Ann. Inst. Statist. Math.
Vol. 47, No. 3, 581–600 (1995)

HIGHER ORDER ASYMPTOTIC THEORY FOR NORMALIZING TRANSFORMATIONS OF MAXIMUM LIKELIHOOD ESTIMATORS

MASANOBU TANIGUCHI[1] AND MADAN L. PURI[2*]

[1] *Department of Mathematical Science, Faculty of Engineering Science, Osaka University, Toyonaka 560, Japan*
[2] *Department of Mathematics, Indiana University, Bloomington, IN 47405, U.S.A.*

(Received March 10, 1994; revised January 26, 1995)

Abstract. Suppose that $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ is a collection of m-dimensional random vectors X_i forming a stochastic process with a parameter θ. Let $\hat{\theta}$ be the MLE of θ. We assume that a transformation $A(\hat{\theta})$ of $\hat{\theta}$ has the k-th-order Edgeworth expansion ($k = 2, 3$). If A extinguishes the terms in the Edgeworth expansion up to k-th-order ($k \geq 2$), then we say that A is the k-th-order normalizing transformation. In this paper, we elucidate the k-th-order asymptotics of the normalizing transformations. Some conditions for A to be the k-th-order normalizing transformation will be given. Our results are very general, and can be applied to the i.i.d. case, multivariate analysis and time series analysis. Finally, we also study the k-th-order asymptotics of a modified signed log likelihood ratio in terms of the Edgeworth approximation.

Key words and phrases: Normalizing transformation, higher-order asymptotic theory, variance stabilizing transformation, multivariate analysis, time series analysis, Edgeworth expansion, saddlepoint expansion, MLE, observed information, signed log likelihood ratio.

1. Introduction

In the area of multivariate analysis, there have been a lot of suggestions on how to transform statistics in order to get some desirable properties. An important example of transformations is Fisher's z-transformation $Z(r)$ for the sample correlation coefficient r in a bivariate normal sample. Hotelling (1953) evaluated the higher order asymptotic moments of $Z(r)$, and showed that $Z(r)$ becomes the asymptotic variance stabilizing transformation. Furthermore, he gave a transformation of r which extinguishes the second-order bias. In terms of the Edgeworth approximation, Konishi (1978) showed that $Z(r)$ extinguishes a part of the second-order terms of the asymptotic expansion. Also, Konishi (1981) discussed the transformations of a statistic based upon the elements of the sample covariance

* Research supported by the Office of Naval Research Contract N00014-91-J-1020.

matrix which extinguish the second-order terms of the Edgeworth expansions. Furthermore, Fang and Krishnaiah (1982) gave the Edgeworth expansions of certain functions of the elements of the noncentral Wishart matrix. They obtained analogous results for functions of the elements of the sample covariance matrix when the underlying distribution is a mixture of multivariate normal distributions.

In the area of time series analysis the first study of higher order asymptotic properties of a transformed statistic is Phillips (1979). He gave the Edgeworth expansion of a transformation of the least squares estimator for the coefficient of an AR(1) process, and showed that Fisher's z-transformation extinguishes a part of the second-order terms of the Edgeworth expansion. Taniguchi *et al.* (1989) investigated Edgeworth type expansions of certain transformations of some statistics of Gaussian ARMA processes. They also gave transformations which make the second-order part of the Edgeworth expansions vanish.

Most of the discussions in the above depend on the individual models. There are few literatures which develop the higher order asymptotic theory for transformations of statistics (c.f. Konishi (1987, 1991)). In this paper, we shall develop the higher order asymptotic theory for transformations of the MLE in general statistical models. Our results are applicable to the i.i.d. case, time series analysis and the non-identically distributed case. Let $\hat{\theta}$ be the MLE of θ in a general statistical model, and let $A(\hat{\theta})$ be a transformation of $\hat{\theta}$. Suppose that a standardized version $\tilde{A}$ of $A(\hat{\theta})$ has the following Edgeworth expansion with respect to the sample size n;

$$P(\tilde{A} < x) = \Phi(x) + n^{-1/2}C_A^{(2)}(x) + n^{-1}C_A^{(3)}(x) + o(n^{-1}), \tag{1.1}$$

where $\Phi(x)$ is the distribution function of the standard normal distribution, and $C_A^{(2)}(x)$ and $C_A^{(3)}(x)$ are the second and third order terms of the Edgeworth expansion. Throughout this paper we say that A is the second-order normalizing transformation if $C_A^{(2)}(x) \equiv 0$, and that A is the third-order normalizing transformation if $C_A^{(2)}(x) \equiv C_A^{(3)}(x) \equiv 0$.

In Section 2, we shall elucidate the second-order asymptotics of the second-order normalizing transformations in our general setting. A relation between the variance stabilizing transformation and the second-order normalizing transformation will be discussed. In Section 3, we will show that the second-order normalizing transformation does not become the third-order one generally. Then some conditions for A to be the third-order normalizing transformation are given. However, they are found to be very restrictive. Thus, in Section 4, without these restrictive conditions we propose another type of third-order normalizing transformation in terms of the observed information, which supplies some additional informations which the MLE $\hat{\theta}$ can not recover.

Barndorff-Nielsen (1991) and Jensen (1992) discussed a modified signed log likelihood ratio, which becomes a kind of normalizing transformation in the sense of the saddlepoint approximation. In Section 5, we elucidate the third-order asymptotics of their modified signed log likelihood ratio by using the Edgeworth approximation.

2. Second-order asymptotic theory for normalizing transformations

In this section, we view the normalizing transformation of the MLE from the second-order asymptotic theory.

Let $\boldsymbol{X}_n = (X_1, X_2, \dots, X_n)$ 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 $p_n(\boldsymbol{x}_n;\theta)$ denote the probability density function of $\boldsymbol{X}_n$ with respect to a carrier measure, where $\boldsymbol{x}_n \in \boldsymbol{R}^{mn}$ and $\theta = (\theta^1, \dots, \theta^p)' \in \Theta \subset \boldsymbol{R}^p$, is an unknown parameter vector. Henceforth it is assumed that $p_n(\cdot;\theta)$ is differentiable with respect to θ up to necessary order. Define

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

where $i,j,k = 1,\dots,p$, $\ell_n(\theta) = \log p_n(\boldsymbol{X}_n;\theta)$ and $\partial/\partial\theta^i$ is abbreviated to ∂_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 (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:

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

and J-th-order ($J \geq 3$) cumulants of Z_i, Z_{ij} and Z_{ijk} are all $O(n^{-J/2+1})$. If θ is scalar we use g, J, $K, \dots, Z_1, Z_2, Z_3$ for the quantities g_{ij}, J_{ijk}, $K_{ijk}, \dots$, Z_i, Z_{ij}, Z_{ijk}, respectively.

Now we are interested in the estimation of a parameter $S = S(\theta) : \boldsymbol{R}^p \to \boldsymbol{R}^1$, which is continuously twice differentiable with respect to $\theta \in \Theta$. We estimate $S(\theta)$ by $S(\hat{\theta})$ where $\hat{\theta}$ is the MLE of θ.

Example 1. Suppose that X_i; $i = 1,\dots,n$, are i.i.d. as

$$N_2\left(\mu, \begin{pmatrix} \theta^1 & \theta^3 \\ \theta^3 & \theta^2 \end{pmatrix}\right). \tag{2.1}$$

If we set $S(\theta) = \theta^3/\{\theta^1\theta^2\}^{1/2}$ then $S(\theta)$ and $S(\hat{\theta})$ are the correlation coefficient and the sample correlation coefficient, respectively.

Next, we consider a transformation of $S(\hat{\theta})$. Let $A : \boldsymbol{R}^1 \to \boldsymbol{R}^1$ be a continuously three times differentiable function. Define $S_i = \partial_i S(\theta)$, $S_{jk} = \partial_j \partial_k S(\theta)$ and $\mathcal{S} = S_i g^{ij} S_j$, where g^{ij} is the (i,j) component of the inverse matrix of $\{g_{ij}\}$. Here, we adopt the Einstein summation convention. We denote the i-th order derivative of A by $A^{(i)}$. To seek the second-order normalizing transformation A of $S(\hat{\theta})$ we make the standardization

$$\tilde{a} = \sqrt{n}\{A^{(1)}\sqrt{\mathcal{S}}\}^{-1}\left[A\{S(\hat{\theta})\} - A\{S(\theta)\} - \frac{c}{n}\right], \tag{2.2}$$

where c is a correction constant, and $A^{(1)}$ and $\mathcal{S}$ are taken at $\hat{\theta} = \theta$. To avoid many sophisticated regularity conditions which depend on the model concerned, throughout this paper we assume the validity of the Edgeworth expansions (see Bhattacharya and Ghosh (1978) for the i.i.d. case and Taniguchi (1987) for the dependent case). The following theorem describes the second-order normalizing transformation. We put the proofs of theorems and propositions in the Appendix, if they are not straightforward.

THEOREM 2.1. *If A and c satisfy the differential equations*

$$\begin{cases} \dfrac{A^{(2)}}{A^{(1)}} = -\dfrac{1}{\mathcal{S}^2} S_i S_j S_{km} g^{ik} g^{jm} \\ \qquad\qquad + \dfrac{1}{3\mathcal{S}^2} S_i S_j S_k g^{ii'} g^{jj'} g^{kk'} (2K_{i'j'k'} + 3J_{i'j'k'}) \\ \dfrac{c}{A^{(1)}} = S_i \left\{ g^{ij} g^{km} J_{mjk} + \dfrac{1}{2} g^{ij} g^{km} R_{jkm} \right\} + \dfrac{1}{2}\left(\dfrac{A^{(2)}}{A^{(1)}} S_i S_j + S_{ij} \right) g^{ij}, \end{cases} \tag{2.3}$$

then A is the second-order normalizing transformation, i.e.,

$$P(\tilde{a} < x) = \Phi(x) + o(n^{-1/2}),$$

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

If the asymptotic variance $\tilde{v}$ of $A\{S(\hat{\theta})\}$ is independent of θ we say that A is an asymptotic variance stabilizing transformation.

PROPOSITION 2.1. *If A satisfies the differential equations*

$$\frac{A^{(2)}}{A^{(1)}} = (S_m \mathcal{S})^{-1}\{S_i S_j g^{ii'} g^{jj'} (K_{i'j'm} + J_{i'j'm} + J_{j'i'm})/2 - S_{mi} S_j g^{ij}\}, \quad m = 1, \ldots, p, \tag{2.4}$$

then A is the asymptotic variance stabilizing transformation.

Also we obtain the following proposition.

PROPOSITION 2.2. *If*

$$S_i S_j S_k g^{ii'} g^{jj'} g^{kk'} K_{i'j'k'} = 0 \tag{2.5}$$

is satisfied, then the variance stabilizing transformation of $S(\hat{\theta})$ *given by* (2.4) *is equivalent to the second-order normalizing transformation of* $S(\hat{\theta})$.

Konishi (1987) discussed the asymptotic theory for transformations of a class of estimators in the i.i.d. case. He gave a sufficient condition that the second-order normalization and variance stabilization are simultaneously achieved. In the case where $\{X_i\}$ are i.i.d. and $\hat{\theta}$ is the MLE, the condition (2.5) is essentially equivalent to Konishi's one.

Remark 1. In general, we can not guarantee that the differential equations (2.3) and (2.4) are always solvable with respect to S. If $p = 1$ (i.e., θ is scalar), making the transformation $\theta \to S$, we can easily show that (2.3) and (2.4) become

$$\begin{cases} \dfrac{A^{(2)}}{A^{(1)}} = \{2K(S) + 3J(S)\}/3g(S), \\ \dfrac{c}{A^{(1)}} = -K(S)/6g(S)^2, \end{cases} \tag{2.3$'$}$$

$$\frac{A^{(2)}}{A^{(1)}} = \frac{K(S) + 2J(S)}{2g(S)}, \tag{2.4$'$}$$

respectively, where $g(S)$, $J(S)$ and $K(S)$ are the corresponding quantities g, J, K obtained by replacing the derivatives with respect to θ with $\partial = \partial/\partial S$. Of course (2.3)$'$ and (2.4)$'$ are solvable with respect to S. Because most of the discussions on this topic are on the wider family $\boldsymbol{S} = \{p_n(\cdot;\theta); \dim\theta > 1\}$ not on the "curved family" $\boldsymbol{M} = \{p_n(\cdot; S); \dim S = 1\}$, we develop our theory on $\boldsymbol{S}$ with respect to θ. We give some examples below.

Example 2. In Example 1 the variance stabilizing transformation of the sample correlation coefficient $r = S(\hat{\theta}) = \hat{\theta}^3/\{\hat{\theta}^1\hat{\theta}^2\}^{1/2}$ is given by Fisher's z-transformation

$$A(r) = \frac{1}{2}\log\{(1+r)/(1-r)\}. \tag{2.6}$$

It is not difficult to show that (2.5) is satisfied, whence $A(r)$ is the second-order normalizing transformation.

Remark 2. Suppose that X_i; $i = 1, \dots, n$, are i.i.d. as

$$N_2\left(\mu, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Then the first-order asymptotics of the MLE $\hat{\rho}$ of ρ is different from that of r in Example 2. This is due to the fact that the family of distributions

$$\boldsymbol{M} = \left\{ N_2 \left(\mu, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \right\}$$

is a curved subfamily of

$$\boldsymbol{S} = \left\{ N_2 \left(\mu, \begin{pmatrix} \theta^1 & \theta^3 \\ \theta^3 & \theta^2 \end{pmatrix} \right) \right\},$$

and ρ is not orthogonal to θ^1 and θ^2 with respect to the Fisher information. It should be noted that most of the discussions on the correlation coefficient are on the wider family $\boldsymbol{S}$ not on $\boldsymbol{M}$. Example 2 discussed $A\{S(\hat{\theta})\}$ not $A\{\hat{\rho}\}$.

Eigenvalue is one of the most important indices in multivariate analysis. The following is an example of the second-order normalizing transformation of the eigenvalues of the sample covariance matrix.

Example 3. Suppose that X_i; $i = 1, \ldots, n$, are i.i.d. as $N_p(\mu, \Sigma)$. Let $\lambda_1 > \cdots > \lambda_p$ be the eigenvalues of Σ, and let $\ell_1 \geq \cdots \geq \ell_p$ be the eigenvalues of S, where

$$S = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})', \quad \text{with} \quad \bar{X} = n^{-1} \sum_{i=1}^{n} X_i.$$

We can see that the second-order normalizing transformation of the α-th eigenvalue ℓ_α is given by $A(x) \propto x^{1/3}$ (see also Krishnaiah and Lee (1979), Konishi (1981), Fang and Krishnaiah (1982)).

Amari (1985) developed a differential geometry of statistical inference for a curved exponential family. His approach gave us a unified view of higher order asymptotic theory. Here we grasp the second-order normalizing transformation from his point of view. For $\tilde{\theta}^i = \sqrt{n}(\hat{\theta}^i - \theta^i)$, $i = 1, \ldots, p$, $\mathrm{cum}\{\tilde{\theta}^i, \tilde{\theta}^j, \tilde{\theta}^k\}$ are expressed as a linear combination of the third-order tensor K_{ijk} and 1-connection J_{ijk} (see (A.4) in Appendix). By transformation $\theta \to \xi = (\xi_\alpha)$, the third-order tensor K_{ijk} changes to

$$\bar{B}^i_\alpha \bar{B}^j_\beta \bar{B}^k_\gamma K_{ijk}, \tag{2.7}$$

where $\bar{B}^i_\alpha = \partial \theta^i / \partial \xi_\alpha$. On the other hand, by this transformation the 1-connection J_{ijk} changes to

$$\bar{B}^i_\alpha \bar{B}^j_\beta (\partial_j \bar{B}^k_\gamma) g_{ki} + \bar{B}^i_\alpha \bar{B}^j_\beta \bar{B}^k_\gamma J_{ijk}. \tag{2.8}$$

Therefore, even if the third-order cumulants of $\tilde{\theta}^j$ do not vanish in the original model we can define a transformation which extinguishes the transformed third-order cumulants in terms of some differential equations. In Theorem 2.1, (2.3) elucidate a relation between the normalizing transformations and the differential geometrical structure of the wider family $\boldsymbol{S}$.

3. Third-order asymptotic theory for normalizing transformation

In this section we shall discuss the third-order asymptotics of normalizing transformations. We restrict ourselves to the situation where θ is scalar and $S(\theta) = \theta$. That is, the third-order asymptotics of the transformation $A(\hat{\theta})$ of the MLE $\hat{\theta}$ will be elucidated. First, we make the following standardization

$$V_n = \sqrt{ng}\left\{A(\hat{\theta}) - A(\theta) - \frac{c}{n}\right\}/A^{(1)}(\theta). \tag{3.1}$$

Define

$$\begin{aligned}
C_1^{(1)} &= \frac{-J-K}{2g^{3/2}} + \frac{A^{(2)}}{2A^{(1)}g^{1/2}} - \frac{cg^{1/2}}{A^{(1)}},\\
C_{11}^{(3)} &= -\frac{\Delta}{g} + \frac{-L-4N-H}{g^2} + \frac{7J^2+14JK+5K^2}{2g^3}\\
&\quad + \frac{A^{(3)}}{A^{(1)}g} + \frac{(A^{(2)})^2}{2(A^{(1)})^2 g} + \frac{-4JA^{(2)}-3KA^{(2)}}{A^{(1)}g^2},\\
C_{111}^{(1)} &= \frac{-3J-2K}{g^{3/2}} + \frac{3A^{(2)}}{A^{(1)}g^{1/2}},\\
C_{1111}^{(1)} &= -\frac{4L+12N+3H}{g^2} + \frac{12(2J+K)(J+K)}{g^3}\\
&\quad + \frac{4A^{(3)}}{A^{(1)}g} + \frac{12(A^{(2)})^2}{g(A^{(1)})^2} - \frac{12A^{(2)}(3J+2K)}{A^{(1)}g^2}.
\end{aligned}$$

Then we have,

PROPOSITION 3.1. (i) *The transformation A is the second-order normalizing transformation if and only if $C_1^{(1)} = C_{111}^{(1)} = 0$.*

(ii) *The transformation A is the third-order normalizing transformation if and only if $C_1^{(1)} = C_{111}^{(1)} = C_{11}^{(3)} = C_{1111}^{(1)} = 0$.*

Remark 3. We can see that $C_{111}^{(1)} = 0$ is equivalent to

$$\frac{A^{(2)}}{A^{(1)}} = \frac{3J+2K}{3g}. \tag{3.2}$$

The second-order normalizing transformation is given by solving the differential equation (3.2). Differentiation of (3.2) with respect to θ yields

$$\frac{A^{(3)}}{A^{(1)}} = \frac{3M+3L+9N+2H}{3g} + \frac{(3J+2K)(-3J-K)}{9g^2}. \tag{3.3}$$

Substituting (3.2) and (3.3) into $C_{1111}^{(1)}$ we obtain

$$C_{1111}^{(1)} = \frac{12(Mg-J^2)}{3g^3} + \frac{4K^2-3Hg}{9g^3}, \tag{3.4}$$

for the second-order normalizing transformation A.

We evaluate (3.4) for the following time series models. The fundamental quantities g, J, K, M and H are given in Taniguchi ((1991), p. 39).

Example 4. (i) Let $\{X_t\}$ be a Gaussian $AR(1)$ process with the spectral density $f_\theta(\lambda) = \frac{\sigma^2}{2\pi}\frac{1}{|1-\theta e^{i\lambda}|^2}$, $|\theta| < 1$. Then, for the second-order normalizing transformation A of $\hat{\theta}$,

$$C_{1111}^{(1)} = \frac{2-6\theta^2}{1-\theta^2}, \tag{3.5}$$

which is not equal to 0 if $\theta^2 \neq \frac{1}{3}$.

(ii) Let $\{X_t\}$ be a Gaussian $MA(1)$ process with the spectral density $f_\theta(\lambda) = \frac{\sigma^2}{2\pi}|1-\theta e^{i\lambda}|^2$, $|\theta| < 1$. Then, for the second-order normalizing transformation A of $\hat{\theta}$,

$$C_{1111}^{(1)} = \frac{6(3-\theta^2)}{1-\theta^2}, \tag{3.6}$$

which is not equal to 0 for all $|\theta| < 1$.

(iii) Let $\{X_t\}$ be a Gaussian $ARMA(p,q)$ process with the spectral density

$$f_\theta(\lambda) = \frac{\theta}{2\pi}\frac{|\sum_{j=0}^{q}\alpha_j e^{ij\lambda}|^2}{|\sum_{j=0}^{p}\beta_j e^{ij\lambda}|^2}, \qquad (\alpha_0 = \beta_0 = 1).$$

Then, for the second-order normalizing transformation A of $\hat{\theta}$,

$$C_{1111}^{(1)} = -\frac{4}{9}.$$

Summarizing the discussion above, we have

THEOREM 3.1. *The second-order normalizing transformation does not become the third-order normalizing transformation in general.*

The next interest is the following question: for what statistical models does the second-order normalizing transformation imply the third-order normalizing transformation? The following theorem answers this.

THEOREM 3.2. (i) *If A is the variance stabilizing transformation, and if $K = 0$, then A is the second-order normalizing transformation.*

(ii) *In addition to the condition* (i) *if $H = 0$, $Mg - J^2 = 0$ and $\Delta = 0$ hold, then A is the third-order normalizing transformation.*

Here it may be noted that $\tilde{M}^{1/2}/g$, with $\tilde{M} = Mg - J^2$, is a counterpart of Efron's statistical curvature in our situation (e.g., Efron(1975)).

4. Third-order normalizing transformation in terms of the observed information

As we saw in the previous section, the condition for A to be the third-order normalizing transformation is fairly restrictive. In this section, without restrictive assumptions, we will make a third-order normalizing transformation. It is well known that the observed information

$$\hat{I} = -\frac{1}{n}\partial^2 \ell_n(\hat{\theta}),$$

can supply some additional information which the MLE $\hat{\theta}$ can not recover. From this motivation, we consider the standardized transformation

$$\tilde{e} = \sqrt{ng}\{A^{(1)}(\theta)\}^{-1}\left\{A(\hat{\theta}) - A(\theta) - \frac{c}{n}\right\}\left\{h(\hat{I}) + \frac{d}{n}\right\}, \tag{4.1}$$

where c and d are correction constants, and $h(\cdot)$ is a continuously three times differentiable function satisfying $h(g) = 1$. We also denote the i-th order derivative of $h(\cdot)$ by $h^{(i)}(\cdot)$. The following theorem describes the third-order normalizing transformation in the form (4.1).

THEOREM 4.1. *Suppose that A and h satisfy the differential equations*

$$\begin{aligned} &-2\frac{A^{(3)}}{A^{(1)}} + \left(\frac{A^{(2)}}{A^{(1)}}\right)^2 \frac{3\tilde{M}}{(2J+K)^2} + \left(\frac{A^{(2)}}{A^{(1)}}\right)\left\{\frac{2(3J+K)\tilde{M}}{g(2J+K)^2} + \frac{K}{g}\right\} \\ &+ \frac{2L+6M+6N+H}{g} - \frac{(3J+2K)(9J+4K)\tilde{M}}{3g^2(2J+K)^2} \\ &- \frac{18J^2+9JK+2K^2}{3g^2} = 0, \end{aligned} \tag{4.2}$$

$$h^{(1)}(g) = \frac{3J+2K}{6g(2J+K)} - \frac{A^{(2)}}{2A^{(1)}(2J+K)}, \tag{4.3}$$

and that c and d are chosen so that $C_1^{(1)} = 0$ and $C_{11}^{(3)} = 0$ of (A.31) *and* (A.32) *in the Appendix, respectively. Then $\tilde{e}$ becomes the third-order normalizing transformation, i.e.,*

$$P\{\tilde{e} < x\} = \Phi(x) + o(n^{-1}). \tag{4.4}$$

Here it should be noted that if the curvature $\tilde{M}$ vanishes, the differential equation (4.2) is reduced to a very simple form. We next give some examples for Theorem 4.1.

Example 4. Let $\{X_t\}$ be a Gaussian $AR(1)$ process with the spectral density $f_\theta(\lambda) = \frac{\sigma^2}{2\pi}|1-\theta e^{i\lambda}|^{-2}$. Then (4.2) becomes the following Riccati type differential equation

$$\frac{dy}{d\theta} + y^2\frac{7\theta^2-3}{4\theta^2} - \frac{3\theta}{1-\theta^2}y = 0, \tag{4.5}$$

where $y = A^{(2)}(\theta)/A^{(1)}(\theta)$. From (4.5) and (4.3) we can get

$$A^{(1)}(\theta) = \sqrt{\frac{\theta^2+3}{1-\theta^2}}, \tag{4.6}$$

$$h^{(1)}(g) = \frac{1-\theta^4}{2(\theta^2+3)}. \tag{4.7}$$

Also, from (A.31) and (A.32) we obtain

$$c = -\theta\sqrt{\frac{\theta^2+3}{1-\theta^2}}, \qquad d = -\frac{2\theta^2(17+10\theta^2+\theta^4)}{(1-\theta^2)(\theta^2+3)^2}.$$

Since we can approximate $\tilde{e}$ up to third-order by the Taylor expansion in terms of the derivatives of $A(\theta)$ and $h(g)$, it is enough for us to give the explicit form of $A^{(1)}(\theta)$ and $h^{(1)}(g)$ (in fact we can not give their integrals explicitly).

Example 5. Let $\{X_t\}$ be a Gaussian $ARMA(1,1)$ process with the spectral density

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

It follows from (4.2) and (4.3) that

$$A(\theta) = \theta^{(6+2\sqrt{6})/3},$$
$$h(x) = \frac{\sqrt{x}}{3\sqrt{g}} + \frac{2}{3} + \left(\frac{\sqrt{2}}{8} + \frac{\sqrt{3}}{6}\right)(g^{-3/2} - x^{-1}g^{-1/2}).$$

Here we give a comment concerning confidence intervals of θ. If $\sqrt{g}/A^{(1)}(\theta)$ and c in (3.1) are independent of θ, or if $\sqrt{g}/A^{(1)}(\theta)$, c and d in (4.1) are independent of θ, then we can easily make confidence intervals of $A(\theta)$ based on the results in Sections 3 and 4. However, if not so it seems to be difficult to construct them because we can not guarantee that $A(\cdot)$ is monotone.

5. A modified signed log likelihood ratio and another normalizing method

Barndorff-Nielsen (1991) discussed a modified signed log likelihood ratio, which becomes a kind of normalizing transformation. For a family of exponential distributions Jensen (1992) investigated some asymptotic properties of

$$r^* = r - \frac{1}{r}\log\frac{r}{u}, \tag{5.1}$$

where

$$r = \operatorname{sgn}(\hat{\theta}-\theta)\{2\ell_n(\hat{\theta}) - 2\ell_n(\theta)\}^{1/2},$$
$$u = \sqrt{n}(\hat{\theta}-\theta)\hat{I}^{1/2}.$$

Here r is called the signed log likelihood ratio. Using a saddlepoint approximation, he showed

$$P(r^* \geq x) = 1 - \Phi(x) + o(n^{-1}). \tag{5.2}$$

In this section, we will give the third-order Edgeworth expansion for r^*. Then we check whether r^* is the third-order normalizing transformation or not in the sense of the Edgeworth approximation. First, we give the stochastic expansion of r^*.

PROPOSITION 5.1. *Under our general assumptions in Section 2 (i.e., we are not restricted to the family of exponential distributions), r^* has the stochastic expansion.*

$$\begin{aligned} r^* = {} & Z_1 \left\{ g + \frac{\Delta}{n} \right\}^{-1/2} + n^{-1/2} \left\{ \frac{Z_1 Z_2}{2g^{3/2}} + \frac{(3J+K)(g - Z_1^2)}{6g^{5/2}} \right\} \\ & + n^{-1} \left[\frac{3}{8g^{5/2}} Z_1 Z_2^2 - \frac{5(3J+K)}{12g^{7/2}} Z_1^2 Z_2 + \frac{1}{6g^{5/2}} Z_1^2 Z_3 \right. \\ & + \frac{(3J+K)^2}{9g^{9/2}} Z_1^3 - \frac{4L+3M+6N+H}{24g^{7/2}} Z_1^3 \\ & + \left\{ \frac{4L+3M+6N+H}{8g^{5/2}} - \frac{7(3J+K)^2}{36g^{7/2}} \right\} Z_1 \\ & \left. + \frac{3J+K}{4g^{5/2}} Z_2 - \frac{Z_3}{6g^{3/2}} \right] + o_P(n^{-1}). \end{aligned} \tag{5.3}$$

If $p_n(\boldsymbol{x}_n;\theta)$ belongs to the family $\mathcal{F}_e$ of exponential distributions, we get the following proposition.

PROPOSITION 5.2. *If $p_n(\boldsymbol{x}_n;\theta) \in \mathcal{F}_e$ and if θ is the natural parameter, then r^* becomes the third-order normalizing transformation in the sense of the Edgeworth approximation, i.e.,*

$$P(r^* \leq x) = \Phi(x) + o(n^{-1}).$$

However, if we are not restricted to the family of exponential distributions we obtain the following.

PROPOSITION 5.3. *Under our general assumptions in Section 2 (i.e., $p_n(\boldsymbol{x}_n;\theta) \notin \mathcal{F}_e$), r^* does not generally become the third-order normalizing transformation in the sense of the Edgeworth approximation.*

Pázman(1990) discussed another normalizing method in the case of a nonlinear regression:

$$\begin{cases} \boldsymbol{X}_n = \boldsymbol{m}(\theta) + \boldsymbol{e}, \\ \boldsymbol{e} \sim N(\boldsymbol{0}, \Sigma), \\ \theta \in \Theta \subset \boldsymbol{R}^p, \end{cases} \tag{5.4}$$

where the systematic part $\boldsymbol{m}(\theta) : \Theta \to \boldsymbol{R}^n$ is three times continuously differentiable and Σ is a known nonsingular variance matrix ($n \times n$ matrix). Then it is proved that, if the Fisher information $g(\theta)$ is constant,

$$q^* = [\boldsymbol{m}(\hat{\theta}) - \boldsymbol{m}(\theta)]'\Sigma^{-1}\frac{\partial}{\partial\theta}\boldsymbol{m}(\hat{\theta}), \tag{5.5}$$

is "almost exactly" normal $N(0, g_n(\theta)/n)$, where $\hat{\theta}$ is the MLE of θ, and $g_n(\theta) = \{\frac{\partial}{\partial\theta}\boldsymbol{m}(\theta)\}'\Sigma^{-1}\{\frac{\partial}{\partial\theta}\boldsymbol{m}(\theta)\}$ (for detail, see Pázman (1990)). In view of the Edgeworth approximation we have,

PROPOSITION 5.4. *Assume that $g_n(\theta)$ is independent of θ and that Assumption 1 holds. Then q^* becomes the third-order normalizing transformation in the sense of the Edgeworth expansion.*

Acknowledgements

The first author is grateful to Professors Barndorff-Nielsen and J. L. Jensen for their discussion at the University of Aarhus. Thanks are also due to the two referees for their instructive comments.

Appendix

In this section, we give the proofs of some of the theorems and propositions.

PROOF OF THEOREM 2.1. The stochastic expansion of $a = \sqrt{n}[A\{S(\hat{\theta})\} - A\{S(\theta)\}]$ is given by

$$\begin{aligned} a = {} & A^{(1)}S_i\sqrt{n}(\hat{\theta}^i - \theta^i) \\ & + \frac{1}{2\sqrt{n}}(A^{(2)}S_iS_m + A^{(1)}S_{im})\sqrt{n}(\hat{\theta}^i - \theta^i)\sqrt{n}(\hat{\theta}^m - \theta^m) \\ & + o_P(n^{-1/2}). \end{aligned} \tag{A.1}$$

It is not difficult to show that

$$E\{\sqrt{n}(\hat{\theta}^i - \theta^i)\} = \frac{1}{\sqrt{n}}\left[g^{ij}g^{km}J_{mjk} + \frac{1}{2}g^{ij}g^{km}R_{jkm}\right] + o(n^{-1/2}), \tag{A.2}$$

$$\mathrm{cum}\{\sqrt{n}(\hat{\theta}^i - \theta^i), \sqrt{n}(\hat{\theta}^j - \theta^j)\} = g^{ij} + o(n^{-1/2}). \tag{A.3}$$

$$\begin{aligned} & \mathrm{cum}\{\sqrt{n}(\hat{\theta}^i - \theta^i), \sqrt{n}(\hat{\theta}^j - \theta^j), \sqrt{n}(\hat{\theta}^k - \theta^k)\} \\ & = -\frac{1}{\sqrt{n}}g^{ii'}g^{jj'}g^{kk'}\{2K_{i'j'k'} + J_{j'i'k'} + J_{i'j'k'} + J_{k'i'j'}\} \\ & \quad + o(n^{-1/2}). \end{aligned} \tag{A.4}$$

and

$$\mathrm{cum}\{\sqrt{n}(\hat{\theta}^{i_1} - \theta^{i_1}), \ldots, \sqrt{n}(\hat{\theta}^{i_J} - \theta^{i_J})\} = O(n^{-J/2+1}), \qquad (J \geq 3) \tag{A.5}$$

(c.f. Amari (1985) for the i.i.d. curved exponential family, Takeuchi and Morimune (1985) for the curved linear model, Taniguchi (1991) for the time series model). From (A.1)–(A.5) it follows that

$$
\begin{aligned}
\text{(A.6)}\quad E\{\tilde{a}\} &= \frac{1}{\sqrt{n}} s^{-1/2}\Big[S_i\Big\{g^{ij}g^{km}J_{mjk} + \frac{1}{2}g^{ij}g^{km}R_{jkm}\Big\} \\
&\qquad + \frac{1}{2}\left(\frac{A^{(2)}}{A^{(1)}}S_iS_j + S_{ij}\right)g^{ij} - \frac{C}{A^{(1)}}\Big] + o(n^{-1/2}) \\
&= \frac{1}{\sqrt{n}}C_1 + o(n^{-1/2}), \qquad \text{(say)},
\end{aligned}
$$

$$\text{(A.7)}\quad \mathrm{cum}\{\tilde{a},\tilde{a}\} = 1 + o(n^{-1/2}),$$

$$
\begin{aligned}
\text{(A.8)}\quad \mathrm{cum}\{\tilde{a},\tilde{a},\tilde{a}\} &= \frac{\sqrt{s}}{\sqrt{n}}\Big[\frac{3A^{(2)}}{A^{(1)}} + \frac{3}{\mathcal{S}^2}S_iS_jS_{km}g^{ik}g^{jm} \\
&\qquad - \frac{1}{\mathcal{S}^2}S_iS_jS_kg^{ii'}g^{jj'}g^{kk'}(2K_{i'j'k'} + 3J_{i'j'k'})\Big] \\
&\quad + o(n^{-1/2}), \\
&= \frac{1}{\sqrt{n}}C_3 + o(n^{-1/2}), \qquad \text{(say)},
\end{aligned}
$$

and the J-th-order ($J \geq 3$) cumulants satisfy

$$\text{(A.9)}\qquad \mathrm{cum}^{(J)}\{\tilde{a},\dots,\tilde{a}\} = O(n^{-J/2+1}).$$

Applying a general formula (e.g., Taniguchi ((1991), p. 15)) to $\tilde{a}$ we obtain

$$\text{(A.10)}\qquad P(\tilde{a} < x) = \Phi(x) - \frac{1}{\sqrt{n}}\phi(x)\left[C_1 + \frac{C_3}{6}(x^2-1)\right] + o(n^{-1/2}),$$

where $\phi(x) = \Phi'(x)$. If we set $C_1 = C_3 = 0$, we get (2.3).

PROOF OF PROPOSITION 2.1. Since $\tilde{v} = A^{(1)}S_ig^{ij}S_jA^{(1)}$, we can see that the relation $\partial_m\tilde{v} = 0$, $(m = 1,\dots,p)$ yields,

$$0 = 2A^{(1)}A^{(2)}S_mS_ig^{ij}S_j + 2A^{(1)^2}\{S_{mi}g^{ij}S_j\} + A^{(1)^2}S_i(\partial_mg^{ij})S_j, \qquad j = 1,\dots,p,$$

which lead to (2.4) by noting the relation

$$\partial_mg^{ij} = -g^{ii'}g^{jj'}(K_{i'j'm} + J_{i'j'm} + J_{j'i'm}).$$

PROOF OF PROPOSITION 2.2. From (2.4) we have

$$
\begin{aligned}
\text{(A.11)}\qquad (S_m\mathcal{S})A^{(2)}/A^{(1)} &= S_iS_jg^{ii'}g^{jj'}(K_{i'j'm} + J_{i'j'm} + J_{j'i'm})/2 \\
&\quad - S_{mi}S_jg^{ij}, \qquad m = 1,\dots,p.
\end{aligned}
$$

Multiplying (A.11) by $g^{mk}S_k$, we obtain

$$\text{(A.11)}' \quad \mathcal{S}^2 A^{(2)}/A^{(1)} = S_i S_j S_k g^{ii'} g^{jj'} g^{mk}(K_{i'j'm} + J_{i'j'm} + J_{j'i'm})/2 - S_j S_k S_{mi} g^{ij} g^{mk}.$$

If (2.5) is satisfied we can see that (A.11)$'$ and the above differential equation of (2.3) are identical.

PROOF OF PROPOSITION 3.1. Since the statement of (i) was proved in Section 2, we will prove (ii). The stochastic expansion of V_n is given by

$$V_n = \sqrt{ng}(\hat{\theta} - \theta) + \frac{A^{(2)}\sqrt{g}}{2\sqrt{n}A^{(1)}}\{\sqrt{n}(\hat{\theta} - \theta)\}^2 - \frac{\sqrt{g}C}{\sqrt{n}A^{(1)}} + \frac{A^{(3)}\sqrt{g}}{6nA^{(1)}}\{\sqrt{n}(\hat{\theta} - \theta)\}^3 + o_P(n^{-1}).$$

Putting $U_n = \sqrt{n}(\hat{\theta} - \theta)$ we can show that

$$\text{(A.12)} \quad E\{U_n\} = -\frac{J + K}{2\sqrt{n}g^2} + o(n^{-1}),$$

$$\text{(A.13)} \quad \mathrm{Var}\{U_n\} = g^{-1} - \frac{\Delta}{g^2 n} + \frac{7J^2 + 14JK + 5K^2}{2g^4 n} - \frac{L + 4N + H}{g^3 n} + o(n^{-1}),$$

$$\text{(A.14)} \quad \mathrm{cum}\{U_n, U_n, U_n\} = -\frac{3J + 2K}{g^3\sqrt{n}} + o(n^{-1}),$$

$$\text{(A.15)} \quad \mathrm{cum}\{U_n, U_n, U_n, U_n\} = \frac{12(2J + K)(J + K)}{g^5 n} - \frac{4L + 12N + 3H}{g^4 n} + o(n^{-1}),$$

$$\text{(A.16)} \quad \mathrm{cum}^{(r)}\{U_n, \ldots, U_n\} = O(n^{-r/2+1}), \quad \text{for} \quad r \geq 5$$

(c.f. Amari (1985) for the i.i.d. curved exponential family, Taniguchi ((1991), p. 41) for the time series model). From (A.12)–(A.16) we can evaluate the cumulants of V_n as follows:

$$\text{(A.17)} \quad E\{V_n\} = \frac{1}{\sqrt{n}} C_1^{(1)} + o(n^{-1}),$$

$$\text{(A.18)} \quad \mathrm{Var}\{V_n\} = 1 + \frac{1}{n} C_{11}^{(3)} + o(n^{-1}),$$

$$\text{(A.19)} \quad \mathrm{cum}\{V_n, V_n, V_n\} = \frac{1}{\sqrt{n}} C_{111}^{(1)} + o(n^{-1}),$$

$$\text{(A.20)} \quad \mathrm{cum}\{V_n, V_n, V_n, V_n\} = \frac{1}{n} C_{1111}^{(1)} + o(n^{-1}),$$

$$\text{(A.21)} \quad \mathrm{cum}^{(r)}\{V_n, \ldots, V_n\} = O(n^{-r/2+1}), \quad \text{for} \quad r \geq 5.$$

Applying a general formula (e.g., Taniguchi ((1991), p. 15)) to V_n we obtain

$$\begin{aligned}(A.22)\quad P(V_n < x) = \Phi(x) - \phi(x)\Bigg[&\frac{C_1^{(1)}}{\sqrt{n}} + \frac{1}{2}\left(\frac{C_{11}^{(3)}}{n} + \frac{C_1^{(1)}C_1^{(1)}}{n}\right)x \\ &+ \frac{C_{111}^{(1)}}{6\sqrt{n}}(x^2-1) \\ &+ \left(\frac{C_{1111}^{(1)}}{24n} + \frac{C_1^{(1)}C_{111}^{(1)}}{6n}\right)(x^3-3x) \\ &+ \frac{(C_{111}^{(1)})^2}{72n}(x^5 - 10x^3 + 15x)\Bigg] + o(n^{-1}),\end{aligned}$$

which implies (ii).

Proof of Theorem 3.2. The proof of (i) follows directly from Proposition 2.2.

(ii) Substituting (3.2), (3.3), $K = 0$, $\Delta = 0$ and $H = 0$, we have

$$(A.23)\qquad C_{11}^{(3)} = \frac{1}{g^3}(Mg - J^2) - \frac{N}{g^2}.$$

Differentiation of $K = 0$ with respect to θ yields $3N + H = 0$, which implies $N = 0$. Therefore, we can see that $C_{11}^{(3)} = 0$ and $C_{1111}^{(1)} = 0$ under the assumption. Regarding $C_1^{(1)}$, we can set $C_1^{(1)} = 0$ if we choose the constant c appropriately, whence the assertion is proved.

Proof of Theorem 4.1. It can be shown that $\sqrt{n}(\hat{\theta}-\theta)$ has the stochastic expansion

$$\begin{aligned}(A.24)\quad \sqrt{n}(\hat{\theta}-\theta) = &\frac{Z_1}{g_n} + \frac{1}{\sqrt{n}g^2}\left\{Z_1Z_2 - \frac{3J+K}{g}Z_1^2\right\} \\ &+ \frac{1}{ng^3}\Bigg\{Z_1Z_2^2 + \frac{1}{2}Z_1^2Z_3 - \frac{3(3J+K)}{2g}Z_1^2Z_2 \\ &+ \frac{(3J+K)^2}{2g^2}Z_1^3 - \frac{4L+3M+6N+H}{6g}Z_1^3\Bigg\} + o_P(n^{-1}),\end{aligned}$$

where $g_n = g + \frac{\Delta}{n}$ (c.f. Taniguchi ((1991), p. 41)). Expanding $\tilde{e}$ around $\hat{\theta} = \theta$ and $\hat{I} = g$, and substituting (A.24) for $\sqrt{n}(\hat{\theta}-\theta)$, we obtain

$$\begin{aligned}(A.25)\qquad \tilde{e} = &\frac{\sqrt{g}}{g_n}Z_1 + \frac{1}{\sqrt{n}}[a_1Z_1Z_2 + a_2Z_1^2 + a_3] \\ &+ \frac{1}{n}[b_1Z_1^3 + b_2Z_1^2Z_2 + b_3Z_1^2Z_3 + b_4Z_1Z_2^2 + b_5Z_1 + b_6Z_2] \\ &+ o_P(n^{-1}),\end{aligned}$$

where

$$
\begin{aligned}
a_1 &= g^{-3/2} - h^{(1)}(g)g^{-1/2},\\
a_2 &= \frac{A^{(2)}}{2A^{(1)}g^{3/2}} + \frac{(3J+K)}{g^{3/2}}\left\{h^{(1)}(g) - \frac{1}{2g}\right\},\\
a_3 &= -\frac{cg^{1/2}}{A^{(1)}},\\
b_1 &= \frac{4L+3M+6N+H}{2g^{5/2}}\left\{h^{(1)}(g) - \frac{1}{3g}\right\}\\
&\quad + \frac{(3J+K)^2}{2g^{5/2}}\left\{h^{(2)}(g) - \frac{2h^{(1)}(g)}{g} + \frac{1}{g^2}\right\}\\
&\quad + \frac{A^{(2)}(3J+K)}{2A^{(1)}g^{5/2}}\left\{h^{(1)}(g) - \frac{1}{g}\right\} + \frac{A^{(3)}}{6A^{(1)}g^{5/2}},\\
b_2 &= \frac{A^{(2)}}{2A^{(1)}g^{3/2}}\left\{\frac{2}{g} - h^{(1)}(g)\right\} + \frac{(3J+K)}{2g^{3/2}}\left\{-2h^{(2)}(g) + \frac{5h^{(1)}(g)}{g} - \frac{3}{g^2}\right\}\\
b_3 &= \left\{\frac{1}{g} - 2h^{(1)}(g)\right\} \Big/ (2g^{3/2}),\\
b_4 &= \left\{h^{(2)}(g) - \frac{2h^{(1)}(g)}{g} + \frac{2}{g^2}\right\} \Big/ (2g^{1/2}),\\
b_5 &= \frac{h^{(1)}(g)}{g^{1/2}}\left\{\Delta - \frac{(3J+K)c}{A^{(1)}}\right\} + \frac{d}{g^{1/2}},\\
b_6 &= \frac{cg^{1/2}h^{(1)}(g)}{A^{(1)}}.
\end{aligned}
$$

It follows from Assumption 1 and (A.25) that

$$E\{\tilde{e}\} = \frac{1}{\sqrt{n}}C_1^{(1)} + o(n^{-1}), \tag{A.26}$$

$$\mathrm{cum}\{\tilde{e},\tilde{e}\} = 1 + \frac{1}{n}C_{11}^{(3)} + o(n^{-1}), \tag{A.27}$$

$$\mathrm{cum}\{\tilde{e},\tilde{e},\tilde{e}\} = \frac{1}{\sqrt{n}}C_{111}^{(1)} + o(n^{-1}), \tag{A.28}$$

$$\mathrm{cum}\{\tilde{e},\tilde{e},\tilde{e},\tilde{e}\} = \frac{1}{n}C_{1111}^{(1)} + o(n^{-1}), \tag{A.29}$$

$$\mathrm{cum}^{(r)}\{\tilde{e},\dots,\tilde{e}\} = O(n^{-r/2+1}), \qquad \text{for} \quad r \geq 5, \tag{A.30}$$

where

$$c_1^{(1)} = \frac{-J-K}{2g^{3/2}} + \frac{A^{(2)}}{2A^{(1)}g^{1/2}} + \frac{h^{(1)}(g)(2J+K)}{g^{1/2}} - \frac{cg^{1/2}}{A^{(1)}}, \tag{A.31}$$

$$C_{11}^{(3)} = \frac{A^{(3)}}{A^{(1)}g} + \frac{\{A^{(2)}\}^2}{2\{A^{(1)}\}^2 g} + \frac{A^{(2)}}{A^{(1)}}\left\{h^{(1)}(g)\frac{10J+5K}{g} - \frac{3K+4J}{g^2}\right\} \tag{A.32}$$

$$
\begin{aligned}
&+ \frac{h^{(1)}(g)c}{A^{(1)}}(-4J-2K) \\
&+ h^{(2)}(g)\left\{M + \frac{11J^2+12JK+3K^2}{g}\right\} \\
&+ h^{(1)}(g)^2\left\{M + \frac{7J^2+8JK+2K^2}{g}\right\} \\
&+ h^{(1)}(g)\left\{2\Delta + \frac{5M+6L+16N+3H}{g}\right. \\
&\qquad\qquad \left. - \frac{15J^2+21JK+6K^2}{g^2}\right\} \\
&- \frac{\Delta}{g} + 2d + \frac{-4N-L-H}{g^2} + \frac{5K^2+14JK+7J^2}{2g^3},
\end{aligned}
$$

$$
\text{(A.33)} \quad C^{(1)}_{111} = \frac{3A^{(2)}}{A^{(1)}g^{1/2}} + \frac{6(2J+K)h^{(1)}(g)}{g^{1/2}} - \frac{3J+2K}{g^{3/2}},
$$

$$
\begin{aligned}
\text{(A.34)} \quad C^{(1)}_{1111} = &\frac{4A^{(3)}}{A^{(1)}g} + \frac{12\{A^{(2)}\}^2}{\{A^{(1)}\}^2 g} \\
&+ \frac{A^{(2)}}{A^{(1)}g^2}\{-36J-24K+h^{(1)}(g)(120gJ+60gK)\} \\
&+ h^{(2)}(g)\frac{12(4J^2+4JK+K^2)}{g} \\
&+ \{h^{(1)}(g)\}^2\left\{12M + \frac{12(15J^2+16JK+4K^2)}{g}\right\} \\
&+ h^{(1)}(g)\left\{\frac{12(2L+M+5N+H)}{g}\right. \\
&\qquad\qquad \left. + \frac{12(-11J^2-14JK-4K^2)}{g^2}\right\} \\
&+ \frac{-4L-12N-3H}{g^2} + \frac{12(2J^2+3JK+K^2)}{g^3}.
\end{aligned}
$$

If we set $C^{(1)}_{111} = 0$, then

$$
\text{(A.35)} \qquad h^{(1)}(g) = \frac{3J+2K}{6g(2J+K)} - \frac{A^{(2)}}{2A^{(1)}(2J+K)}.
$$

Substitution of (A.35) into $C^{(1)}_{1111}$ yields the differential equation (4.2). For A and h satisfying (4.2) and (4.3) we can solve the equations $C^{(1)}_1 = 0$ and $C^{(3)}_{11} = 0$ with respect to c and d. From Proposition 3.1(ii), the proof is completed.

PROOF OF PROPOSITION 5.1. Expanding $\ell_n(\theta) - \ell_n(\hat\theta)$ at $\theta = \hat\theta$ in a Taylor series we obtain

$$
\text{(A.36)} \quad r = \operatorname{sgn}(\hat\theta - \theta)\{2\ell_n(\hat\theta) - 2\ell_n(\theta)\}^{1/2}
$$

$$= \sqrt{n}(\hat{\theta}-\theta)\left\{\hat{I}+\frac{1}{3}(\hat{\theta}-\theta)\partial^3\ell_n(\hat{\theta})/n-\frac{1}{12}(\hat{\theta}-\theta)^2\partial^4\ell_n(\hat{\theta})/n\right\}^{1/2}$$
$$+\, o_P(n^{-1}).$$

Since the stochastic expansion of $\sqrt{n}(\hat{\theta}-\theta)$ is given by (A.24), expansion of $\hat{I}$ at $\hat{\theta}=\theta$ and substitution of (A.24) into $\hat{I}$ yield

$$\text{(A.37)}\qquad \sqrt{\hat{I}} = \left\{g+\frac{\Delta}{n}\right\}^{1/2}+\frac{1}{\sqrt{n}}\left[\frac{3J+K}{2g^{3/2}}Z_1-\frac{Z_2}{2g^{1/2}}\right]$$
$$+\frac{1}{n}\left[\frac{3(3J+K)}{4g^{5/2}}Z_1Z_2-\frac{1}{2g^{3/2}}Z_1Z_3-\frac{Z_2^2}{8g^{3/2}}\right.$$
$$\left.+\left\{\frac{4L+3M+6N+H}{4g^{5/2}}-\frac{3(3J+K)^2}{8g^{7/2}}\right\}Z_1^2\right]$$
$$+\, o_P(n^{-1}).$$

Substituting (A.37) and (A.24) into (A.36) we can show the stochastic expansion (5.3).

Proof of Proposition 5.2. If $p_n(\boldsymbol{x}_n;\theta)\in\mathcal{F}_e$ and if θ is the natural parameter, we can see that $Z_2=0$, a.s. Therefore $J=L=M=N=0$, which implies that r^* in (5.3) is reduced to

$$\text{(A.38)}\qquad r_e^* = Z_1\left\{g+\frac{\Delta}{n}\right\}^{-1/2}+\frac{1}{\sqrt{n}}\left\{\frac{K}{6g^{5/2}}(g-Z_1^2)\right\}$$
$$+\frac{1}{n}\left[\frac{K^2}{9g^{9/2}}Z_1^3-\frac{H}{24g^{7/2}}Z_1^3+\left(\frac{H}{8g^{5/2}}-\frac{7K^2}{36g^{7/2}}\right)Z_1\right]$$
$$+\, o_P(n^{-1}).$$

We can evaluate the asymptotic cumulants of r_e^* as follows:

$$\text{(A.39)}\qquad E\{r_e^*\}=o(n^{-1}),$$
$$\text{(A.40)}\qquad \mathrm{cum}\{r_e^*,r_e^*\}=1+o(n^{-1}),$$
$$\text{(A.41)}\qquad \mathrm{cum}\{r_e^*,r_e^*,r_e^*\}=o(n^{-1}),$$
$$\text{(A.42)}\qquad \mathrm{cum}\{r_e^*,r_e^*,r_e^*,r_e^*\}=o(n^{-1}),$$
$$\text{(A.43)}\qquad \mathrm{cum}^{(k)}\{r_e^*,\ldots,r_e^*\}=O(n^{-k/2+1}),\quad \text{for}\quad k\geq 5.$$

The assertion follows from (A.39)–(A.43).

Proof of Proposition 5.3. From the stochastic expansion (5.3) it is shown that

$$\text{(A.44)}\qquad E\{r^*\}=\frac{1}{\sqrt{n}}\frac{J}{2g^{3/2}}+o(n^{-1}).$$

Since $p_n(\boldsymbol{x}_n;\theta) \notin \mathcal{F}_e$, $J \neq 0$ in general. In view of Proposition 3.1, r^* does not become the third-order normalizing transformation.

PROOF OF PROPOSITION 5.4. Since we assumed that $g_n = g_n(\theta)$ is constant, and that $\boldsymbol{e} \sim N(\boldsymbol{0}, \Sigma)$, we can show

$$J = K = N = 0, \qquad M = -L = \frac{\partial^2}{\partial\theta^2}{}' \boldsymbol{m}(\theta)\Sigma^{-1}\frac{\partial^2}{\partial\theta^2}\boldsymbol{m}(\theta).$$

The stochastic expansion of q^* is given by

$$q^* = \sqrt{n}(\hat{\theta}-\theta)\sqrt{g_n} - \frac{M}{6\sqrt{gn}}\{\sqrt{n}(\hat{\theta}-\theta)\}^3 + o_p(1). \tag{A.45}$$

For $U = \sqrt{n}(\hat{\theta}-\theta)$, it is shown (c.f., Taniguchi (1991)) that

$$\begin{aligned} &E(U) = o(n^{-1}), \qquad \mathrm{cum}\{U,U\} = \{g_n/n\}^{-1} + M/g^3n + o(n^{-1}),\\ &\mathrm{cum}\{U,U,U\} = o(n^{-1}), \qquad \mathrm{cum}\{U,U,U,U\} = M/g^4n + o(n^{-1}),\end{aligned}$$

which, together with (A.45), lead to

$$\begin{aligned} &E(q^*) = o(n^{-1}), \qquad \mathrm{cum}\{q^*,q^*\} = 1 + o(n^{-1}),\\ &\mathrm{cum}\{q^*,q^*,q^*\} = o(n^{-1}), \qquad \mathrm{cum}\{q^*,q^*,q^*,q^*\} = o(n^{-1}),\\ &\mathrm{cum}^{(k)}\{q^*,\dots,q^*\} = O(n^{-k/2+1}), \qquad \text{for} \quad k \geq 5.\end{aligned}$$

The assertion follows from the above.

REFERENCES

Amari, S. (1985). Differential geometrical methods in statistics, *Lecture Notes in Statist.*, **28**, Springer, New York.

Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio, *Biometrika*, **78**, 557–563.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansions, *Ann. Statist.*, **6**, 434–451.

Efron, B. (1975). Defining the curvature of a statistical problem, *Ann. Statist.*, **3**, 1189–1242.

Fang, C. and Krishnaiah, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations, *J. Multivariate Anal.*, **12**, 39–63.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms, *J. Roy. Statist. Soc. Ser. B*, **15**, 193–232.

Jensen, J. L. (1992). The modified signed likelihood statistic and saddlepoint approximations, *Biometrika*, **79**, 693–703.

Konishi, S. (1978). An approximation to the distribution of sample correlation coefficient, *Biometrika*, **65**, 654–656.

Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika*, **68**, 647–651.

Konishi, S. (1987). Transformations of statistics in multivariate analysis, *Advances in Multivariate Statistical Analysis* (A. K. Gupta ed.), 213–231, Reidel, Dordrecht.

Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals, *Ann. Statist.*, **19**, 2209–2225.

Krishnaiah, P. R. and Lee, J. C. (1979). On the asymptotic joint distributions of certain functions of the eigenvalues of four random matrices, *J. Multivariate Anal.*, **9**, 248–258.

Pázman, A.(1990). Almost exact distributions of estimators II—flat nonlinear regression models, *Statistics*, **21**, 21–33.

Phillips, P. C. B. (1979). Normalizing transformations and expansions for functions of statistics, Research notes, Birmingham University.

Takeuchi, K. and Morimune, K. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system, *Econometrica*, **53**, 177–200.

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 Statist.*, **68**, Springer, Heidelberg.

Taniguchi, M. and Krishnaiah, P. R. (1987). Asymptotic distributions of functions of the eigenvalues of sample covariance matrix and canonical correlation matrix in multivariate time series, *J. Multivariate Anal.*, **22**, 156–176.

Taniguchi, M., Krishnaiah, P. R. and Chao, R. (1989). Normalizing transformations of some statistics of Gaussian ARMA processes, *Ann. Inst. Statist. Math.*, **41**, 187–197.