The Annals of Statistics
1986, Vol. 14, No. 3, 856–873

LIKELIHOOD AND OBSERVED GEOMETRIES

BY O. E. BARNDORFF-NIELSEN

Aarhus University

In the differential geometric approach to parametric statistics, developed by Chentsov, Efron, Amari, and others, the parameter space is set up as a differentiable manifold with expected information as metric tensor and with a family of affine connections, the α-connections, determined from the expected information and the skewness tensor of the score vector. The usefulness of this approach is particularly notable in connection with Edgeworth expansions of estimators. Motivated by the conditionality viewpoint, an "observed" parallel to that theory is established in the present paper using observed information and an "observed skewness" tensor instead of the above expected quantities. The formula $c|\hat{j}|^{1/2}\bar{L}$ for the conditional distribution of the maximum likelihood estimator is expanded (to third order) asymptotically and the "observed geometries" are shown to have a role in this type of expansion similar to that of the "expected geometries" in the Edgeworth expansions mentioned above. In these new developments "mixed derivatives of the log model function," defined by means of an auxiliary statistic complementing the maximum likelihood estimator, take the place of moments of derivatives of the log likelihood function.

1. Introduction. A number of recent investigations have shown that in the study of inference for parametric statistical models, particularly as regards higher-order asymptotics, it is useful and illuminating to set the model, $\mathscr{M}$ say, up as a differentiable manifold equipped with a Riemannian metric and a family of affine connections, the so-called α-connections. In that approach, the parameter space of the model serves for the coordinate representation of $\mathscr{M}$, the metric tensor employed is the expected information matrix

$$i_{rs} = -E\{\partial_r \partial_s l\} \tag{1.1}$$

and the family of α-connections is determined by (1.1) and the so-called skewness tensor

$$T_{rst} = E\{\partial_r l\, \partial_s l\, \partial_t l\} \tag{1.2}$$

which is a covariant tensor of rank 3. Here l denotes the log likelihood function of the model, and with ω, of dimension d, as the parameter of the model, we write $\omega = (\omega^1, \dots, \omega^d)$ and $\partial_r = \partial/\partial\omega^r$. The indices $r, s, t, \dots$ run over $1, 2, \dots, d$. In this framework it is, for instance, possible to give geometrical interpretations to various of the terms arising in conditional and unconditional Edgeworth expansions for the distribution of the maximum likelihood estimator $\hat{\omega}$ under curved

Received January 1985; revised September 1985.
AMS 1980 *subject classifications*. Primary 62F99; secondary 62A10.
Key words and phrases. Alpha connections, ancillarity, asymptotic expansion, Bartlett factors, conditionality, constant normal fractile, covariant differentiation, exponential model, hyperboloid model, inverse Gaussian–Gaussian model, location–scale model, maximum likelihood, observed information, tensors.

exponential models. For these theoretical developments see Efron (1975), Amari (1982a, b, 1983, 1984, 1985), and Amari and Kumon (1983) and the references given there.

For many purposes the observed information matrix j, i.e.,

$$j_{rs} = -\partial_r \partial_s l,$$

is more natural to work with than the expected information i, and it therefore seemed of interest to enquire whether the model $\mathscr{M}$ can also be rigged with some kind of "observed geometrical structures" paralleling the "expected geometrical structures" given by (i, T) as defined by (1.1) and (1.2). We shall show that this is indeed the case and that the resulting geometries are intimately connected with a certain type of asymptotic expansion deriving from the formula $c|\hat{j}|^{1/2}\bar{L}$ [Barndorff-Nielsen (1980, 1983)] for the conditional distribution of the maximum likelihood estimator.

These new types of statistical geometries and expansions notably do not involve integrations over the sample space, as is required in (1.1) and (1.2) and in the calculation of the cumulants that occur in the Edgeworth expansions. Instead they employ what may be referred to as mixed derivatives of the log model function.

Furthermore, whereas the studies of expected geometries have been largely concerned with curved exponential families, the approach taken here makes it equally natural to consider other parametric models, and in particular transformation models.

The viewpoint of conditional inference has been instrumental for the constructions in question. However, the observed geometrical calculus, as discussed below, does not presuppose the existence of exact or approximate ancillaries but only operates with an auxiliary statistic a complementing the maximum likelihood estimate $\hat{\omega}$. Only when it comes to applications to problems of inference does distribution constancy—and hence ancillarity—of a become essential.

Let the model $\mathscr{M}$ be given by $(\mathscr{X}, p(x; \omega), \Omega)$ where $\mathscr{X}$ is the sample space, Ω is the parameter space, and $p(x; \omega)$ is the model function, i.e., for a given value of the parameter ω the function $p(x; \omega)$ is the probability density function of the observation $x \in \mathscr{X}$ relative to a fixed dominating measure μ on $\mathscr{X}$. Suppose the minimal sufficient statistic t for $\mathscr{M}$ is of dimension k. We then speak of $\mathscr{M}$ as a (k, d)-model (d being the dimension of the parameter ω). Let $(\hat{\omega}, a)$ be a one-to-one transformation of t, where $\hat{\omega}$ is the maximum likelihood estimator of ω and a, of dimension $k - d$, is an auxiliary statistic.

In most applications it will be essential to construct a so as to be distribution constant either exactly or to the relevant asymptotic order. And then, according to the conditionality principle the conditional model for $\hat{\omega}$ given a is considered the appropriate basis for inference on ω.

However, distribution constancy of a is not assumed in the construction of the observed geometries.

There will be no loss of generality in viewing the log likelihood $l = l(\omega)$ in its dependence on the observation x as being a function of the minimal sufficient $(\hat{\omega}, a)$ only. Henceforth we shall think of l in this manner and we will indicate this by writing $l = l(\omega; \hat{\omega}, a)$. Similarly, in the case of observed information we write $j = j(\omega; \hat{\omega}, a)$, etc. We may now take partial derivatives of l with respect to the coordinates $\hat{\omega}^r$ of $\hat{\omega}$ as well as with respect to ω^r. Letting $\hat{\partial} = \partial/\partial\hat{\omega}^r$ we introduce the notation

$$l_{r_1 \ldots r_p; s_1 \ldots s_q} = \partial_{r_1} \ldots \partial_{r_p} \hat{\partial}_{s_1} \ldots \hat{\partial}_{s_q} l \tag{1.3}$$

and refer to these quantities as mixed derivatives of the log model function. The function of ω and a obtained from (1.3) by substituting ω for $\hat{\omega}$ will be denoted by $\not{l}_{r_1 \ldots r_p; s_1 \ldots s_q}$. Thus, for instance,

$$\not{l}_{rs;t} = \not{l}_{rs;t}(\omega) = \not{l}_{rs;t}(\omega; a) = l_{rs;t}(\omega; \omega, a).$$

Similarly,

$$\not{j} = \not{j}(\omega) = \not{j}(\omega; a) = j(\omega; \omega, a).$$

The observed geometries, which will be introduced and illustrated in Section 2, are expressed in terms of the mixed derivatives

$$\not{l}_{r_1 \ldots r_p; s_1 \ldots s_q}. \tag{1.4}$$

So are the terms of an asymptotic expansion of

$$p^*(\hat{\omega}; \omega|a) = c|\hat{j}|^{1/2}\bar{L}, \tag{1.5}$$

to be derived in Section 3. In (1.5) $\bar{L}$ denotes the normed likelihood function, i.e.,

$$\bar{L} = e^{l - \hat{l}},$$

$|\hat{j}|$ is the determinant of the observed information, and $c = c(\omega, a)$ is a norming constant determined so as to make the integral of $p^*(\hat{\omega}; \omega|a)$ with respect to $\hat{\omega}$ for fixed a equal to 1.

For a ancillary the model function p^*, given by (1.5), may be considered as an approximation to the actual model function $p(\hat{\omega}; \omega|a)$ for the maximum likelihood estimator $\hat{\omega}$ conditional on a. As such it is, in wide generality, correct to order $O(n^{-3/2})$ at least, under repeated sampling with n denoting sample size. In fact, $p^*(\hat{\omega}; \omega|a)$ equals $p(\hat{\omega}; \omega|a)$ exactly for a considerable range of models, including all transformation models, cf. Barndorff-Nielsen (1980, 1983, 1984b) and Barndorff-Nielsen and Blæsild (1984). Some further discussion and applications of (1.5) may be found in Barndorff-Nielsen and Cox (1984a, b), Barndorff-Nielsen (1984a, 1985a, b) and McCullagh (1984a). In particular, in Barndorff-Nielsen and Cox (1984a) a simple relation is established between the norming constant c of (1.5) and the Bartlett adjustment factors for log likelihood ratio tests of hypotheses about ω. [See also Barndorff-Nielsen and Cox (1984b).] We comment on this relation in Section 3.

Besides being in a certain sense "closer to the actual data at hand," the "observed" quantities and formulas are in various respects simpler to work with than their expected counterparts. For instance, in certain cases Bartlett adjustment factors are more readily calculable in terms of the observed quantities. Another example is provided by formula (3.15), cf. the discussion following that formula.

Some connections between expected and observed geometries and profile likelihood, L-sufficiency, marginal likelihood, and transformation models have been studied in Barndorff-Nielsen and Jupp (1984, 1985).

2. Observed geometries. We shall be interested in how various quantities behave under reparametrizations of the model $\mathscr{M}$. Let ψ, of dimension d, be the parameter of some parametrization of $\mathscr{M}$, alternative to that indicated by ω. Coordinates of ψ will be denoted by ψ^a, ψ^b, etc. and we write ∂_a for $\partial/\partial\psi^a$ and $\omega^i_{/a}$ for $\partial\omega^i/\partial\psi^a$, $\omega^i_{/ab}$ for $\partial^2\omega^i/\partial\psi^a\,\partial\psi^b$, etc. Furthermore, we write $l(\psi)$ for the log likelihood under the parametrization by ψ, though formally this is in conflict with the notation $l(\omega)$, and correspondingly we let $l_a = \partial_a l = \partial_a l(\psi)$, etc.; similarly for other parameter-dependent quantities. Finally, the symbol ^ over such a quantity indicates that the maximum likelihood estimate has been substituted for the parameter.

Using this notation and that established in Section 1, and adopting the summation convention that if a suffix occurs repeatedly in a single expression then summation over that suffix is understood, we have

$$l_a = l_r\omega^r_{/a}, \tag{2.1}$$

$$l_{ab} = l_{rs}\omega^r_{/a}\omega^s_{/b} + l_r\omega^r_{/ab}, \tag{2.2}$$

$$l_{abc} = l_{rst}\omega^r_{/a}\omega^s_{/b}\omega^t_{/c} + l_{rs}\omega^r_{/ab}\omega^s_{/c}[3] + l_r\omega^r_{/abc}, \tag{2.3}$$

etc., where [3] signifies a sum of three similar terms determined by permutation of the indices a, b, c. On substituting $\hat\omega$ for ω in (2.2) we obtain the well-known relation

$$\hat{j}_{ab} = \hat{j}_{rs}\hat\omega^r_{/a}\hat\omega^s_{/b}$$

which, now by substitution of ω for $\hat\omega$, may be reexpressed as

$$\mathscr{j}_{ab} = \mathscr{j}_{rs}\omega^r_{/a}\omega^s_{/b} \tag{2.4}$$

or, written more explicitly,

$$\mathscr{j}_{ab}(\psi; a) = \mathscr{j}_{rs}(\omega; a)\frac{\partial\omega^r}{\partial\psi^a}\frac{\partial\omega^s}{\partial\psi^b}.$$

Equation (2.4) shows that $\mathscr{j}$ is a metric tensor on $\mathscr{M}$, for any given value of the auxiliary statistic a. Moreover, in wide generality $\mathscr{j}$ will be positive definite on $\mathscr{M}$, and we assume henceforth that this is the case. In fact, for any $\hat\omega \in \Omega$ we have $\hat{\mathscr{j}} = \hat{j}$, i.e., observed information at the maximum likelihood point, which is generally positive definite (though counterexamples do exist).

Equipped with $\mathscr{j}$ as metric tensor $\mathscr{M}$ becomes a Riemannian manifold. Notice however that this Riemannian geometry depends on the value of the auxiliary a. We call $\mathscr{j}$ the observed metric on $\mathscr{M}$.

The Riemannian connection determined by $\mathscr{j}$ has connection symbols $\overset{0}{\mathscr{f}}{}^t_{rs}$ given by $\overset{0}{\mathscr{f}}{}^t_{rs} = \mathscr{j}^{tu}\overset{0}{\mathscr{f}}_{rsu}$ and

$$\overset{0}{\mathscr{f}}_{rst} = \tfrac{1}{2}(\partial_r\mathscr{j}_{st} - \partial_t\mathscr{j}_{rs} + \partial_s\mathscr{j}_{tr}).$$

Employing the notation established in Section 1 we have $\partial_t j_{rs} = -\partial_t \ell_{rs} = -\ell_{rst} - \ell_{rs;t}$, etc., so that

$$\overset{0}{\Gamma}_{rst} = \ell_{rs;t} - \tfrac{1}{2}(\ell_{rst} + \ell_{rs;t}[3]). \tag{2.5}$$

As we shall now show, the quantity

$$T_{rst} = -(\ell_{rst} + \ell_{rs;t}[3]) \tag{2.6}$$

is a covariant tensor of rank 3, i.e.,

$$T_{abc} = T_{rst}\omega^r_{/a}\omega^s_{/b}\omega^t_{/c}. \tag{2.7}$$

First, from (2.3) we have

$$\ell_{abc} = \ell_{rst}\omega^r_{/a}\omega^s_{/b}\omega^t_{/c} + \ell_{rs}\omega^r_{/ab}\omega^s_{/c}[3]. \tag{2.8}$$

Further, from (2.2) we obtain, on differentiating with respect to $\hat{\psi}^c$ and then substituting parameter for estimate,

$$\ell_{ab;c} = \ell_{rs;t}\omega^r_{/a}\omega^s_{/b}\omega^t_{/c} + \ell_{r;t}\omega^r_{/ab}\omega^t_{/c}. \tag{2.9}$$

Finally, differentiating the likelihood equation

$$\ell_r = 0$$

we find

$$\ell_{rs} + \ell_{r;s} = 0, \tag{2.10}$$

or

$$\ell_{r;s} = j_{rs}. \tag{2.11}$$

Combination of (2.6), (2.8), (2.9), and (2.11) yields (2.7).

It follows from the tensorial nature of T and from (2.5) and (2.11) that for any real α an affine connection $\overset{\alpha}{\Gamma}$ on $\mathscr{M}$ may be defined by

$$\overset{\alpha}{\Gamma}{}^t_{rs} = j^{tu}\overset{\alpha}{\Gamma}_{rsu}$$

with

$$\overset{\alpha}{\Gamma}_{rst} = \ell_{rs;t} + \frac{1-\alpha}{2}T_{rst}. \tag{2.12}$$

In particular, we have

$$\overset{1}{\Gamma}_{rst} = \ell_{rs;t}, \qquad \overset{-1}{\Gamma}_{rst} = \ell_{t;rs},$$

where to obtain the latter expression we have used

$$\ell_{rst} + \ell_{rs;t} + \ell_{rt;s} + \ell_{r;st} = 0$$

which follows on differentiation of (2.10). It may also be noted that

$$\partial_t j_{rs} = \overset{1}{\Gamma}_{rts} + \overset{-1}{\Gamma}_{str} = \overset{1}{\Gamma}_{str} + \overset{-1}{\Gamma}_{rts}$$

and

$$\overset{\alpha}{\not{\Gamma}}_{rst} = \frac{1+\alpha}{2}\overset{1}{\not{\Gamma}}_{rst} + \frac{1-\alpha}{2}\overset{-1}{\not{\Gamma}}_{rst}.$$

The connections $\overset{\alpha}{\not{\Gamma}}$, which we shall refer to as the observed α-connections, are analogues of the expected α-connections $\overset{\alpha}{\Gamma}$ of Chentsov (1972) and Amari (1982a), which are given by

$$\overset{\alpha}{\Gamma}{}^{t}_{rs} = i^{tu}\overset{\alpha}{\Gamma}_{rsu}$$

and

$$\overset{\alpha}{\Gamma}_{rst} = E(l_{rs}l_t) + \frac{1-\alpha}{2}T_{rst},$$

where T is the skewness tensor (1.2). The analogy between $\overset{\alpha}{\Gamma}$ and $\overset{\alpha}{\not{\Gamma}}$ becomes more apparent by rewriting T_{rst} as

$$T_{rst} = -E\{l_{rst} + l_{rs}l_t[3]\},$$

the validity of which follows on differentiation of the formula

$$E\{l_{rs} + l_r l_s\} = 0, \tag{2.13}$$

which, in turn, may be compared to (2.10).

Under the specifications of a of primary statistical interest, one has that in broad generality the observed geometries converge to the corresponding expected geometries as the sample size tends to infinity.

For (k, k) exponential models

$$p(x;\theta) = a(\theta)b(x)e^{\theta\cdot t(x)}, \tag{2.14}$$

we have $\not{j} = i$ and $\overset{\alpha}{\not{\Gamma}} = \overset{\alpha}{\Gamma}$, $\alpha \in R$. More generally, for a curved subfamily of (2.14), given by restricting θ to be of the form $\theta = \theta(\omega)$ where the dimension d of the parameter ω is less than k, the quantities $\not{j}$ and $\not{T}$ possess, under mild regularity conditions, asymptotic expansions the first terms of which are given by

$$\not{j}_{rs} = i_{rs} - \theta^i_{/rs}\theta^j_{/\lambda}\kappa_{ij}a^\lambda + \cdots \tag{2.15}$$

and

$$\not{T}_{rst} = T_{rst} - \{\kappa_{ijk}\theta^i_{/rs}\theta^j_{/t}\theta^k_{/\lambda}[3] + \kappa_{ij}\theta^i_{/rs}\theta^j_{/t\lambda}[3] + \kappa_{ij}\theta^i_{/rst}\theta^j_{/\lambda}\}a^\lambda + \cdots. \tag{2.16}$$

Here suffices i and j run from 1 to k, θ^i, θ^j denote coordinates of θ, $\kappa_{ij} = \partial_i\,\partial_j\kappa$, where $\kappa = \kappa(\theta) = -\log a(\theta)$ is the cumulant transform of t, and $\partial_i = \partial/\partial\theta^i$, and a^λ, $\lambda = 1,\dots,k-d$, are coordinates of an ancillary complement of $\hat{\omega}$. For instance, in the repeated sampling situation and letting a_0 denote the affine ancillary, as defined in Barndorff-Nielsen (1980), we may take $a = n^{-1/2}a_0$ and the expansions (2.15) and (2.16) are asymptotic in powers of $n^{-1/2}$. [For further comparison with Amari (1982a) it may be noted that the coefficient in the first-order correction term of (2.15) may be written as $\theta^{i}_{/rs}\theta^{j}_{/\lambda}\kappa_{ij} = n\overset{e}{H}_{rs\lambda}$ where $\overset{e}{H}_{rs\lambda}$ is Amari's notation for the exponential curvature, or α-curvature with $\alpha = 1$, of the curved exponential model viewed as a manifold imbedded in the full (k, k) model.]

We now briefly consider four examples. In the first three the model is transformational and the auxiliary statistic a is taken to be the maximal invariant statistic, and thus a is exactly ancillary. In the fourth example a is only approximately ancillary. Examples 2.1, 2.3, and 2.4 are concerned with curved exponential models whereas the model in Example 2.2—the location-scale model—is exponential only if the error distribution is normal.

EXAMPLE 2.1. *Constant normal fractile.* For known $\alpha \in (0,1)$ and $c \in (-\infty, \infty)$, let $\mathscr{N}_{\alpha, c}$ denote the class of normal distributions having the real number c as α-fractile, i.e.,

$$\mathscr{N}_{\alpha, c} = \{N(\mu, \sigma^2): (c - \mu)/\sigma = u_\alpha\},$$

where u_α denotes the α-fractile of the standard normal distribution, and let $x_1, \dots, x_n$ be a sample from a distribution in $\mathscr{N}_{\alpha, c}$. The model for $x = (x_1, \dots, x_n)$ thus defined is a $(2,1)$ exponential model, except for $u_\alpha = 0$ when it is a $(1,1)$ model. Henceforth we suppose that $u_\alpha \neq 0$, i.e., $\alpha \neq \frac{1}{2}$. The model is also a transformation model relative to the subgroup G of the group of one-dimensional affine transformations given by

$$G = \{[c(1 - \lambda), \lambda]: \lambda > 0\},$$

the group operation being

$$[c(1 - \lambda), \lambda][c(1 - \lambda'), \lambda'] = [c(1 - \lambda\lambda'), \lambda\lambda']$$

and the action of G on the sample space being

$$[c(1 - \lambda), \lambda](x_1, \dots, x_n) = (c(1 - \lambda) + \lambda x_1, \dots, c(1 - \lambda) + \lambda x_n).$$

(Note that G is isomorphic to the multiplicative group.)

Letting

$$a = (\bar{x} - c)/s',$$

where $\bar{x} = (x_1 + \cdots + x_n)/n$ and

$$s'^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

we have that a is maximal invariant and, parametrizing the model by $\zeta = \log \sigma$, that the maximum likelihood estimate is

$$\hat{\zeta} = \log(bs'),$$

where

$$b = b(a) = (u_\alpha/2)a + \sqrt{1 + \{(u_\alpha/2)^2 + 1\}a^2}.$$

Furthermore, $(\hat{\zeta}, a)$ is a one-to-one transformation of the minimal sufficient statistic $(\bar{x}, s')$ and a is exactly ancillary.

The log likelihood function may be written as

$$l(\zeta) = l(\zeta; \hat{\zeta}, a) = n\left[\hat{\zeta} - \zeta - \tfrac{1}{2}\left\{b^{-2}e^{2(\hat{\zeta}-\zeta)} + \left(u_\alpha + ab^{-1}e^{\hat{\zeta}-\zeta}\right)^2\right\}\right],$$

from which it is evident that the model for $\hat{\zeta}$ given a is a location model.

Indicating differentiation with respect to ζ and $\hat{\zeta}$ by subscripts ζ and $\hat{\zeta}$, respectively, we find

$$l_\zeta = n\left\{-1 + b^{-2}e^{2(\hat{\zeta}-\zeta)} + ab^{-1}\left(u_\alpha + ab^{-1}e^{\hat{\zeta}-\zeta}\right)e^{\hat{\zeta}-\zeta}\right\},$$

and hence

$$\begin{aligned}
\mathit{j} &= n\left\{2b^{-2} + ab^{-1}\left(u_\alpha + 2ab^{-1}\right)\right\},\\
\mathit{l}_{\zeta\zeta\zeta} &= n\left\{4b^{-2} + ab^{-1}\left(u_\alpha + 4ab^{-1}\right)\right\},\\
\mathit{l}_{\zeta\zeta;\hat{\zeta}} &= -n\left\{4b^{-2} + ab^{-1}\left(u_\alpha + 4ab^{-1}\right)\right\} = \overset{1}{\mathit{\Gamma}},\\
\mathit{l}_{\zeta;\hat{\zeta}\hat{\zeta}} &= n\left\{4b^{-2} + ab^{-1}\left(u_\alpha + 4ab^{-1}\right)\right\} = \overset{-1}{\mathit{\Gamma}} = -\overset{1}{\mathit{\Gamma}},
\end{aligned}$$

and the observed skewness tensor is

$$\mathit{T} = n\left\{8b^{-2} + 2ab^{-1}\left(u_\alpha + 4ab^{-1}\right)\right\}.$$

Note also that

$$\overset{\alpha}{\mathit{\Gamma}} = \alpha\overset{1}{\mathit{\Gamma}}.$$

We mention in passing that another normal submodel, that specified by a known coefficient of variation μ/σ, has properties similar to those exhibited by Example 2.1.

EXAMPLE 2.2. *Location-scale model.* Let data x consist of a sample $x_1, \ldots, x_n$ from a location-scale model, i.e., the model function is

$$p(x; \mu, \sigma) = \sigma^{-n}\prod_{i=1}^{n} f\left(\frac{x_i - \mu}{\sigma}\right)$$

for some known probability density function f. We assume that $\{x\colon f(x) > 0\}$ is an open interval and that $g = -\log f$ has a positive and continuous second-order derivative on that interval. This ensures that the maximum likelihood estimate $(\hat{\mu}, \hat{\sigma})$ exists uniquely with probability 1 [cf., for instance, Burridge (1981)].

Taking as the auxiliary a Fisher's configuration statistic

$$a = (a_1, \ldots, a_n) = \left(\frac{x_1 - \hat{\mu}}{\hat{\sigma}}, \ldots, \frac{x_n - \hat{\mu}}{\hat{\sigma}}\right),$$

which is an exact ancillary, we find

$$\mathit{j}(\mu, \sigma) = \sigma^{-2}\begin{bmatrix} \Sigma g''(a_\nu) & \Sigma a_\nu g''(a_\nu) \\ \Sigma a_\nu g''(a_\nu) & n + \Sigma a_\nu^2 g''(a_\nu) \end{bmatrix}$$

and, in an obvious notation,

$$\begin{aligned}
\not{l}_{\mu\mu;\hat{\mu}} &= -\sigma^{-3}\Sigma g'''(a_\iota),\\
\not{l}_{\mu\mu;\hat{\sigma}} &= -\sigma^{-3}\Sigma a_\iota g'''(a_\iota),\\
\not{l}_{\mu\sigma;\hat{\mu}} &= -\sigma^{-3}\{2\Sigma g''(a_\iota) + \Sigma a_i g'''(a_\iota)\},\\
\not{l}_{\mu\sigma;\hat{\sigma}} &= -\sigma^{-3}\{2\Sigma a_\iota g''(a_\iota) + \Sigma a_i^2 g'''(a_i)\},\\
\not{l}_{\sigma\sigma;\hat{\mu}} &= -\sigma^{-3}\{4\Sigma a_\iota g''(a_\iota) + \Sigma a_i^2 g'''(a_\iota)\},\\
\not{l}_{\sigma\sigma;\hat{\sigma}} &= -\sigma^{-3}\{2n + 4\Sigma a_i^2 g''(a_\iota) + \Sigma a_i^3 g'''(a_\iota)\},\\
\not{l}_{\mu\mu\mu} &= \sigma^{-3}\Sigma g'''(a_\iota),\\
\not{l}_{\mu\mu\sigma} &= \sigma^{-3}\{2\Sigma g''(a_\iota) + \Sigma a_i g'''(a_\iota)\},\\
\not{l}_{\mu\sigma\sigma} &= \sigma^{-3}\{4\Sigma a_i g''(a_\iota) + \Sigma a_i^2 g'''(a_\iota)\},\\
\not{l}_{\sigma\sigma\sigma} &= \sigma^{-3}\{4n + 6\Sigma a_i^2 g''(a_\iota) + \Sigma a_i^3 g'''(a_\iota)\}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\not{T}_{\mu\mu\mu} &= 2\sigma^{-3}\not{l}_{\mu\mu\mu}((0,1);a),\\
\not{T}_{\mu\mu\sigma} &= -2\sigma^{-3}\not{j}_{\mu\mu}((0,1);a) + 2\sigma^{-3}\not{l}_{\mu\mu\sigma}((0,1);a),\\
\not{T}_{\sigma\sigma\mu} &= -4\sigma^{-3}\not{j}_{\mu\sigma}((0,1);a) + 2\sigma^{-3}\not{l}_{\sigma\sigma\mu}((0,1);a),\\
\not{T}_{\sigma\sigma\sigma} &= -6\sigma^{-3}\not{j}_{\sigma\sigma}((0,1);a) + 2\sigma^{-3}\not{l}_{\sigma\sigma\sigma}((0,1);a).
\end{aligned}$$

EXAMPLE 2.3. *Hyperboloid model.* Let $(u_1, v_1), \dots, (u_n, v_n)$ be a sample from the hyperboloid distribution

$$p(u,v;\chi,\varphi) = (2\pi)^{-1}\lambda e^{\lambda}\sinh u \exp[-\lambda\{\cosh\chi\cosh u - \sinh\chi\sinh u\cos(v-\varphi)\}]. \tag{2.17}$$

Here $0 \le u < \infty$, $0 \le v < 2\pi$ and the parameters χ and φ vary in $[0,\infty)$ and $[0, 2\pi)$, respectively, while $\lambda > 0$ is a precision parameter which we consider as known.

This distribution is analogous to the von Mises–Fisher or Langevin distribution for three-dimensional unit vectors, but pertains to observations on the positive unit hyperboloid in R^3 rather than the unit sphere. The distribution was introduced in Barndorff-Nielsen (1978) and its most important properties, including those on which we build below, have been unravelled by Jensen (1981); see also Blæsild and Jensen (1981).

The hyperboloid model (2.17) is a transformation model, the acting group being the special pseudo-orthogonal group $SO^{\uparrow}(1,2)$, and

$$a = \{(\Sigma\cosh u_i)^2 - (\Sigma\sinh u_i\cos v_i)^2 - (\Sigma\sinh u_\iota\sin v_\iota)^2\}$$

is maximal invariant after minimal sufficient reduction. Furthermore, the maximum likelihood estimate $(\hat{\chi}, \hat{\varphi})$ of (χ, φ) exists uniquely, with probability 1, $(a, \hat{\chi}, \hat{\varphi})$ is minimal sufficient, and the conditional distribution of $(\hat{\chi}, \hat{\varphi})$ given the ancillary a is again hyperboloidic, as in (2.17) but with u, v, and λ replaced by $\hat{\chi}$, $\hat{\varphi}$, and $a\lambda$ (the von Mises–Fisher distribution having a similar property). It may also be noted that $s = a - n$ follows the gamma distribution

$$\frac{\lambda^{n-1}}{\Gamma(n-1)} s^{n-2} e^{-\lambda s}.$$

It follows that the log likelihood function is

$$\begin{aligned} l(\chi, \varphi) &= l(\chi, \varphi; \hat{\chi}, \hat{\varphi}, a) \\ &= -a\lambda\{\cosh\chi\cosh\hat{\chi} - \sinh\chi\sinh\hat{\chi}\cos(\hat{\varphi} - \varphi)\} \end{aligned}$$

and hence

$$\overset{\alpha}{\not{\Gamma}}_{\chi\chi\chi} = \overset{\alpha}{\not{\Gamma}}_{\chi\chi\varphi} = \overset{\alpha}{\not{\Gamma}}_{\chi\varphi\chi} = \overset{\alpha}{\not{\Gamma}}_{\varphi\varphi\varphi} = 0,$$

$$\overset{\alpha}{\not{\Gamma}}_{\chi\varphi\varphi} = a\lambda\cosh\chi\sinh\chi,$$

$$\overset{\alpha}{\not{\Gamma}}_{\varphi\varphi\chi} = -a\lambda\cosh\chi\sinh\chi,$$

whatever the value of α. Thus, in this case, the α-geometries are identical.

We note again that whereas the auxiliary statistic a is taken so as to be ancillary in the various examples discussed—exact distribution constant in the three examples above and asymptotical distribution constant in the one to follow—ancillarity is no prerequisite for the general theory developed in this paper.

Furthermore, let a be any statistic which depends on the minimal sufficient statistic t, say, only and suppose that the mapping from t to $(\hat{\omega}, a)$ is defined and one-to-one on some subset $\mathscr{T}_0$ of the full range $\mathscr{T}$ of values of t though not, perhaps, on all of $\mathscr{T}$. We can then endow the model $\mathscr{M}$ with observed geometries, in the manner described above, for values of t in $\mathscr{T}_0$. The next example illustrates this point.

The above considerations allow us to deal with questions of nonuniqueness and nonexistence of maximum likelihood estimates and nonexistence of exact ancillaries, especially in asymptotic considerations.

EXAMPLE 2.4. *Inverse Gaussian–Gaussian model.* Let $x(\cdot)$ and $y(\cdot)$ be independent Brownian motions with a common diffusion coefficient $\sigma^2 = 1$ and drift coefficients $\mu > 0$ and ξ, respectively. We observe the process $x(\cdot)$ until it first hits a level $x_0 > 0$ and at the time u when this happens we record the value $v = y(u)$ of the second process. The joint distribution of u and v is then given by

$$\begin{aligned} p(u, v; \mu, \xi) &= (2\pi)^{-1} x_0 e^{x_0\mu} u^{-2} \exp\left[-\tfrac{1}{2}(x_0^2 + v^2)u^{-1}\right] \\ &\quad \times \exp\left[-\tfrac{1}{2}\mu^2 u + \xi v - \tfrac{1}{2}\xi^2 u\right]. \end{aligned} \tag{2.18}$$

Suppose that $(u_1, v_1), \ldots, (u_n, v_n)$ is a sample from the distribution (2.18) and let $t = (\bar{u}, \bar{v})$, where $\bar{u}$ and $\bar{v}$ are the arithmetic means of the observations. Then t is minimal sufficient and follows a distribution similar to (2.18), specifically

$$p(\bar{u}, \bar{v}; \mu, \xi) = (2\pi)^{-1} x_0 n e^{n x_0 \mu} \bar{u}^{-2} \exp\left[-\frac{n}{2}\left(x_0^2 + \bar{v}^2\right)\bar{u}^{-1}\right] \tag{2.19}$$
$$\times \exp\left[-\frac{n}{2}\mu^2\bar{u} + n\xi\bar{v} - \frac{n}{2}\xi^2\bar{u}\right].$$

Now, assume ξ equal to μ. The model (2.19) is then a (2, 1) exponential model, still with t as minimal sufficient statistic. The maximum likelihood estimate of μ is undefined if $t \notin \mathscr{T}_0$, where

$$\mathscr{T}_0 = \{t = (\bar{u}, \bar{v})\colon x_0 + \bar{v} \geq 0\},$$

whereas for $t \in \mathscr{T}_0$, $\hat{\mu}$ exists uniquely and is given by

$$\hat{\mu} = \tfrac{1}{2}(x_0 + \bar{v})\bar{u}^{-1}. \tag{2.20}$$

The event $t \notin \mathscr{T}_0$ happens with a probability that decreases exponentially fast with the sample size n and may therefore be ignored for most statistical purposes.

Defining, formally, $\hat{\mu}$ to be given by (2.20) even for $t \notin \mathscr{T}_0$ and letting

$$a = \Phi^{-}\left(\bar{u}; 2nx_0^2, 2n\hat{\mu}^2\right),$$

where $\Phi^{-}(\cdot; \chi, \psi)$ denotes the distribution function of the inverse Gaussian distribution with density function

$$\varphi^{-}(x; \chi, \psi) = (2\pi)^{-1/2}\sqrt{\chi}\, e^{\sqrt{\chi\psi}} x^{-3/2} \exp\left[-\tfrac{1}{2}(\chi x^{-1} + \psi x)\right], \tag{2.21}$$

we have that the mapping $t \to (\hat{\mu}, a)$ is one-to-one from $\mathscr{T} = \{t = (\bar{u}, \bar{v})\colon \bar{u} > 0\}$ onto $(-\infty, +\infty) \times (0, \infty)$ and that a is asymptotically ancillary and has the property that $p^*(\hat{\mu}; \mu | a) = c|\hat{j}|^{1/2}\bar{L}$ approximates the actual conditional density of $\hat{\mu}$ given a to order $O(n^{-3/2})$, cf. Barndorff-Nielsen (1984a).

Letting $\Phi_{-}(\cdot; \chi, \psi)$ denote the inverse function of $\Phi^{-}(\cdot; \chi, \psi)$ we may write the log likelihood function for μ as

$$\begin{aligned} l(\mu) &= l(\mu; \hat{\mu}, a) \\ &= n\{(x_0 + \bar{v})\mu - \bar{u}\mu^2\} \\ &= n\Phi_{-}\left(a; 2nx_0^2, 2n\hat{\mu}^2\right)\{2\hat{\mu}\mu - \mu^2\}. \end{aligned} \tag{2.22}$$

From this we find

$$l_{\mu\mu} = -2n\Phi_{-}\left(a; 2nx_0^2, 2n\hat{\mu}^2\right),$$

so that

$$j = 2h\Phi_{-}\left(a; 2nx_0^2, 2n\mu^2\right),$$
$$l_{\mu\mu\mu} = 0$$

and

$$\not{l}_{\mu\mu;\hat{\mu}} = 8n^2\mu\left(\varphi^- \circ \Phi_-/\Phi_\psi^-\right)\left(a; 2nx_0^2, 2n\mu^2\right)$$
$$= \overset{1}{\not{\Gamma}}_{\mu\mu\mu} = -\tfrac{1}{2}\overset{-1}{\not{\Gamma}}_{\mu\mu\mu},$$

where Φ_ψ^- denotes the derivative of $\Phi^-(x; \chi, \psi)$ with respect to ψ. By the well-known result [Shuster (1968)]

$$\Phi^-(x; \chi, \psi) = \Phi\left(\psi^{1/2}x^{1/2} - \chi^{1/2}x^{-1/2}\right) + e^{2\sqrt{\chi\psi}}\Phi\left(-\left(\psi^{1/2}x^{1/2} + \chi^{1/2}x^{-1/2}\right)\right),$$

where Φ is the distribution function of the standard normal distribution, Φ_ψ^- could be expressed in terms of Φ and $\varphi = \Phi'$.

For any $m = 2, 3, \ldots$ a covariant tensor on $\mathscr{M}$ of rank m is given by

$$E\{\partial_{r_1} l\, \partial_{r_2} l \cdots \partial_{r_m} l\}, \tag{2.23}$$

the first two of which, i.e., i and T, determine the expected geometries studied by Amari and others, as discussed briefly in the foregoing. The tensors $\not{j}$ and $\not{T}$ are observed analogues of i and T and it seems natural to enquire whether, similarly, there exists observed analogues of (2.23) for $m > 3$.

For $m = 4$ an approach like that used above to derive $\not{T}$ does not, in any obvious way at least, lead to a fourth-rank tensor. However, one may proceed otherwise by noting that $\not{T}$ equals the covariant derivative of $\not{j}$ relative to the $\overset{1}{\not{\Gamma}}$ connection. In fact, denoting the operation of covariant differentiation with respect to ω^r and relative to $\overset{\alpha}{\not{\Gamma}}$ by $\overset{\alpha}{\not{D}}_r$ we have

$$\begin{aligned}\overset{\alpha}{\not{D}}_t \not{j}_{rs} &= \partial_t \not{j}_{rs} - \overset{\alpha}{\not{\Gamma}}{}_{rt}^{u} \not{j}_{us} - \overset{\alpha}{\not{\Gamma}}{}_{st}^{u} \not{j}_{ru} \\ &= \alpha \not{T}_{rst}\end{aligned} \tag{2.24}$$

and hence $\not{T}_{rst} = \overset{1}{\not{D}}_t \not{j}_{rs}$. Similarly, with $\overset{\alpha}{D}$ indicating covariant differentiation as determined by the expected connection $\overset{\alpha}{\Gamma}$, we have

$$\overset{\alpha}{D}_t i_{rs} = \alpha T_{rst}. \tag{2.25}$$

Formulas (2.24) and (2.25) are special cases of a more general differential-geometric result due to Lauritzen (1984). Taking now the covariant derivative of $\not{T}_{rst}$ we obtain

$$\begin{aligned}\overset{1}{\not{D}}_u \not{T}_{rst} &= \partial_u \not{T}_{rst} - \overset{1}{\Gamma}{}_{ru}^{v} \not{T}_{stv} - \overset{1}{\Gamma}{}_{su}^{v} \not{T}_{rtv} - \overset{1}{\Gamma}{}_{tu}^{v} \not{T}_{rsv} \\ &= -\not{l}_{rstu} - \not{l}_{rst;u}[4] - \not{l}_{rs;tu} - \not{l}_{rt;su} - \not{l}_{st;ru} \\ &\quad - \not{l}_{ru;w} \not{T}_{stv} \not{j}^{vw} - \not{l}_{su;w} \not{T}_{rtv} \not{j}^{vw} - \not{l}_{tu;w} \not{T}_{rsv} \not{j}^{vw}.\end{aligned}$$

In contrast to $\not{T}_{rst}$ this expression is not symmetric in the four indices. To obtain symmetry we introduce

$$(2.26)\qquad \begin{aligned} \not{T}_{rstu} &= \tfrac{1}{4}\overset{1}{\not{D}}_u \not{T}_{rst}[4] \\ &= -\{\not{l}_{rstu} + \not{l}_{rst;\,u}[4] + \tfrac{1}{2}(\not{l}_{rs;\,tu} + \not{l}_{rs;\,w}\not{T}_{tuv}\not{j}^{vw})[6]\} \end{aligned}$$

which is a covariant tensor of rank 4. This may be compared to each of the "expected tensors"

$$(2.27)\qquad \begin{aligned} T_{rstu} &= \tfrac{1}{4}\overset{1}{D}_u T_{rst}[4] \\ &= -E\{l_{rstu} + l_{rst}l_u[4] + \tfrac{1}{2}(l_{rs}l_{tu} + l_{rs}l_t l_u + l_{rs}l_v T_{tuw}i^{vw}[6])\} \end{aligned}$$

and

$$(2.28)\qquad M_{rstu} = E\{\partial_r l\, \partial_s l\, \partial_t l\, \partial_u l\}.$$

The latter may, as appears by differentiation of (2.13) twice, be rewritten as

$$(2.29)\qquad M_{rstu} = -E\{l_{rstu} + l_{rst}l_u[4] + l_{rs}l_{tu}[3] + l_{rs}l_t l_u[6]\}.$$

The tensors (2.26) and (2.27) are closely analogous. In particular, they are identical for (k, k) exponential models, with common value $-E\{l_{rstu}\}$. In contrast to this, M_{rstu} does not equal the common value of $\not{T}_{rstu}$ and T_{rstu} for such models. But if instead of the fourth moment of the score vector, i.e., M_{rstu}, we consider the fourth cumulant, i.e.,

$$K_{rstu} = M_{rstu} - i_{rs}i_{tu}[3],$$

then we find that this is also a tensor, that

$$K_{rstu} = -E\{l_{rstu} + l_{rst}l_u[4] + \tfrac{1}{2}(l_{rs}l_{tu} + l_{rs}l_t l_u + (l_{rs}l_t l_u - i_{rs}i_{tu}))[6]\},$$

and that for (k, k) exponential models $\not{T}_{rstu} = T_{rstu} = K_{rstu}$.

More generally, for $m = 2, 3, \ldots$ let $K_{r_1 \ldots r_m}$ denote the mth-order cumulant of the score vector $\partial l = (\partial_1 l, \ldots, \partial_d l)$. From the tensorial nature of the moments (2.23) of ∂l and from the general formulae relating moments and cumulants, cf. for instance Speed (1983) or McCullagh (1984b), we find that $K_{r_1 \ldots r_m}$ is a covariant tensor of rank m. Furthermore, writing $\not{T}_{rs}$ for $\not{j}_{rs}$, T_{rs} for i_{rs}, and defining $\not{T}_{r_1 \ldots r_m}$ and $T_{r_1 \ldots r_m}$ recursively by

$$\not{T}_{r_1 \ldots r_{m+1}} = \frac{1}{m+1}\overset{1}{\not{D}}_{r_{m+1}}\not{T}_{r_1 \ldots r_m}[m+1]$$

and

$$T_{r_1 \ldots r_{m+1}} = \frac{1}{m+1}\overset{1}{D}_{r_{m+1}}T_{r_1 \ldots r_m}[m+1],$$

we have $\not{T}_{r_1 \ldots r_m} = T_{r_1 \ldots r_m} = K_{r_1 \ldots r_m}$ for (k, k) exponential models and for any $m \geq 2$.

3. Expansion of $c|\hat{j}|^{1/2}\bar{L}$. We shall derive an asymptotic expansion of (1.5), by Taylor expansion of $c|\hat{j}|^{1/2}\bar{L}$ in $\hat{\omega}$ around ω, for fixed value of the auxiliary a.

The various terms of this expansion are given by mixed derivatives [cf. (1.4)] of the log model function. It should be noted that for arbitrary choice of the auxiliary statistic a the quantity $c|\hat{j}|^{1/2}\bar{L}$ constitutes a probability (density) function on the domain of variation of $\hat{\omega}$ and the expansions below are valid. However, $c|\hat{j}|^{1/2}\bar{L}$ furnishes an approximation to the actual conditional distribution of $\hat{\omega}$ given a, as discussed in Section 1, only for suitable ancillary specification of a.

To expand $c|\hat{j}|^{1/2}\bar{L}$ in $\hat{\omega}$ around ω we first write $\bar{L}$ as $\exp\{l - \hat{l}\}$ and expand l in ω around $\hat{\omega}$. By Taylor's formula,

$$l - \hat{l} = \sum_{\nu=2}^{\infty} \frac{1}{\nu!}(\omega - \hat{\omega})^{r_1} \cdots (\omega - \hat{\omega})^{r_\nu}(\partial_{r_1} \cdots \partial_{r_\nu} l)(\hat{\omega})$$

whence, expanding each of the terms $(\partial_{r_1} \cdots \partial_{r_\nu} l)(\hat{\omega})$ around ω,

$$\begin{aligned} l - \hat{l} = &\sum_{\nu=2}^{\infty} \frac{(-1)^\nu}{\nu!}(\hat{\omega} - \omega)^{r_1} \cdots (\hat{\omega} - \omega)^{r_\nu} \\ &\times \sum_{\rho=0}^{\infty} \frac{1}{\rho!}(\hat{\omega} - \omega)^{s_1} \cdots (\hat{\omega} - \omega)^{s_\rho} \partial_{s_1} \cdots \partial_{s_\rho} \not{l}_{r_1 \ldots r_\nu}. \end{aligned} \tag{3.1}$$

Consequently, writing δ for $\hat{\omega} - \omega$ and $\delta^{rs\cdots}$ for $(\hat{\omega} - \omega)^r(\hat{\omega} - \omega)^s \cdots$, we have

$$\begin{aligned} l - \hat{l} = &-\tfrac{1}{2}\delta^* \not{j} \delta + \tfrac{1}{2}\delta^{rst}\left(\not{l}_{rs;\,t} + \tfrac{2}{3}\not{l}_{rst}\right) \\ &+ \tfrac{1}{24}\delta^{rstu}\left(6\not{l}_{rs;\,tu} + 8\not{l}_{rst;\,u} + 3\not{l}_{rstu}\right) + \cdots . \end{aligned} \tag{3.2}$$

Next, we wish to expand $\log\{|\hat{j}|/|\not{j}|\}^{1/2}$ in $\hat{\omega}$ around ω. To do this we observe that if A is a $d \times d$ matrix whose elements a_{rs} depend on ω then

$$\begin{aligned} \partial_t \log|A| &= |A|^{-1}\, \partial_t |A| \\ &= a^{sr}\, \partial_t a_{rs}, \end{aligned}$$

where a^{rs} denotes the (r, s)-element of the inverse of A. Furthermore, using

$$\partial_t a^{rs} = -a^{rv} a^{ws}\, \partial_t a_{vw},$$

which is obtained by differentiating $a_{ru} a^{us} = \delta_r^s$ with respect to ω^t and solving for a^{rs}, we find

$$\partial_t \partial_u \log|A| = -a^{sv} a^{wr}\, \partial_u a_{vw}\, \partial_t a_{rs} + a^{sr} \partial_t\, \partial_u a_{rs}.$$

It follows that

$$\begin{aligned} \log\{|\hat{j}|/|\not{j}|\}^{1/2} = &-\tfrac{1}{2}\delta^t \not{j}^{rs}\left(\not{l}_{rst} + \not{l}_{rs;\,t}\right) \\ &-\tfrac{1}{4}\delta^{tu}\{\not{j}^{rs}\left(\not{l}_{rstu} + \not{l}_{rst;\,u} + \not{l}_{rsu;\,t} + \not{l}_{rs;\,tu}\right) \\ &\qquad + \not{j}^{rv}\not{j}^{sw}\left(\not{l}_{rst} + \not{l}_{rs;\,t}\right)\left(\not{l}_{vwu} + \not{l}_{vw;\,u}\right)\} + \cdots . \end{aligned} \tag{3.3}$$

By means of (3.2) and (3.3) we therefore find

$$c|\hat{j}|^{1/2}\bar{L} = (2\pi)^{d/2} c\varphi_d(\hat{\omega} - \omega; \not{j})\{1 + A_1 + A_2 + \cdots\}, \tag{3.4}$$

where $\varphi_d(\cdot\,; j)$ denotes the density function of the d-dimensional normal distribution with mean 0 and precision (i.e., inverse variance–covariance matrix) j and where

$$A_1 = -\tfrac{1}{2}\delta^{t} j^{rs}(\ell_{rs;\,t} + \ell_{rst}) + \tfrac{1}{2}\delta^{rst}(\ell_{rs;\,t} + \tfrac{2}{3}\ell_{rst}) \tag{3.5}$$

and

$$\begin{aligned} A_2 = \tfrac{1}{24}\big[& -3\delta^{tu}\{2j^{rs}(\ell_{rstu} + \ell_{rst;\,u} + \ell_{rsu;\,t} + \ell_{rs;\,tu}) \\ & \qquad + (2j^{rv}j^{sw} - j^{rs}j^{vw})(\ell_{rs;\,t} + \ell_{rst})(\ell_{vw;\,u} + \ell_{vwu})\} \\ & + \delta^{rstu}\{(3\ell_{rstu} + 8\ell_{rst;\,u} + 6\ell_{rs;\,tu}) \\ & \qquad - 6j^{vw}(\ell_{vw;\,u} + \ell_{vwu})(\ell_{rs;\,t} + \tfrac{2}{3}\ell_{rst})\} \\ & + 3\delta^{rstuvw}(\ell_{rs;\,t} + \tfrac{2}{3}\ell_{rst})(\ell_{uv;\,w} + \tfrac{2}{3}\ell_{uvw})\big], \end{aligned} \tag{3.6}$$

A_1 and A_2 being of order $O(n^{-1/2})$ and $O(n^{-1})$, respectively, under ordinary repeated sampling.

By integration of (3.4) with respect to $\hat{\omega}$ we obtain

$$(2\pi)^{d/2}c = 1 + C_1 + \cdots, \tag{3.7}$$

where C_1 is obtained from A_2 by changing the sign of A_2 and making the substitutions

$$\begin{aligned} \delta^{rs} &\to j^{rs}, \\ \delta^{rstu} &\to j^{rs}j^{tu}[3], \\ \delta^{rstuvw} &\to j^{rs}j^{tu}j^{vw}[15], \end{aligned}$$

the 3 and 15 terms in the two latter expressions being obtained by appropriate permutations of the indices (thus, for example, $\delta^{rstu} \to j^{rs}j^{tu} + j^{rt}j^{su} + j^{ru}j^{st}$).

Combination of (3.4) and (3.7) finally yields

$$c|\hat{j}|^{1/2}\bar{L} = \varphi(\hat{\omega} - \omega; j)\{1 + A_1 + (A_2 + C_1) + \cdots\} \tag{3.8}$$

with an error term which in wide generality is of order $O(n^{-3/2})$ under repeated sampling. In comparison with an Edgeworth expansion it may be noted that the expansion (3.8) is in terms of mixed derivatives of the log model function, rather than in terms of cumulants, and that the error of (3.8) is relative, rather than absolute.

In particular, under repeated sampling and if the auxiliary statistic is (approximately or exactly) ancillary such that

$$p(\hat{\omega}; \omega|a) = p^*(\hat{\omega}; \omega|a)\{1 + O(n^{-3/2})\}$$

(cf. Section 1) we generally have

$$p(\hat{\omega}; \omega|a) = \varphi_d(\hat{\omega} - \omega; j)\{1 + A_1 + (A_2 + C_1) + O(n^{-3/2})\}. \tag{3.9}$$

For one-parameter models, i.e., for $d = 1$, the expansion (3.8) with A_1, A_2, and C_1 as given above reduces to an expansion derived in Barndorff-Nielsen and Cox (1984a). Using that expansion confidence limits for ω, valid to order $O(n^{-3/2})$, have been derived in Barndorff-Nielsen (1985a). In the former of those two papers a relation valid to order $O(n^{-3/2})$ was established, for general d, between the norming constant c of (1.4) and the Bartlett adjustment factors for likelihood ratio tests of hypotheses about ω. By means of this relation such adjustment factors may be simply calculated from the expression for C_1.

EXAMPLE 3.1. Suppose $\mathscr{M}$ is a (k, k) exponential model with model function (2.14). Then the expression for C_1 takes the form

$$C_1 = \tfrac{1}{24}\left\{3\kappa_{rstu}\kappa^{rs}\kappa^{tu} - \kappa_{rst}\kappa_{uvw}\left(2\kappa^{ru}\kappa^{sv}\kappa^{tw} + 3\kappa^{rs}\kappa^{tu}\kappa^{vw}\right)\right\},$$

where, for $\partial_r = \partial/\partial\theta^r$ and $\kappa(\theta) = -\log a(\theta)$,

$$\kappa_{rs\ldots} = \partial_r\,\partial_s \ldots \kappa(\theta)$$

and where κ^{rs} is the inverse matrix of κ_{rs}.

From (3.8) we find the following expansion for the mean value of $\hat{\omega}$:

$$E_\omega \hat{\omega}^\alpha = \omega^\alpha + \mu_1^\alpha + \mu_2^\alpha + \cdots,$$

where μ_1^α is of order $O(n^{-1})$, μ_2^α is of order $O(n^{-2})$, and

$$\mu_1^\alpha = -\tfrac{1}{2}\mathscr{j}^{\alpha r}\mathscr{j}^{st}\mathscr{l}_{r;st} = -\tfrac{1}{2}\mathscr{j}^{\alpha r}\mathscr{j}^{st}\overset{-1}{\mathscr{T}}_{str}. \tag{3.10}$$

Hence, from (3.8) and writing δ' for $\delta - \mu_1$,

$$\begin{aligned} c|\hat{j}|^{1/2}\bar{L} &= \varphi_d(\hat{\omega} - \omega - \mu_1; \mathscr{j})\left\{1 + \left(A_1 - \delta^r \mathscr{j}_{rs}\mu_1^s\right) + \cdots\right\} \\ &= \varphi_d(\hat{\omega} - \omega - \mu_1; \mathscr{j})\left\{1 + \tfrac{1}{2}h^{rst}(\delta'; \mathscr{j})\left(\mathscr{l}_{rs;t} + \tfrac{2}{3}\mathscr{l}_{rst}\right) + \cdots\right\}, \end{aligned} \tag{3.11}$$

where the error term is of order $O(n^{-1})$ and where $h^{r_1\cdots r_n}(\cdot; \mathscr{j})$ denotes the tensorial Hermite polynomial [as defined by Amari and Kumon (1983)], relative to the tensor $\mathscr{j}_{rs}$. Using (2.12) we may rewrite the last quantity in (3.11) as

$$\mathscr{l}_{rs;t} + \tfrac{2}{3}\mathscr{l}_{rst} = -\overset{-1/3}{\mathscr{T}}_{rst} + \mathscr{R}_{rst}, \tag{3.12}$$

where

$$\mathscr{R}_{rst} = \tfrac{4}{3}\left(\mathscr{l}_{rs;t} - \tfrac{1}{2}\left(\mathscr{l}_{rt;s} + \mathscr{l}_{st;r}\right)\right). \tag{3.13}$$

Since

$$h^{rst}(\delta'; \mathscr{j}) = \delta'^r\delta'^s\delta'^t - \mathscr{j}^{rs}\delta'^t[3] \tag{3.14}$$

we find

$$h^{rst}(\delta'; \mathscr{j})\mathscr{R}_{rst} = 0$$

and hence (3.11) reduces to

$$c|\hat{j}|^{1/2}\bar{L} = \varphi_d(\hat{\omega} - \omega - \mu_1; \mathscr{j})\left\{1 - \tfrac{1}{2}h^{rst}(\delta'; \mathscr{j})\overset{-1/3}{\mathscr{T}}_{rst} + \cdots\right\}, \tag{3.15}$$

the error term being $O(n^{-1})$.

Suppose, in particular, that the model is an exponential (k, d) model. We may then compare (3.15) with the Edgeworth expansion for an efficient, bias adjusted estimate of ω given an ancillary statistic, provided by formulas (3.33) and (3.25) in Amari and Kumon (1983). It appears that $h^{rst}(\delta'; \jmath) \overset{-1/3}{\not{\Gamma}}_{rst}$ of (3.15) is the counterpart of Amari and Kumon's $\overset{-1/3}{\Gamma}_{abc}h^{abc} - \overset{e}{H}_{ab\kappa}h^{ab}h^{\kappa} + \overset{m}{H}_{\kappa\lambda a}h^{a}h^{\kappa\lambda}$. Thus (3.15) offers some simplification over the corresponding expression provided by the Amari and Kumon paper.

Note that, again by the symmetry of (3.14), if

$$\overset{-1/3}{\not{\Gamma}}_{rst}[3] = 0 \tag{3.16}$$

for all r, s, t then the first-order correction term in (3.15) is 0. Furthermore, for any one-parameter model $\mathscr{M}$ the quantity $\overset{\alpha}{\not{\Gamma}}$ with $\alpha = -\frac{1}{3}$ can be made to vanish by choosing that parametrization for which ω is the geodesic coordinate for the $-\frac{1}{3}$ observed connection. (Note that generally this parametrization will depend on the value of the ancillary a.) An analogous result holds for the Edgeworth expansion derived by Amari and Kumon (1983), referred to above. The parametrization making the $\alpha = -\frac{1}{3}$ expected connection $\overset{\alpha}{\Gamma}$ vanish has the interpretation of a skewness reducing parametrization, cf. Kass (1984).

Acknowledgments. I am much indebted to Peter E. Jupp, Steffen L. Lauritzen, and Hans Anton Salomonsen for valuable discussions. Helpful comments from an associate editor and two referees are also gratefully acknowledged.

REFERENCES

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

AMARI, S.-I. (1982b). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* **69** 1–17.

AMARI, S.-I. (1983). Comparisons of asymptotically efficient tests in terms of geometry of statistical structures. *Bull. Internat. Statist. Inst.* **50** 1190–1206.

AMARI, S.-I. (1984). Differential geometrical theory of statistics—towards new developments. To appear in *Differential Geometry in Statistical Inference*. IMS monograph.

AMARI, S.-I. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics* **28**. Springer, Heidelberg.

AMARI, S.-I. and KUMON, M. (1983). Differential geometry of Edgeworth expansion in curved exponential family. *Ann. Inst. Statist. Math.* **35** 1–24.

BARNDORFF-NIELSEN, O. E. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.* **5** 151–157.

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

BARNDORFF-NIELSEN, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365.

BARNDORFF-NIELSEN, O. E. (1984a). On conditionality resolution and the likelihood ratio for curved exponential families. *Scand. J. Statist.* **11** 157–170.

BARNDORFF-NIELSEN, O. E. (1984b). Differential and integral geometry in statistical inference. To appear in *Differential Geometry in Statistical Inference*. IMS monograph.

BARNDORFF-NIELSEN, O. E. (1985a). Confidence limits from $c|\hat{\jmath}|^{1/2}\bar{L}$ in the single-parameter case. *Scand. J. Statist.* **12** 83–87.

BARNDORFF-NIELSEN, O. E. (1985b). Inference on full or partial parameters based on the standardized signed log likelihood ratio. To appear in *Biometrika* **73**.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1984). Combination of reproductive models. Research Report 107, Dept. Theor. Statist., Aarhus Univ.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984a). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46** 483–495.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984b). The effect of sampling rules on likelihood statistics. *Internat. Statist. Rev.* **52** 309–326.

BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1984). Differential geometry, profile likelihood and L-sufficiency. Research Report 118, Dept. Theor. Statist., Aarhus Univ.

BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1985). Profile likelihood, marginal likelihood and differential geometry of composite transformation models. Research Report 122, Dept. Theor. Statist., Aarhus Univ.

BLÆSILD, P. and JENSEN, J. L. (1981). Multivariate distributions of hyperbolic type. In *Statistical Distributions in Scientific Work.* (C. Taillie, G. P. Patil and B. A. Baldessari, eds.) **4** 45–66. D. Reidel, Dordrecht.

BURRIDGE, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *J. Roy. Statist. Soc. Ser. B* **43** 41–45.

CHENTSOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference.* Nauka, Moscow (in Russian). (English translation in *Translation of Mathematical Monographs* **53** (1982). Amer. Math. Soc., Providence, R.I.)

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

JENSEN, J. L. (1981). On the hyperboloid distribution. *Scand. J. Statist.* **8** 193–206.

KASS, R. E. (1984). Canonical parameterizations and zero parameter-effects curvature. *J. Roy. Statist. Soc. Ser. B* **46** 86–92.

LAURITZEN, S. L. (1984). Statistical manifolds. To appear in *Differential Geometry in Statistical Inference.* IMS monograph.

MCCULLAGH, P. (1984a). Local sufficiency. *Biometrika* **71** 233–244.

MCCULLAGH, P. (1984b). Tensor notation and cumulants of polynomials. *Biometrika* **71** 461–476.

SHUSTER, J. J. (1968). A note on the inverse Gaussian distribution function. *J. Amer. Statist. Assoc.* **63** 1514–1516.

SPEED, T. P. (1983). Cumulants and partition lattices. *Austral. J. Statist.* **25** 378–388.

DEPARTMENT OF THEORETICAL STATISTICS
INSTITUTE OF MATHEMATICS
AARHUS UNIVERSITY
NY MUNKEGADE
DK-8000 AARHUS C
DENMARK