The Annals of Statistics
1991, Vol. 19, No. 2, 961–981

ASYMPTOTIC THEORY OF SEQUENTIAL ESTIMATION: DIFFERENTIAL GEOMETRICAL APPROACH

BY ICHI OKAMOTO, SHUN-ICHI AMARI AND KEI TAKEUCHI

University of Tokyo

Sequential estimation continues observations until the observed sample satisfies a prescribed criterion. Its properties are superior on the average to those of nonsequential estimation in which the number of observations is fixed a priori. A higher-order asymptotic theory of sequential estimation is given in the framework of geometry of multidimensional curved exponential families. This gives a design principle of the second-order efficient sequential estimation procedure. It is also shown that a sequential estimation can be designed to have a covariance stabilizing effect at the same time.

1. Introduction. The higher-order asymptotic theory of statistical inference has been developed mostly for these ten years by many researchers, for example, Rao (1962), Akahira and Takeuchi (1981), Pfanzagl (1982), Chibisov (1974), Ghosh and Subramanyam (1974) and Bickel, Chibisov and van Zwet (1981). It has given rise to the differential geometrical theory of statistics [see, e.g., Efron (1975), Amari (1985)], which is proved to provide statistics with a new framework [Barndorff-Nielsen, Cox and Reid (1986), Kass (1989), Amari (1987), Amari, Barndorff-Nielsen, Kass, Lauritzen and Rao (1987), Barndorff-Nielsen (1988), Vos (1989)]. This framework may be called information geometry and is applicable not only to statistics but also to more wide areas of information sciences such as control systems theory or time series analysis [Amari (1987)], information theory [Amari and Han (1989), Amari (1989)].

Main results of the higher-order asymptotics of estimation are summarized as follows.

1. Information loss: The square $(H_M^e)^2$ of the e-curvature (which is a generalization of Efron's statistical curvature) of a statistical model gives the amount of loss of Fisher information by summarizing observed data into the m.l.e. or other second-order efficient estimators.
2. Estimation error: Let $\hat{u}^*$ be the bias-corrected version of an efficient estimator. Its covariance matrix is asymptotically expanded as

$$\text{Covariance matrix} = \frac{1}{N}E_1 + \frac{1}{N^2}E_2 + O(N^{-3}),$$

where N is the number of observations and E_1 is the inverse of Fisher

Received November 1989; revised April 1990.

AMS 1980 *subject classifications*. Primary 62F10; secondary 62F12.

Key words and phrases. Asymptotic theory, conformal transformation, covariance stabilization, differential geometry, higher-order asymptotics, sequential estimation, statistical curvature, stopping rule.

information matrix. The second-order term E_2 of the covariance is further decomposed into the sum of three nonnegative terms

$$E_2 = \tfrac{1}{2}(\Gamma_M^m)^2 + (H_M^e)^2 + \tfrac{1}{2}(H_A^m)^2.$$

The first term is the square of the m-connection of the model (the Bhattacharya bound), the second is the square of the e-curvature of the model and the third is the square of the m-curvature of the estimating submanifold of the estimator (which vanishes for the m.l.e. or other second-order efficient estimators).

3. Observed information and ancillary: The observed information (which is the negative of the second derivative of log-likelihood) is different from the (expected) Fisher information by the amount of the e-curvature direction components of the asymptotic ancillary statistic. The covariance of an estimator is evaluated more accurately by the observed information rather than the expected one [see Efron and Hinkley (1978); Cox, (1980); Barndorff-Nielsen, (1980); Amari (1985); see also Barndorff-Nielsen (1988)].

Among the second-order covariance term E_2 in 2, the Bhattacharya bound term (Γ_M^m) can be made equal to 0 locally by choosing an adequate parameterization. The mixture curvature term (H_A^m) can be made equal to 0 by taking the second-order efficient estimator such as the m.l.e. Therefore, the statistical curvature (H_M^e) of the model represents the essential information structure inherent to the statistical model, as is also manifested in 1. However, as is shown in 3, the information loss due to the statistical curvature is included in the asymptotic ancillary statistic, so that it might be recovered by some statistical procedure which makes use of the ancillary. This suggests a sequential estimation procedure of continuing new observations until the observed Fisher information reaches a prescribed (large) value, instead of fixing the number N of observation. Such an estimator is expected to have a uniformly better characteristic on the average [see, e.g., Sørensen (1986)]. Takeuchi and Akahira (1988) formulated this scheme rigorously and analyzed the higher-order efficiency of sequential estimation procedures in the scalar parameter case [see also Akahira and Takeuchi (1989)]. They showed that the e-curvature term in the second-order covariance can be eliminated by a second-order efficient sequential estimator, where the expected number of observations is put equal to N for comparison. The m.l.e. with the best stopping rule gives such a sequential estimator. This shows that sequential estimators are superior to (nonsequential) estimators in the asymptotic sense.

It is important to know the reason why the e-curvature term vanishes from the geometrical viewpoint. We can then easily generalize the results to the multiparameter case by using the geometrical method and moreover we can analyze characteristics of more general sequential estimation procedures. A statistical manifold is uniformly enlarged by N times when we use N observations, keeping the intrinsic features of the manifold unchanged. However, in a sequential estimation procedure with an adequate stopping rule, the observed number N is a random variable depending on the position of a statistical manifold. This causes a nonuniform expansion of a statistical manifold. Such an expansion is called the conformal transformation in geometry, since it changes the scale locally and isotropically but it does not change the shape of a figure (it does not change the orthogonality). It is easy to understand intuitively that, given a curve, the curvature decreases if the positive side (inner side) of the curve in the enveloping plane is enlarged while the outer side is not. The result of Takeuchi and Akahira is interpreted such that it is possible to reduce the e-curvature of a statistical model to 0 by a conformal transformation. This implies that a stopping rule of a sequential procedure gives a conformal transformation by which the e-curvature of a statistical model is modified. The conformal geometry gives an adequate framework for the analysis of the sequential inferential procedures if we extend the concept of the conformal transformation to the statistical manifold (Riemannian manifold with a dual couple of affine connections).

The present paper is devoted to the mathematical analysis of the higher-order asymptotics of sequential estimation procedures in the multiparameter case in the framework of conformal geometry (see Appendix 1). The conformal geometry of a statistical manifold (manifold with dual affine connections) itself is a new interesting geometrical problem which will be studied elsewhere.

We use a multidimensional curved exponential family as a statistical model and show that the mean exponential curvature can be eliminated by the conformal transformation associated with a stopping rule. This implies that the e-curvature term is always reduced by taking an adequate stopping rule and that the characteristics of estimators are improved by sequential procedures. The e-curvature can be reduced to 0 in the scalar parameter case, but this is not always so in the general multiparameter case. It is the mean e-curvature that can be reduced to 0. In addition to the elimination of the e-curvature, there is another possibility of improving the characteristics of estimators by sequential procedures. By choosing an adequate stopping rule, it is possible to get more covariance stabilized inferential procedures. Greenwood and Shiryaev (1985, 1988) also treated this effect in an AR model of time series.

2. Geometry of curved exponential family.

2.1. *Curved exponential family.* Let $S = \{p(x, \theta)\}$ be a full regular minimally represented exponential family of distributions, where

$$p(x, \theta) = \exp\{\theta^i x_i - \Psi(\theta)\}$$

is a density function of a vector random variable $x = (x_1, \dots, x_n)$ parameterized by a vector parameter $\theta = (\theta^1, \dots, \theta^n)$, $\theta \in \Theta$, with respect to a common measure μ. The Einstein summation convention is assumed throughout the paper, so that summation is automatically taken over indices repeated twice so that $\theta^i x_i$ automatically implies $\Sigma\, \theta^i x_i$. We also assume that Θ is homeomorphic to $\mathbf{R}^n$.

The family S can be regarded as a statistical manifold [Amari (1985)]. The natural or canonical parameter θ is a coordinate system to specify a point, that is, a distribution $p(x, \theta) \in S$. The expectation parameter $\eta = (\eta_i)$, defined by

$$\eta_i = E_\theta[x_i] = \partial_i \Psi(\theta),$$

also plays a role of another coordinate system, where E_θ (subscript θ is often omitted) denotes the expectation with respect to $p(x, \theta)$ and $\partial_i = \partial/\partial\theta^i$.

The geometry of S is determined by the following two tensor quantities

$$g_{ij}(\theta) = E\left[\partial_i l(x, \theta)\, \partial_j l(x, \theta)\right] = \partial_i \partial_j \Psi(\theta),$$
$$T_{ijk}(\theta) = E\left[\partial_i l(x, \theta)\, \partial_j l(x, \theta)\, \partial_k l(x, \theta)\right] = \partial_i \partial_j \partial_k \Psi(\theta),$$

in terms of the θ-coordinate system, where $l(x, \theta) = \log[p(x, \theta)]$. The first is the Fisher information metric and the second is called the skewness tensor by Lauritzen (1987), from which a pair of dual affine connections are defined [Nagaoka and Amari (1982); Amari (1985)].

A family $M = \{\bar{p}(x, u)\}$ of probability distributions parameterized by an m-dimensional vector parameter u is said to be an (n, m)-curved exponential family embedded in S, when

$$\bar{p}(x, u) = p(x, \theta(u)) = \exp\{\theta^i(u) x_i - \Psi(\theta(u))\}, \qquad u \in U,$$

where U is homeomorphic to an m-dimensional Euclidean space $\mathbb{R}^m$ $(m < n)$. We assume that $\theta(u)$ is a smooth injection from U to S, so that M is regarded as a submanifold of S. We use indices i, j, k, to denote quantities in terms of the coordinate system θ or η of S and indices a, b, c, and so on to denote quantities in M. Therefore, the parameter u is written as u^a $(a = 1, 2, \ldots, m)$ in the component form. The geometry of M is defined similarly and is identical with that induced from the enveloping manifold S.

2.2. *Estimators.* Let

$$\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(N)})$$

be a set of N independent vector observations from the same distribution $p(x,\theta) \in S$, where the sample size N is fixed. The probability density function of $\mathbf{x}$ is written as

$$p(\mathbf{x}, \theta) = \prod_{i=1}^{N} p(x^{(i)}, \theta).$$

It is easy to show that the metric tensor and the skewness tensor become Ng_{ij} and NT_{ijk}, respectively. This shows that the geometry of S^N based on N observations is similar to S based on one observation, except that the basic quantities are enlarged by N times.

Let

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}.$$

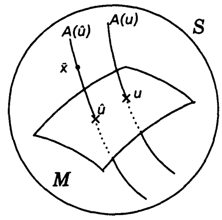

FIG. 1.

Then, x is a minimal sufficient statistic. Let us define a point (i.e., a distribution) $\hat{\eta}$ in S, whose η-coordinates are put equal to $\bar{x}$,

$$\hat{\eta} = \bar{x}.$$

Let $\hat{\theta}$ be the corresponding θ-coordinates of the same point. Obviously, $\hat{\theta}$ (or $\hat{\eta}$) is the maximum likelihood estimator in S. We call this probability distribution $p(x, \hat{\theta})$ the observed point in S determined from the observation.

When the true distribution is $\bar{p}(x, u) = p(x, \theta(u))$ belonging to M, we need to estimate u. We consider an estimator

$$\hat{u} = e(\bar{x}),$$

which is a function of the sufficient statistics $\bar{x}$, or equivalently the observed point $\hat{\eta} \in S$. Given a function e, we can define the following subset of S

$$A(u) = \{\eta \in S | e(\eta) = u\}$$

attached to a point $u \in M$. The set $A(\hat{u})$ consists of those points in S such that the value of the estimator is $\hat{u}$ when and only when the observed point belongs to $A(\hat{u})$. One may write

$$A(\hat{u}) = e^{-1}(\hat{u}).$$

We call $A(\hat{u})$ the estimating submanifold or ancillary submanifold attached to $\hat{u}$. We assume that each $A(\hat{u})$ forms a smooth $(n - m)$-dimensional submanifold of S, the $A(u)$'s forming a smooth foliation (a smooth partition) of S. An estimator $\hat{u}$ is then regarded as a projection from S to M through $A(u)$'s (see Figure 1).

An estimator $\hat{u}$ is characterized geometrically by the properties of the family of estimating submanifolds $A(u)$ associated with it. Let us denote by $P_M(S)$ the set of all the smooth projections from the S to M, which is equivalent to the set of all the estimators we treat here. We characterize projections or estimators in terms of the associated $A(u)$'s.

DEFINITION 2.1.

(i) A projection from S to M is said to be *consistent*, when $\theta(u) \in A(u)$ for any u, that is, when $A(u)$ includes the distribution specified by u. The set of consistent projections is denoted by $P_{M0}(S)$.

(ii) A consistent projection is said to be orthogonal when $A(u)$ is orthogonal to M at their intersection $\theta(u)$. The set of *orthogonal* projections is denoted by $P_{M1}(S)$.

(iii) An orthogonal projection is said to be locally *m-flat* when $A(u)$ has zero m-embedding curvature on M. The set of locally m-flat projections is denoted by $P_{M2}(S)$.

The definitions of the orthogonality and m-flatness are given in Appendix 1. Let us introduce a coordinate system $v = (v^{m+1}, v^{m+2}, \dots, v^n)$ to each submanifold $A(u) \in P_{M1}(S)$ such that the origin $v = 0$ is at the intersection of $A(u)$ and M. Since u is a coordinate system of M, the combined system $w = (u, v)$ gives a new coordinate system of the entire S, where u designates that the point w is in $A(u)$ and v denotes the position of w in $A(u)$. We use indices α, β, γ, and so on for quantities related to the coordinate system w and indices κ, λ, μ, and so on for quantities related to the coordinate system v. We may write $w = (w^\alpha) = (u^a, v^\kappa)$. The coordinate transformation from w to θ is written as

$$\theta^i = \theta^i(w^\alpha),$$

with the Jacobian matrix

$$B^i_\alpha = \frac{\partial \theta^i}{\partial w^\alpha}.$$

The geometrical quantities of S, M and $A(u)$ can be easily obtained with the help of the new coordinate system. We give the definitions and some results in Appendix 1.

3. Sequential estimation and conformal transformation. In a sequential estimation procedure, the number N of observations is a random variable determined from the result of observations. In other words, it has a stopping rule which decides whether to stop observation or to continue it further depending on the result of the past observations. A stopping rule is given by a scalar function $\nu(\eta)$ called the expansion factor or the gauge.

Let K be a large number playing the role of the average number of observations and let $\nu(\eta)$ be a smooth positive scalar function defined on S in the η-coordinate system. We adopt a stopping rule by which the random sample size N satisfies

$$E_\theta[N] = K\nu(\eta) + O(1), \tag{3.1}$$

where $\theta = \theta(\eta)$. This implies that, when the true distribution is $p(x, \theta)$, the expectation of the sample size N is required to be approximately equal to $K\nu(\eta)$. The gauge function $\nu(\eta)$ is called an expansion factor, because the manifold S is expanded about $K\nu(\eta)$ times at η by this stopping rule, where K is a large number such that asymptotic results hold as K tends to infinity. This nonuniform expansion gives rise to a conformal transformation of S.

A rough idea of a stopping rule is as follows. Let $\bar{x}$ be the observed point by N independent observations. Since $\bar{x}$ converges to the true point η as N increases, the required expected number $E[N] = K\nu(\eta)$ is approximated by $K\nu(\bar{x})$. Therefore, the stopping rule is such that we stop observation when the number N of observations becomes larger than $K\nu(\bar{x})$. More precisely, we assume that the number of observations is determined by our stopping rule such that

$$N = K\nu(\bar{x}) + c(\bar{x}) + \varepsilon \tag{3.2}$$

holds, where $\bar{x}$ is the observed point by N observations, $c(\bar{x})$ is a function of order 1 and ε is a small order term asymptotically independent of $\bar{x}$ satisfying $E[\varepsilon] = o(1)$. These terms c and ε are introduced for N to satisfy the requirement (3.1), because

$$E[\nu(\bar{x})] = \nu(\eta) \tag{3.3}$$

does not hold exactly. The term c is due to the bias of $\bar{x}$ from the true η and will be explicitly given later. The term ε includes a rounding error, because N is an integer. Following Takeuchi and Akahira (1988), we further assume that the tth moments of N ($t = 1, 2, 3, 4$) are of order K^t,

$$E_u[N^t] = O(K^t). \tag{3.4}$$

Since the higher order asymptotic theory of statistical inference is constructed on the geometry S of a statistical model in a unified manner [Amari (1985); Amari, Barndorff-Nielsen, Kass, Lauritzen and Rao (1987)], we show how the geometry of S is changed by introducing a stopping rule and then give an example.

Let

$$p(\mathbf{x}, \theta) = \prod_{i=1}^{N} p(x^{(i)}, \theta), \tag{3.5}$$

where N is a random variable under an expansion factor $\nu(\eta)$. By this sequential rule, we have an extended statistical manifold

$$'S = \{p(\mathbf{x}, \theta)\}.$$

The metric and the skewness tensor of $'S$ is calculated by using Lemma 1 in Appendix 2 as

$$'g_{ij} = K\nu g_{ij},$$

$$'T_{ijk} = K\nu\left(T_{ijk} + 3g_{(ij}s_{k)}\right),$$

where

$$s_i(\theta) = \partial_i \log \nu(\theta),$$

$$3g_{(ij}s_{k)} = g_{ij}s_k + g_{jk}s_i + g_{ki}s_j.$$

Normalizing the magnitude, the geometry of $'S$ is defined as

$$'g_{ij} = g_{ij}, \tag{3.6}$$

$$'T_{ijk} = T_{ijk} + 3g_{(ij}s_{k)}. \tag{3.7}$$

The α-connection of $'S$ is defined therefrom. This defines a conformal transformation of the statistical manifold S induced by the gauge or expansion factor $\nu(\eta)$ (see Appendix 1).

Let us decompose s_i into the a- and κ-components,

$$s_a = \partial_a \log \nu(w), \tag{3.8}$$

$$s_\kappa = \partial_\kappa \log \nu(w) \tag{3.9}$$

in terms of the w-coordinates. The m-connection of $'M$ is then given by

$$'\Gamma^m_{abc} = \Gamma^m_{abc} + g_{ca}s_b + g_{cb}s_a. \tag{3.10}$$

When $A(u)$'s are orthogonal to M, the e-curvature of M is changed into

$$'H^{(e)}_{ab\kappa} = H^{(e)}_{ab\kappa} - g_{ab}s_\kappa, \tag{3.11}$$

while the m-curvature of $A(u)$ is kept invariant

$$'H^{(m)}_{\kappa\lambda a} = H^{(m)}_{\kappa\lambda a}. \tag{3.12}$$

EXAMPLE OF STOPPING RULE. Takeuchi and Akahira (1988) developed the second-order efficient sequential m.l.e. for a one-dimensional model by using the following rule: Given $\nu(u)$, stop the observation when the following criterion is satisfied,

$$-\frac{1}{g(\hat{u})}\sum_{i=1}^{N} l''(x^{(i)}, \theta(\hat{u})) = K\nu(\hat{u}) + c(\hat{u}) + \varepsilon, \tag{3.13}$$

where $\hat{u}$ is the m.l.e., $l'' = d^2/du^2\, l(x, \theta(u))$ is the second derivative of the log-likelihood, $g(u)$ is the Fisher information and $c(u)$ and ε are determined from

$$E_{\hat{u}}[N] = K\nu(\hat{u}). \tag{3.14}$$

The previous rule implies that observation stops when an observed Fisher information reaches the prescribed amount, because $-\sum l''(x^{(i)}, \theta(\hat{u}))$ is the amount of observed Fisher information. The Fisher information is a tensor in the multiparameter case, so that we cannot apply this rule (3.13) directly to the latter case. However, we can generalize it by evaluating the trace of the observed information as

$$-\frac{1}{m}\sum_{i=1}^{N} \partial_a \partial_b l(x^{(i)}, \theta(\hat{u}))g^{ab}(\hat{u}) = K\nu(\hat{u}) + c(\hat{u}) + \varepsilon, \tag{3.15}$$

where (g^{ab}) is the inverse of the Fisher information matrix (g_{ab}). Expanding (3.15), we have

$$N = K\nu(\hat{u})\left(1 + H^{(e)}_\kappa \hat{v}^\kappa + O_p(1)\right), \tag{3.16}$$

where $(\hat{u}, \hat{v})$ is the w-coordinates of the observed point $\bar{x}$ and

$$H_{\kappa}^{(e)} = \frac{1}{m} H_{ab\kappa}^{(e)} g^{ab} \tag{3.17}$$

is the mean e-curvature. Therefore, when the expansion factor $\nu(u)$ is specified on M, the stopping rule determines the expansion factor $\nu(w)$ in the entire S (or in a neighborhood of M) as

$$\nu(w) = \nu(u)\big(1 + H_{\kappa}^{(e)}(u) v^{\kappa}\big),$$

extending the given $\nu(u)$ in a tubular neighborhood of M. The vectors s_a and s_κ are

$$s_a = \partial_a \log \nu(u),$$
$$s_\kappa = H_{\kappa}^{(e)}$$

on M. We can calculate the geometrical quantities of $'S$ by using them.

4. Characteristics of sequential estimators. We study higher-order properties of sequential estimators and show how they depend on the expansion factor $\nu(w)$ or the associated stopping rule. This gives a design principle to the sequential estimator. We assume that estimators are functions of observed points $\bar{x}$ accompanied by smooth estimating submanifolds $A(u)$. The observed point $\bar{x}$ is not the sufficient statistic in the sequential case, but the loss of information by using only $\bar{x}$ instead of the whole sequence of observations is proved to be of order $1/K$. The second-order asymptotic properties of an estimator are then proved to depend only on $A(u)$'s and $\nu(w)$.

We first give a theorem on the consistency and first-order efficiency of sequential procedures. Since the result is similar to the nonsequential case, the proof is omitted.

THEOREM 4.1. (i) *A procedure is said to be consistent when the estimator $\hat{u}$ converges in probability to the true parameter u as K tends to infinity. A procedure is consistent if the estimator $\hat{u}$ belongs to $P_{M0}(S)$, that is, each $A(u)$ includes the point $\eta(u)$.*

(ii) *A consistent procedure is said to be first-order efficient when $\sqrt{K}\,\nu(\hat{u} - u)$ converges in distribution to the normal distribution $N(0, g)$ with $g = (g^{ab})$, the inverse of the Fisher information matrix as K tends to infinity. A procedure is first-order efficient if and only if the estimator belongs to $P_{M1}(S)$, that is, the associated manifolds $A(u)$'s, are orthogonal to M.*

REMARK. In many cases $A(u)$ depends on the number of observations N. In these cases, it is necessary that $A(u)$ includes $\eta(u)$ as $K \to \infty$ and that $A(u)$ is orthogonal to M as $K \to \infty$. We do not mention it, because the situation is the same as the nonsequential case [see Amari (1985)].

In the following, we treat first-order efficient procedures. We first give the asymptotic bias of the estimator in order to obtain its bias-corrected version;

we then give the second-order asymptotic covariance of the bias-corrected estimator.

THEOREM 4.2. *The asymptotic bias of a sequential estimator is given by*

$$E\left[\sqrt{K}\,\nu(\hat{u}^a - u^a)\right] = b^a(u) + O(K^{-1}), \tag{4.1}$$

where

$$b^a(u) = -\frac{1}{2\sqrt{K}\,\nu}\left(\Gamma^{(m)a}_{\alpha\beta}g^{\alpha\beta} + 2s^a\right). \tag{4.2}$$

We obtain the bias-corrected estimator $\hat{u}^*$ *as*

$$\hat{u}^{*a} = \hat{u}^a - \frac{1}{2\sqrt{K}\,\nu}b^a(\hat{u}). \tag{4.3}$$

THEOREM 4.3. *The asymptotic covariance of a bias-corrected first-order efficient sequential estimator is given by*

$$\begin{aligned} &E\left[K\nu(\hat{u}^{*a} - u^a)(\hat{u}^{*b} - u^b)\right] \\ &\quad = g^{ab} + \frac{1}{K\nu}\left\{\frac{1}{2}{}'\Gamma^{(m)2ab}_{M} + {}'H^{(e)2ab}_{M} + \frac{1}{2}H^{(m)2ab}_{A}\right\} + O(K^{-2}), \end{aligned} \tag{4.4}$$

where the primed quantities ${}'\Gamma^{(m)}_{M}$ *and so on are those obtained by the conformal transformation of* S.

PROOF OF THEOREMS 4.2 AND 4.3. Let w be the parameter $(u, 0)$ of the true distribution and let us define

$$\tilde{w}^\alpha = \sqrt{K}\,\nu(\hat{w}^\alpha - w^\alpha),$$
$$\tilde{u}^a = \sqrt{K}\,\nu(\hat{u}^a - u^a),$$
$$\tilde{v}^k = \sqrt{K}\,\nu v^\kappa.$$

Denoting the log-likelihood of $\mathbf{x}$ by $l(\mathbf{x}, \theta(w))$, we have

$$\partial_i l(\mathbf{x}, \theta(w)) = N\bar{x}_i - N\eta_i(w), \tag{4.5}$$

where $\eta_i(w) = \partial_i\Psi(\theta(w))$. Expanding $\bar{x}_i = \eta_i(\hat{w}) = \eta_i(\hat{u}, \hat{v})$, we obtain

$$\begin{aligned} \bar{x}_i = \eta_i(w) &+ B_{\alpha i}\frac{\tilde{w}^\alpha}{\sqrt{K}\,\nu} + \frac{1}{2}C_{\alpha\beta i}\frac{\tilde{w}^\alpha\tilde{w}^\beta}{K\nu} \\ &+ \frac{1}{6}D_{\alpha\beta\gamma i}\frac{\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma}{(K\nu)^{3/2}} + O_p(K^{-2}), \end{aligned} \tag{4.6}$$

where

$$B_{\alpha i} = \partial_\alpha\eta_i,$$
$$C_{\alpha\beta i} = \partial_\alpha\,\partial_\beta\eta_i,$$
$$D_{\alpha\beta\gamma i} = \partial_\alpha\,\partial_\beta\,\partial_\gamma\eta_i.$$

On the other hand, we expand (3.2) into

$$(4.7)\quad N = K\nu(w) + K\partial_\alpha\nu\frac{\tilde{w}^\alpha}{\sqrt{K}\nu} + \frac{1}{2}K\partial_\alpha\partial_\beta\nu\frac{\tilde{w}^\alpha\tilde{w}^\beta}{K\nu} + c(w) + \varepsilon + o_p(1).$$

Substituting (4.6) and (4.7) in (4.5), we have

$$(4.8)\quad \begin{aligned} \partial_i l(\mathbf{x}, \theta(w)) &= \sqrt{K}\nu B_{\alpha i}\tilde{w}^\alpha + \frac{1}{2}(C_{\alpha\beta i} + 2s_{(\alpha}B_{\beta)i})\tilde{w}^\alpha\tilde{w}^\beta \\ &\quad + \frac{1}{2}B_{\alpha i}\partial_\beta\partial_\gamma\nu\frac{\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma}{\nu\sqrt{K}\nu} + B_{\alpha i}(c+\varepsilon)\frac{\tilde{w}^\alpha}{\sqrt{K}\nu} \\ &\quad + \frac{1}{2}C_{\alpha\beta i}s_\gamma\frac{\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma}{\sqrt{K}\nu} + O_p\left(\frac{1}{K}\right). \end{aligned}$$

When we put

$$\tilde{x}_i = \frac{1}{\sqrt{K}\nu}\partial_i l(\mathbf{x}, w),$$

it is rewritten as

$$(4.9)\qquad \tilde{x}_i = B_{\alpha i}\tilde{w}^\alpha + \frac{1}{2\sqrt{K}\nu}(C_{\alpha\beta i} + 2s_{(\alpha}B_{\beta)i})\tilde{w}^\alpha\tilde{w}^\beta + O_p(K^{-1}),$$

$$(4.10)\qquad \tilde{w}^\alpha = g^{\alpha\beta}B^i_\beta\tilde{x}_i - \frac{1}{2\sqrt{K}\nu}\left(C^\alpha_{\beta\gamma} + 2s_{(\alpha}\delta^\alpha_{\beta)}\tilde{w}^\alpha\tilde{w}^\beta\right) + O_p(K^{-1}),$$

where

$$B^i_\beta = \partial_\beta\theta^i,$$

$$C^\alpha_{\beta\gamma} = C_{\beta\gamma i}B^i_\delta g^{\delta\alpha}, \qquad \delta^\alpha_\beta = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases}$$

Since we have

$$g^{\alpha\beta}B^i_\beta\tilde{x}_i = g^{\alpha\beta}\partial_\beta l(\mathbf{x}, w),$$

by Lemma 1 and Lemma 2 in Appendix 2, the moments of $\tilde{w}$ are obtained as

$$(4.11)\qquad E[\tilde{w}^\alpha\tilde{w}^\beta] = \frac{1}{\sqrt{K}\nu}g^{\alpha\beta} + O(K^{-1}),$$

$$(4.12)\qquad E[\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma] = -\frac{1}{\sqrt{K}\nu}{}'T^{\alpha\beta\gamma} + O(K^{-1}),$$

$$(4.13)\qquad E[\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma\tilde{w}^\delta] = 3g^{(\alpha\beta}g^{\gamma\delta)} + O(K^{-1}),$$

where

$$3g^{(\alpha\beta}g^{\gamma\delta)} = g^{\alpha\beta}g^{\gamma\delta} + g^{\alpha\gamma}g^{\beta\delta} + g^{\alpha\delta}g^{\beta\gamma}.$$

Substituting (4.11) in (4.10), we obtain the expectation of $\tilde{w}$ as

$$E[\tilde{w}^\alpha] = -\frac{1}{2\sqrt{K}\nu}\left(C^\alpha_{\beta\gamma}g^{\beta\gamma} + 2s^\alpha\right) + O(K^{-1}). \tag{4.14}$$

Since Amari (1985) showed that

$$C^\alpha_{\beta\gamma} = \Gamma^{(m)\alpha}_{\beta\gamma} = \Gamma^{(m)}_{\beta\gamma\delta}g^{\alpha\delta},$$

(4.14) is rewritten as

$$\begin{aligned} E[\tilde{w}^\alpha] &= -\frac{1}{2\sqrt{K}\nu}\left(\Gamma^{(m)\alpha}_{\beta\gamma}g^{\beta\gamma} + 2s^\alpha\right) + O(K^{-1}) \\ &= \frac{-1}{2\sqrt{K}\nu}\left({}'\Gamma^{(m)\alpha}_{\beta\gamma}g^{\beta\gamma}\right) + O(K^{-1}). \end{aligned} \tag{4.15}$$

This completes the proof of Theorem 4.2.

In order to obtain the covariance of the bias-corrected estimator $\hat{u}^*$, let

$$\tilde{u}^{*a} = \sqrt{K}\nu(\hat{u}^{*a} - u^a).$$

Since

$$E[\tilde{u}^{*a}] = O(K^{-1}),$$

we obtain (4.4) using Lemma 2 in Appendix 2. Hence, we proved Theorem 4.3. □

It is noted that the scalar function $c(\eta)$ is given explicitly by

$$c(\eta) = \tfrac{1}{2}\left(\partial_\beta s_\gamma - \Gamma^{(m)\alpha}_{\beta\gamma}s_\alpha - s_\beta s_\gamma\right)g^{\beta\gamma}, \tag{4.16}$$

which is shown by substituting (4.11) and (4.14) in (4.7).

For practical readers, it is remarked that Akahira and Takeuchi (1989) made a comment on how the stopping rule is designed so that sample size N satisfies (3.2) for given $\nu(w)$.

We now evaluate the procedures from the point of view of information loss. The amount of information loss by summarizing whole the data in an estimator $\hat{u}$ is given by

$$\Delta g_{ab} = E\left[\mathrm{Cov}\left[\partial_a l(\mathbf{x}, \theta(u)), \partial_b l(\mathbf{x}, \theta(u)) \middle| \hat{u}\right]\right]. \tag{4.17}$$

THEOREM 4.4. *The amount of information loss of a first-order efficient sequential estimator is given by*

$$\Delta g_{ab} = \left({}'H^{(e)}_M\right)^2_{ab} + \tfrac{1}{2}\left(H^{(m)}_A\right)^2_{ab}. \tag{4.18}$$

The proof is straightforward and is omitted. It should be noted that the second-order term of the covariance of $\hat{u}^*$ is written as

$$\left({}'\Gamma^{(m)}_M\right)^{2ab} + \Delta g^{ab}.$$

It should be noted that, when $\nu(w)$ is fixed, the loss of information or the covariance of $\hat{u}^*$ is minimized when and only when $A(u)$ is m-flat, that is, $H_A^{(m)} = 0$, implying that the estimator belongs to $P_{M2}(S)$. It should be noted that $H_A^{(m)}$ does not depend on $\nu(w)$ or s_α. Therefore, given $\nu(w)$, an estimator whose estimating function belongs to $P_{M2}(s)$ is second-order efficient. The m.l.e. satisfies $H_A^{(m)} = 0$. We hereafter treat only estimators belonging to $P_{M2}(S)$.

The term $'\Gamma_M^{(m)}$ depends only on s_α. Hence, when $\nu(w)$ is fixed on M to be equal to a given function $\nu(u,0)$, $'\Gamma_M^{(m)}$ is also fixed. On the other hand, $'H_M^{(e)}$ depends only on s_κ. Therefore, the problem of designing a good sequential estimator is divided into two separate problems: one is to choose a stopping rule which is effective for reducing the $'H_M^{(e)}$ term and the other is to choose a function $\nu(u,0)$ effective for stabilizing the first-order covariance $\nu^{-1}(u)g^{ab}(u)$ of estimators. The second problem is how to choose $\nu(u)$ on M and the first one is how to extend it in S to give the entire $\nu(w)$. The first problem is solved in Section 5 and the second one is solved in Section 6.

5. Second-order efficiency of sequential procedures. Now we fix $\nu(w)$ on M, such that

$$E_u[N] = k\nu(u,0).$$

The properties of estimators belonging to $P_{M2}(S)$ depend on

$$s_\kappa = \partial_\kappa \log \nu(u,v)$$

at $v = 0$. This implies that they depend on how we extend $\nu(u,0)$ in a tubular neighborhood of M.

Let us define the mean e-curvature of M by

$$H_\kappa^{(e)} = \frac{1}{m} H_{ab\kappa}^{(e)} g^{ab} \tag{5.1}$$

and the scalar mean e-curvature by

$$\|H_M^{(e)}\| = \sqrt{H_\kappa^{(e)} H_\lambda^{(e)} g^{\kappa\lambda}}\,. \tag{5.2}$$

We further define the conformal e-embedding curvature of M in S by

$$\overline{H}_{ab\kappa}^{(e)} = H_{ab\kappa}^{(e)} - g_{ab} H_\kappa^{(e)}. \tag{5.3}$$

This is a conformal invariant, which does not change under any conformal transformations.

THEOREM 5.1. *The loss of information is minimized by the stopping rule satisfying*

$$s_\kappa = H_\kappa^{(e)}, \tag{5.4}$$

where the minimality is measured by the trace of Δg_{ab}. The minimized loss of *information is given by the square of the e-conformal curvature,*

$$\Delta g_{ab} = \left(\overline{H}^{e}_{M}\right)^2_{ab} + O(K^{-1}).$$

PROOF. We have

$$\begin{aligned}
\operatorname{tr}\left('H^{(e)}_{M}\right)^2_{ab} &= \left('H^{(e)}_{M}\right)^2_{ab} g^{ab} \\
&= \left(H^{(e)}_{M}\right)^2_{ab} g^{ab} + m\left\{\left(s_\kappa - H^{(e)}_\kappa\right)\left(s_\lambda - H^{(e)}_\lambda\right) - H^{(e)}_\kappa H^{(e)}_\lambda\right\} g^{\kappa\lambda} \\
&\geq \operatorname{tr}\left(H^{(e)}_{M}\right)^2_{ab} - m H^{(e)}_\kappa H^{(e)}_\lambda g^{\kappa\lambda} \\
&= \operatorname{tr}\left(H^{(e)}_{M}\right)^2_{ab} - m\|H^{(e)}_{M}\|^2.
\end{aligned} \tag{5.5}$$

The equality holds if and only if

$$s_\kappa = H^{(e)}_\kappa.$$

Hence, Δg_{ab} attains the minimum when

$$'H^{(e)}_{ab\kappa} = H^{(e)}_{ab\kappa} - g_{ab} H^{(e)}_\kappa = \overline{H}^{(e)}_{ab\kappa}. \qquad \square$$

It is significant to compare sequential procedures with nonsequential procedures. Since $(H^{(e)}_M)^2_{ab}$ is the least information loss of nonsequential estimators [Amari (1985)], $m\|H^{(e)}_M\|$ corresponds to the maximal attainable information recovered by a sequential estimation. When the e-mean curvature vector vanishes, a submanifold M is said to be e-minimal. There is no recovery of information by a sequential estimation when the submanifold M is e-minimal. On the other hand, when the conformal e-curvature vanishes, all the loss due to the e-curvature is recovered by the best sequential estimation.

A bias-corrected first-order efficient estimator is said to be *second-order efficient* when the second-order asymptotic covariance of the procedure attains the minimum. The following theorem is a direct consequence of Theorem 5.1.

THEOREM 5.2. *A sequential estimation procedure is second-order efficient if and only if its estimator and stopping rule are defined such that* $s_\kappa = H^{(e)}_\kappa$ *holds in* (3.2). *The asymptotic covariance of a second-order efficient estimator is given by*

$$\begin{aligned}
&E\left[K\nu(\hat{u}^{*a} - u^a)(\hat{u}^{*b} - u^b)\right] \\
&\quad = g^{ab} + \frac{1}{2K\nu}\left('\Gamma^{(m)}_{M}\right)^{2ab} + \frac{1}{K\nu}\left(\overline{H}^{(e)}_{M}\right)^{2ab} + O(K^{-2}).
\end{aligned} \tag{5.6}$$

The Bhattacharya-type bound for the second-order asymptotic covariances of sequential estimation procedures [Takeuchi and Akahira (1988)] is given by

$$g^{ab} + \frac{1}{K\nu}\left('\Gamma^{(m)}_{M}\right)^{2ab}. \tag{5.7}$$

A second-order efficient procedure attains the Bhattacharya-type bound when the conformal e-embedding curvature vanishes on M. In particular, when M

is a one-dimensional manifold, the Bhattacharya-type bound is surely attained by some second-order efficient procedure.

6. Covariance stabilization. The first-order term of the covariance matrix of an efficient estimator is given by N^{-1} times the inverse g^{ab} of the Fisher information matrix. When $g^{ab}(u)$ does not depend on u, in particular, when

$$g^{ab}(u) = \delta^{ab},$$

where δ^{ab} is the Kronecker delta, we say that the covariance of an estimator is stabilized. We say that a model M is covariance stabilizable when there exists a covariance stabilized parameterization.

The following theorem is known in the case of nonsequential estimation [Yoshizawa (1971), Amari (1985)].

THEOREM 6.1. *A model is covariance stabilizable, when and only when the 0-Riemann–Christoffel curvature (RC curvature) $R^{(0)}_{abcd}$ vanishes (see Appendix 1).*

In the case of sequential estimation, we have additional degrees of freedom of covariance stabilization; choosing $\nu(u,0)$ adequately as well as choosing a suitable parameterization.

When we have a function $\nu(u,0)$ on M and a parameterization $u^{a'}$ such that

$$\nu(u,0)g_{a'b'}(u) = \delta_{a'b'},$$

the covariance of an efficient estimator is uniformly stabilized by an efficient sequential estimation procedure. The following theorem comes from conformal geometry.

THEOREM 6.2. *A model M is covariance stabilizable, when and only when M is conformally 0-flat (see Appendix 1).*

Obviously, when the 0-RC curvature vanishes, the model is conformally 0-flat. It is known that the RC curvature vanishes for any one-dimensional model. It is known that any two-dimensional model is conformally 0-flat.

COROLLARY. *Any two-dimensional model is covariance stabilizable by an adequate sequential estimation procedure.*

Even when M is not conformally covariance stabilizable, it is always possible to choose $\nu(u)$ such that the trace or the determinant of νg_{ab} does not depend on u, so that the parameter becomes more stabilizable.

7. Conclusion. We have studied the asymptotic theory of sequential estimation procedures for curved exponential families. It has been shown that statistical manifolds are conformally changed according to the stopping rules in sequential estimation procedures. We have obtained the asymptotic covariance matrix and information loss of sequential estimation procedures up to the second order. We have also proposed second-order efficient procedures. Stopping rules and information losses of second-order efficient procedures are closely related to the geometrical structure of the underlying manifold.

Covariance stabilization is another effect which sequential estimation procedures enjoy. We have also shown the relation between the conformal change and covariance stabilization by the sequential estimation.

APPENDIX 1

Geometrical quantities. Given an exponential family $S = \{p(x, \theta)\}$, the basic geometrical quantities are defined by

$$g_{ij} = E\left[\partial_i l \, \partial_j l\right],$$

$$T_{ijk} = E\left[\partial_i l \, \partial_j l \, \partial_k l\right].$$

The α-connection is then defined by

$$\Gamma^{(\alpha)}_{ijk} = E\left[\partial_i \, \partial_j l \, \partial_k l\right] + \frac{1-\alpha}{2} T_{ijk}.$$

The α-RC curvature is given by

$$R^{(\alpha)}_{ijkl} = \partial_i \Gamma^{(\alpha)}_{jkl} - \partial_j \Gamma^{(\alpha)}_{ikl} + g^{st}\left(\Gamma^{(\alpha)}_{iks}\Gamma^{(\alpha)}_{jtl} - \Gamma^{(\alpha)}_{jks}\Gamma^{(\alpha)}_{itl}\right).$$

An (n, m)-curved exponential family M is given by

$$\theta = \theta(u), \qquad \eta = \eta(u),$$

in the θ- and η-coordinate systems, respectively. Let us divide S into a family of smooth $(n - m)$-dimensional submanifolds $A(u)$. When an $A(u)$ intersects M at $\theta(u)$, we have a w-coordinate system $w = (u, v)$ in the entire S, by introducing a coordinate system $v = (v^\kappa)$ in each $A(u)$. The basic tensors are written as

$$g_{\alpha\beta} = g_{ij} B^i_\alpha B^j_\beta,$$

$$T_{\alpha\beta\gamma} = T_{ijk} B^i_\alpha B^j_\beta B^k_\gamma,$$

in the w-coordinate system and the α-connection is given by

$$\Gamma^{(\alpha)}_{\beta\gamma\delta} = B^i_\beta B^j_\gamma B^k_\delta \Gamma^{(\alpha)}_{ijk} + g_{ij} B^i_\delta \, \partial_\beta B^j_\gamma,$$

where

$$B^i_\alpha = \partial\theta^i / \partial w^\alpha.$$

The tangent space of M is spanned by m tangent vectors,

$$\partial_a = B^i_a \, \partial_i, \qquad a = 1, 2, \ldots, m,$$

and the tangent vectors of $A(u)$ are spanned by $n - m$ tangent vectors

$$\partial_\lambda = B^i_\lambda \partial_i, \qquad \lambda = m + 1, \dots, n.$$

The submanifolds M and $A(u)$ are said to be orthogonal when the tangent vectors are mutually orthogonal,

$$\langle \partial_a, \partial_\lambda \rangle = g_{a\lambda} = g_{ij} B^i_a B^j_\lambda = 0.$$

It should be noted that

$$B_{\alpha i} = g_{ij} B^j_\alpha = \partial \eta_i / \partial w^\alpha$$

holds, if we use the η-coordinate system.

The m-connection of M is given by

$$\Gamma^{(m)}_{abc} = B^i_a (\partial_b B_{ci}).$$

The (embedding) e-curvature of M is given by

$$H^{(e)}_{ab\kappa} = B_{\kappa i} \partial_a B^i_b,$$

and the (embedding) m-curvature of $A(u)$ is given by

$$H^{(m)}_{\kappa\lambda a} = B^i_a \partial_\kappa B_{\lambda i}.$$

When this vanishes at $v = 0$, $A(u)$ is said to be locally m-flat.

The square of these quantities are given by

$$\left(\Gamma^{(m)}_M\right)^2_{ab} = \Gamma^{(m)}_{cda} \Gamma^{(m)}_{efb} g^{ce} g^{df},$$

$$\left(H^{(e)}_M\right)^2_{ab} = H^{(e)}_{ac\kappa} H^{(e)}_{bd\lambda} g^{cd} g^{\kappa\lambda},$$

$$\left(H^{(m)}_A\right)^2_{ab} = H^{(e)}_{\kappa\lambda a} H^{(e)}_{\nu\mu b} g^{\kappa\nu} g^{\lambda\mu},$$

where

$$g_{ab} = B_{ai} B_{bj} g^{ij}, \qquad g_{\kappa\lambda} = B_{\kappa i} B_{\lambda j} g^{ij},$$

and all indices can be lowered or uppered by using these metric tensors or their inverse, for example,

$$\left(\Gamma^{(m)}_M\right)^{2ab} = g^{ac} g^{bd} \left(\Gamma^{(m)}_M\right)^2_{cd}.$$

Suppose that two statistical manifolds (S, g, T) and $('S, 'g, 'T)$ are diffeomorphic and that their fundamental quantities are related by

$$'g_{\alpha\beta} = g_{\alpha\beta},$$

$$'T_{\alpha\beta\gamma} = T_{\alpha\beta\gamma} + 3 g_{(\alpha\beta} s_{\gamma)},$$

where

$$s_\gamma = \partial_\gamma \log \nu(w).$$

The diffeomorphism from (S, g, T) to $('S, 'g, 'T)$ is then said to be a conformal transformation of the statistical manifold induced by the gauge or expansion factor $\nu(w)$. The conformal change of a Riemannian manifold implies that the manifold is expanded or contracted isotropically but that an expansion rate depends on each point [Schouten (1954)]. Our transformation is a statistical counterpart of this change. A conformal transformation changes the α-connection into

$$'\Gamma^{(\alpha)}_{\beta\gamma\delta} = \Gamma^{(\alpha)}_{\beta\gamma\delta} + \frac{3(1-\alpha)}{2} g_{(\beta\gamma} s_{\delta)} - g_{\beta\gamma} s_{\delta}.$$

Accordingly, the m-connection of M, the e-curvature of M and the m-curvature of $A(u)$ are changed into

$$'\Gamma^{(m)}_{abc} = \Gamma^{(m)}_{abc} + g_{ca} s_b + g_{cb} s_a,$$

$$'H^{(e)}_{ab\kappa} = H^{(e)}_{ab\kappa} - g_{ab} s_{\kappa},$$

$$'H^{(m)}_{\kappa\lambda a} = H^{(m)}_{\kappa\lambda a}.$$

The square of these quantities are calculated as before.

The Weyl–Schouten curvature (0-WS curvature) of M with $\dim M \geq 4$ is given by

$$C^{(0)d}_{abc} = C^{(0)d}_{abc} - \frac{1}{m-2}\left(\delta^d_b R^{(0)}_{ca} - \delta^d_c R^{(0)}_{ba} + R^{(0)d}_{b} g_{ca} - R^{(0)d}_{c} g_{ba}\right) + \frac{1}{(m-1)(m-2)}\left(\delta^d_b g_{ca} - \delta^d_c g_{ba}\right),$$

where

$$R^{(0)}_{ab} = R^{(0)}_{cabd} g^{cd} \quad \text{and} \quad m = \dim M.$$

When $\dim M = 3$, the 0-WS curvature of M is given by

$$C^{(0)}_{abc} = \left(\nabla^{(0)}_b R^{(0)}_{ca} - \nabla^{(0)}_c R^{(0)}_{ba}\right) - \tfrac{1}{4}\left(g_{ab}\, \partial_c R^{(0)} - g_{ac}\, \partial_b R^{(0)}\right),$$

where

$$R^{(0)} = R^{(0)}_{ab} g^{ab} \quad \text{and} \quad \nabla^{(0)}_b R^{(0)}_{ca} = \partial_b R^{(0)}_{ca} - \Gamma^{(0)d}_{bc} R^{(0)}_{da} - \Gamma^{(0)d}_{ba} R^{(0)}_{cd}.$$

A manifold M is said to be conformally equivalent to a manifold M', when there exists a conformal transformation of M onto M'. A manifold which is conformally equivalent to a Euclidean space is said be conformally 0-flat. The following theorem is well known in conformal geometry.

THEOREM [SCHOUTEN (1954)].

(i) *If* $\dim M \leq 2$, *then* M *is conformally* 0-*flat.*

(ii) *If* $\dim M = 3$, *then* M *is conformally* 0-*flat when and only when the* 0-WS *curvature* $C^{(0)}_{abc}$ *vanishes.*

(iii) *If* $\dim M \geq 4$, *then* M *is conformally* 0-*flat when and only when the* 0-WS *curvature* $C^{(0)}_{abcd}$ *vanishes.*

It is noted that the statistical model M is covariance stabilizable if and only if M is conformally 0-flat.

APPENDIX 2

The following lemmas are due to Takeuchi and Akahira (1988). These are useful for the calculation of geometrical quantities and asymptotic covariances.

LEMMA 1. *Suppose that T is a function of observed samples* $\mathbf{x}$ *and of he true parameter* θ. *The sample size is assumed to be determined according to some stopping rule and to have a finite expectation. Let* $l(\mathbf{x}, \theta)$ *denote the* log-*likelihood of* $\mathbf{x}$. *Then we have*

$$E_\theta[T\,\partial_i l(\mathbf{x}, \theta)] = \partial_i E_\theta[T] - E_\theta[\partial_i T].$$

LEMMA 2. *Suppose that the estimator* $\hat{u}$ *of the procedure satisfies the following equations*:

$$\sqrt{K}\,\nu(\hat{u}^a - u^a) = \frac{1}{\sqrt{K}\,\nu}\{g^{ab}\,\partial_b l(\mathbf{x}, \theta) + Q^a\} + O(K^{-1})$$

and

$$E\left[\sqrt{K}\,\nu(\hat{u}^a - u^a)\right] = O(K^{-1}).$$

Then we have

$$E[K\nu(\hat{u}^a - u^a)(\hat{u}^b - u^b)] = g^{ab} + \frac{1}{K\nu}\operatorname{Cov}[Q^a, Q^b] + o(K^{-1}),$$

where

$$\operatorname{Cov}[f, g] = E[fg] - E[f]E[g].$$

PROOF. Let

$$\tilde{u}^a = \sqrt{K}\,v(\hat{u}^a - u^a),$$

then

$$\begin{aligned} E[\tilde{u}^a\tilde{u}^b] &= E\left[\left(\tilde{u}^a - \frac{1}{\sqrt{K}\,\nu}g^{ac}\,\partial l_c(\mathbf{x}, u)\right)\left(\tilde{u}^b - \frac{1}{\sqrt{K}\,\nu}g^{bd}\,\partial l_d(\mathbf{x}, u)\right)\right] \\ &\quad + \frac{2}{\sqrt{K}\,\nu}E[\tilde{u}^{(a}g^{b)c}\,\partial_c l(\mathbf{x}, u)] - \frac{1}{K\nu}E[g^{ac}g^{bd}\,\partial_c l(\mathbf{x}, u)\,\partial_d l(\mathbf{x}, u)] \\ &= \frac{1}{K\nu}E[Q^a, Q^b] + \frac{1}{\sqrt{K}\,\nu}g^{ac}E[\tilde{u}^b\,\partial_c l(\mathbf{x}, u)] \\ &\quad + \frac{1}{\sqrt{K}\,\nu}g^{bc}E[\tilde{u}^a\,\partial_c l(\mathbf{x}, u)] - g^{ab}. \end{aligned}$$

Because of Lemma 1,

$$\frac{1}{K\nu} E[\tilde{u}^a \partial_b l(\mathbf{x}, u)] = \delta_b^a + O(K^{-2}),$$

where δ_b^a implies Kronecker's delta. Hence, we complete the proof. □

REFERENCES

AKAHIRA, M. and TAKEUCHI, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency. Lecture Notes in Statist.* **7**. Springer, New York.

AKAHIRA, M. and TAKEUCHI, K. (1989). Third order asymptotic efficiency of the sequential maximum likelihood estimation procedure. *Sequential Anal.* **8** 333–359.

AMARI, S. (1985). *Differential Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, New York.

AMARI, S. (1987a). Differential geometrical theory of statistics I. In *Differential Geometry in Statistical Inference* (S. I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao, eds.) 19–94. IMS, Hayward, Calif.

AMARI, S. (1987b). Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Math. Systems Theory* **20** 53–82.

AMARI, S. (1989). Fisher information under restriction of Shannon information in multi-terminal situations. *Ann. Inst. Statist. Math.* **41** 623–648.

AMARI, S. and HAN, T. S. (1989). Statistical inference under multiterminal rate restrictions: A differential geometrical approach. *IEEE Trans. Inform. Theory* **35** 217–227.

AMARI, S. I., BARNDORFF-NIELSEN, O. E., KASS, R. E. LAURITZEN, S. L. and RAO, C. R. (1987). *Differential Geometry in Statistical Inference*. IMS, Hayward, Calif.

BARNDORFF-NIELSEN, O. (1980). Conditionality resolutions. *Biometrika* **67** 293–310.

BARNDORFF-NIELSEN, O. (1988). *Parametric Statistical Models and Likelihood. Lecture Notes in Statist.* **50**. Springer, New York.

BARNDORFF-NIELSEN, O., COX, D. R. and REID, N. (1986). The role of differential geometry in statistical theory. *Internat. Statist. Rev.* **54** 83–96.

BICKEL, P. J., CHIBISOV, D. M. and VAN ZWET, W. R. (1981). On efficiency of first and second order. *Internat. Statist. Rev.* **49** 169–175.

CHIBISOV, D. M. (1974). Asymptotic expansion for some asymptotically optimal tests. In *Proc. Prague Symp. Asymptotic Statist.* (J. Hajek, ed.) **2** 37–68. Univ. Karlova, Prague.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

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

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

GHOSH, J. K. and SUBRAMANYAM, K. (1974). Second order efficiency of the maximum likelihood estimator. *Sankhyā Ser. A* **36** 325–358.

GREENWOOD, P. E. and SHIRYAEV, A. N. (1985). On uniform weak convergence of semimartingales with applications to estimation of the parameter of an autoregressive model of first order. In *Statistics and Control of Stochastic Processes* (N. V. Krylov, R. Sh. Lipster and A. A. Novikov, eds.) 40–48. Optimization Software, New York.

GREENWOOD, P. E. and SHIRYAEV, A. N. (1988). Sequential estimation for first order autoregressive models. Dept. Mathematics, Univ. London.

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4** 188–234.

LAURITZEN, S. L. (1987). Statistical manifolds. In *Differential Geometry in Statistical Inference* (S. I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao, eds.) 163–216. IMS, Hayward, Calif.

NAGAOKA, I. and AMARI, S. (1982). Differential geometry of smooth families of probability distributions, *METR* 82-7, Univ. Tokyo.

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, New York.

RAO, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **24** 46–72.

SCHOUTEN, J. A. (1954). *Ricci-Calculus: An Introduction to Tensor Analysis and Its Geometrical Applications*, 2nd ed. Springer, Berlin.

SØRENSEN, M. (1986). On sequential maximal likelihood estimation for exponential families of stochastic processes. *Internat. Statist. Rev.* **54** 191–210.

TAKEUCHI, K. and AKAHIRA, M. (1988). Second order asymptotic efficiency in terms of asymptotic variances of the sequential maximum likelihood estimation, procedures. In *Statistical Theory and Data Analysis II: Proceedings of the Second Pacific Area Statistical Conference* (K. Matusita, ed.) 191–196. North-Holland, Amsterdam.

VOS, P. W. (1989). Fundamental equations for statistical submanifolds with applications to the Bartlett correction. *Ann. Inst. Statist. Math.* **41** 429–450.

YOSHIZAWA, T. (1971). A geometry of parameter space and its statistical interpretation. Memo TYH-2, Harvard Univ.

FACULTY OF ENGINEERING
UNIVERSITY OF TOKYO
BUNKYO-KU
TOKYO 113
JAPAN