tilde{w}^1 < y^1, \dots, \tilde{w}^p < y^p) \\
 &= \int_{-\infty}^{y^1} \cdots \int_{-\infty}^{y^p} N(y; G) \left[ 1 + \frac{1}{\sqrt{n}} \mu^\alpha H_\alpha(y) + \frac{1}{6\sqrt{n}} c^{\alpha\beta\gamma} H_{\alpha\beta\gamma}(y) \right. \\
 &\quad + \frac{1}{2n} (c^{\alpha\beta} + \mu^\alpha \mu^\beta) H_{\alpha\beta}(y) + \left\{ \frac{1}{24n} c^{\alpha\beta\gamma\delta} + \frac{1}{6n} \mu^\alpha c^{\beta\gamma\delta} \right\} H_{\alpha\beta\gamma\delta}(y) \\
 &\quad \left. + \frac{1}{72n} c^{\alpha\beta\gamma} c^{\alpha'\beta'\gamma'} H_{\alpha\beta\gamma\alpha'\beta'\gamma'}(y) \right] dy + o\left(\frac{1}{n}\right), \quad (5.19)
 \end{aligned}$$

where  $y = (y^1, \dots, y^p)'$ ,  $G = \{g^{\alpha\beta}\}$ , ( $p \times p$ -matrix), and  $\alpha, \beta, \gamma, \delta$  move over  $\{1, \dots, p\}$ . Letting  $y^{q+1} \rightarrow \infty, \dots, y^p \rightarrow \infty$  in (5.19) we obtain (2.9).

*Proof of Theorem 6.* It follows from (3.2) that

$$\begin{aligned}
 \tilde{u}^{*\alpha} &= \sqrt{n} \left( \hat{u}^\alpha - \frac{1}{n} \mu^\alpha(\hat{u}) - u \right) \\
 &= \sqrt{n} (\hat{u}^\alpha - u) - \frac{1}{\sqrt{n}} \mu^\alpha(u) - \frac{1}{\sqrt{n}} (\hat{u}^\gamma - u^\gamma) \partial_\gamma \mu^\alpha + o_p\left(\frac{1}{n}\right). \quad (5.20)
 \end{aligned}$$

By Proposition 2,

$$E(\tilde{u}^{*\alpha}) = o\left(\frac{1}{n}\right). \quad (5.21)$$

From (5.20) we obtain

$$\begin{aligned}
 \text{cum}\{\tilde{u}^{*\alpha}, \tilde{u}^{*\beta}\} &= \text{cum}\{\tilde{u}^\alpha, \tilde{u}^\beta\} - \frac{1}{n} \text{cum}\{\tilde{u}^\alpha, \tilde{u}^\gamma\} \partial_\gamma \mu^\alpha \\
 &\quad - \frac{1}{n} \text{cum}\{\tilde{u}^\beta, \tilde{u}^\gamma\} \partial_\gamma \mu^\beta + o\left(\frac{1}{n}\right), \quad (5.22)
 \end{aligned}$$

$$\text{cum}\{\tilde{u}^{*\alpha}, \tilde{u}^{*\beta}, \tilde{u}^{*\gamma}\} = \frac{1}{\sqrt{n}} c^{\alpha\beta\gamma} + o\left(\frac{1}{n}\right), \quad (5.23)$$

$$\text{cum}\{\tilde{u}^\alpha, \tilde{u}^{*\beta}, \tilde{u}^{*\gamma}, \tilde{u}^{*\delta}\} = \frac{1}{n} c^{\alpha\beta\gamma\delta} + o\left(\frac{1}{n}\right). \quad (5.24)$$

Since  $\hat{u} \in \mathcal{A}_1$ , the matrix  $\mathcal{F}$  in (2.10) is identical to  $\mathcal{F}_0$ . Returning to Theorem 5, and noting (5.21)–(5.24), we can get the result (3.3).

*Proof of Theorem 7.* Note that

$$H_{ab}(x) = -g_{ab} + g_{aa'} g_{bb'} x^{a'} x^{b'},$$

$$\int \dots \int x^e x^f N(x; \mathcal{F}_0) H_{ab}(x) dx = \delta(a, e) \delta(b, f) + \delta(a, f) \delta(b, e).$$

By virtue of (3.3) and the orthogonality of the Hermite polynomials, we can obtain the desired result.

*Proof of Theorem 8.* (i) From the definition (3.4), (3.5), (3.8), and (3.7) it is easy to see that  $A_1^{ab}$ ,  $A_2^{ab}$ , and  $A_5^{ab}$  are independent of the choice of an estimator and that  $A_4^{ab}$  depends on it. To check the assertion for  $A_3^{ab}$ , we use the followings:

$$\partial_a = B_a^j \partial_j \quad \text{and} \quad \partial_a \partial_\kappa = (\partial_a B_\kappa^i) \partial_i + B_\kappa^i B_a^j \partial_j \partial_i, \quad (5.25)$$

where  $a \in \{1, \dots, q\}$ ,  $\kappa \in \{q+1, \dots, p\}$ . By means of (5.25) we have

$$\begin{aligned} M_{a\kappa a' \kappa'} &= (\partial_a B_\kappa^i) (\partial_{a'} B_{\kappa'}^{i'}) g_{ii'} + B_\kappa^i B_a^j (\partial_{a'} B_{\kappa'}^{i'}) J_{i'ij} \\ &\quad + (\partial_a B_\kappa^i) B_{\kappa'}^{i'} B_a^j J_{ii'j'} + B_\kappa^i B_a^j B_{\kappa'}^{i'} B_a^j M_{ijj'}. \end{aligned} \quad (5.26)$$

$$J_{\alpha a \kappa} = B_\alpha^k (\partial_a B_\kappa^i) g_{ki} + B_\alpha^k B_\kappa^i B_a^j J_{kij}, \quad (5.27)$$

where  $a, a' \in \{1, \dots, q\}$ ,  $k, k' \in \{q+1, \dots, p\}$ , and  $\alpha \in \{1, \dots, p\}$ . It follows from (5.26) and (5.27) that

$$A_3^{ab} = g^{aa'} g^{bb'} B_a^j B_{b'}^{j'} g^{ii'} \{M_{ijj'j'} - J_{kij} J_{k'i'j'} g^{kk'}\}, \quad (5.28)$$

which are evidently independent of the choice of an estimator.

(ii) The assertion follows from (5.28).

*Proof of Theorem 9.* Since  $A_1^{ab}$ ,  $A_2^{ab}$ ,  $A_3^{ab}$ , and  $A_5^{ab}$  are independent of the choice of an estimator we investigate  $A_4^{ab}$  for the MLE. Recalling (3.12), we expand

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int \partial_a \log p_n(x_n; \theta(u, 0)) p_n(x_n; \theta(u, v)) dx_n$$

in a Taylor series with respect to  $v$ . Then we get the following equation:

$$0 = v^\kappa g_{a\kappa} + \frac{1}{2} v^\kappa v^\mu C_{a\kappa\mu} + \dots \quad (5.29)$$

Since  $\hat{u} \in \mathcal{A}_1$ , then  $g_{a\kappa} = 0$ . Therefore  $C_{a\kappa\mu} = 0$  for the MLE (i.e.,  $A_4^{ab} = 0$ ). Hence the theorem is proved in view of Theorem 7.

*Proof of Theorem 10.* Put  $A_3^{ab} = 0$  in (3.3) of Theorem 6. We can see that the resulting formula is independent of the ancillary coordinates, and is exactly the Edgeworth expansion for  $\bar{u}^*$ . However, if  $A_3^{ab} \neq 0$ , the result does not hold because  $A_3^{ab}$  depends on the ancillary coordinates.

#### ACKNOWLEDGMENT

The authors are very grateful to the referee for his instructive comments.

#### REFERENCES

- AKAHIRA, M., AND TAKEUCHI, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*, Springer Lecture Notes in Statistics, Vol. 7. Springer-Verlag, New York/Berlin.
- AMARI, S. (1982). Differential geometry of curved exponential families: Curvatures and information loss. *Ann. Statist.* **10** 357–387.
- AMARI, S. (1983). Differential geometry of spectrum estimation. In *Symposium at Osaka University*, preprint.
- AMARI, S. (1985). *Differential Geometrical Methods in Statistics*, Lecture Notes in Statistics, Vol. 28. Springer-Verlag, New York/Berlin.
- AMARI, S. (1987). Differential geometry of a parametric family of invertible linear systems: Riemannian metric, dual Affine connections, and divergence. *Math. Systems Theory* **20** 53–82.
- ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
- BARNDORFF-NIELSEN, O. E., COX, D. R., AND REID, N. (1986). The role of differential geometry in statistical theory. *Internat. Statist. Rev.* **54** 83–96.
- BARNDORFF-NIELSEN, O. E., AND JUPP, P. E. (1989). Approximating exponential models. *Ann. Inst. Statist. Math.* **41** 247–267.
- BHATTACHARYA, R. N., AND GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.
- BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, New York.
- KOTZ, S., AND JOHNSON, N. L. (1987). *Encyclopedia of Statistical Sciences*, Vol. 8. Wiley, New York.
- RAVISHANKER, N., MELNICK, E. L., AND TAI, C. L. (1990). Differential geometry of ARMA models. *J. Time Ser. Anal.* **11** 259–274.
- SUZUKI, T. (1978). Asymptotic sufficiency up to higher orders and its applications to statistical tests and estimates. *Osaka J. Math.* **15** 575–588.
- TANIGUCHI, M. (1986). Third order asymptotic properties of maximum likelihood estimators for Gaussian ARMA processes. *J. Multivariate Anal.* **18** 1–31.
- TANIGUCHI, M. (1987). Validity of Edgeworth expansions of minimum contrast estimators for Gaussian ARMA processes. *J. Multivariate Anal.* **21** 1–28.
- TANIGUCHI, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics, Vol. 68. Springer-Verlag, New York/Berlin.