The Annals of Statistics
1988, Vol. 16, No. 3, 1009–1043

DIFFERENTIAL GEOMETRY, PROFILE LIKELIHOOD, L-SUFFICIENCY AND COMPOSITE TRANSFORMATION MODELS[1]

BY O. E. BARNDORFF-NIELSEN AND P. E. JUPP

University of Aarhus and University of St. Andrews

Let Ω denote the parameter space of a statistical model and let $\mathscr{K}$ be the domain of variation of the parameter of interest. Various differential-geometric structures on Ω are considered, including the expected information metric and the α-connections studied by Chentsov and Amari, as well as the observed information metric and the observed α-connections introduced by Barndorff-Nielsen. Under certain conditions these geometric objects on Ω can be transferred in a canonical purely differential-geometric way to $\mathscr{K}$. The transferred objects are related to structures on $\mathscr{K}$ obtained from derivatives of pseudolikelihood functions such as the profile likelihood, the modified profile likelihood and the marginal likelihood based on an L-sufficient statistic (cf. Rémon) when such a statistic exists. For composite transformation models it is shown that the modified profile likelihood is very close to the Laplace approximation to a certain integral representation of the marginal likelihood.

1. Introduction. Consider a parametric statistical model $(\mathscr{X}, p(x; \omega), \Omega)$ and let κ be a parameter of interest with range space $\mathscr{K}$. We shall denote the mapping sending $\omega \in \Omega$ to $\kappa \in \mathscr{K}$ by π. In certain circumstances inference on κ may appropriately be drawn separately, based on a marginal likelihood function or some other type of pseudolikelihood function, such as profile likelihood.

The original model induces various geometrical objects on the parameter space Ω considered as a differentiable manifold, such as the expected information metric and the α-connections studied by Chentsov (1972) and Amari (1982a, b, 1985, 1987) [see also Amari and Kumon (1983)] and the observed information metric and the observed α-connections introduced in Barndorff-Nielsen (1987). From that model and the associated geometrical structures it is in certain circumstances possible to derive, by various routes, more or less analogous geometrical structures on the interest parameter space $\mathscr{K}$. Our objective in this paper is to explore when such derivations are feasible and to compare the resulting geometrical structures from the viewpoint of statistics.

We shall be concerned mainly with cases where there exists an L-sufficient statistic for κ. A statistic u on the sample space $\mathscr{X}$ is said to be (minimal)

Received April 1985; revised October 1987.
[1]Research partly supported by the UK Science and Engineering Research Council and by the Danish Science Research Council.
AMS 1980 *subject classifications*. Primary 62F99; secondary 62B05.
Key words and phrases. Alpha connections, ancillary, composite transformation models, connections, cuts, distributional shape, expected information metric, likelihood, marginal likelihood, modified profile likelihood, observed alpha-connections, observed information metric, orthogonal parameters, parameter of interest, profile discrimination information, reproductive exponential models, statistical manifolds, submersion, tau-parallel foliations, tensors.

L-sufficient for κ if (i) the profile likelihood $\tilde{L}(\kappa) = \sup_{\omega|\kappa} L(\omega)$ for κ generates the same σ-algebra on $\mathscr{X}$ as u and (ii) the distribution of u depends on ω through κ only. Three main classes of instances where an L-sufficient statistic exists are (a) models with cuts, (b) exponential models with τ-parallel foliations, including many reproductive exponential models and (c) composite transformation models.

Section 2 summarizes relevant material on profile and marginal likelihoods and on statistical differential geometry. Section 3 is of a purely differential-geometric nature and discusses the conditions under which a submersion π of a manifold Ω to a manifold $\mathscr{K}$ of lower dimension is accompanied by a natural transfer of geometrical objects on Ω, such as tensors and connections, to similar objects on $\mathscr{K}$. We show that if a Riemannian metric and an affine connection transfer to $\mathscr{K}$, then the transfer of the covariant derivative of the metric tensor is equal to the covariant derivative, under the transferred connection, of the transferred metric. A similar property holds for the torsion tensor. It follows, in particular, that the Riemannian connection transfers to a Riemannian connection on $\mathscr{K}$. Further, preservation of sectional curvatures by such a transfer is related to the existence of orthogonal parameters.

In Section 4 we show that the existence of an L-sufficient statistic for κ ensures that observed statistical geometries on the parameter space Ω can be transferred to $\mathscr{K}$. These transferred geometries are equal to observed geometries on $\mathscr{K}$ based on the profile likelihood $\tilde{L}(\kappa)$ for κ. L-sufficiency for all sample sizes ensures transfer of expected geometries. Section 5 provides some discussion of the special cases (a), (b) and (c) already mentioned, as well as of an extended class of generalized linear models. In particular, for full exponential models the existence of a τ-parallel foliation is equivalent to the property of geometries transferring onto an associated subspace of canonical parameters. A more detailed and comprehensive study of the case of composite transformation models is carried out in Sections 6, 7 and 8.

In Section 6 we study the asymptotic relation, in the repeated sampling situation, between the profile likelihood $\tilde{L}(\kappa)$, the modified profile likelihood $\overset{\approx}{L}(\kappa)$ and the marginal likelihood $\check{L}(\kappa) = L(\kappa; u)$ for κ, based on the maximal invariant u of a composite transformation model. In particular, it is shown that in some important classes of cases, $\overset{\approx}{L}(\kappa)$ is equivalent to the Laplace approximation of a certain integral representation of $\check{L}(\kappa)$. The results form the basis for an asymptotic comparison of profile and marginal geometries on the range space $\mathscr{K}$ of κ. Such a comparison is carried out in Section 7, where we also compare profile geometrical objects on $\mathscr{K}$ to the corresponding objects obtained by geometric transfer from Ω and to objects obtained from a "profile" version of discrimination information. Finally, some examples are considered in Section 8.

2. Background material on profile likelihood and statistical geometries.

2.1. *Profile likelihood, modified profile likelihood and marginal likelihood.* The likelihood function of the statistical model $(\mathscr{X}, p(x; \omega), \Omega)$ is denoted by

$L(\omega)$, or by $L(\omega; x)$ when we wish to stress its dependence on the observation on which it is based. Let $l = \ln L$ and let j and i denote observed and expected information, i.e.,

$$j(\omega) = -\frac{\partial^2 l(\omega)}{\partial\omega\,\partial\omega}$$

and $i(\omega) = E_\omega j(\omega)$. The inverse matrices of j and i are called observed and expected formation, respectively. The parameter ω is assumed to be of the form $\omega = (\kappa, \psi)$, where κ is the parameter of interest, and

$$j = \begin{bmatrix} j_{\kappa\kappa} & j_{\kappa\psi} \\ j_{\psi\kappa} & j_{\psi\psi} \end{bmatrix}$$

indicates the corresponding block partition of j, with similar notations for the partitions of i, j^{-1} and i^{-1}.

The profile likelihood for the interest parameter κ is defined by

$$\tilde{L}(\kappa) = \sup_{\omega|\kappa} L(\omega), \tag{2.1}$$

i.e., it is the supremum over ω for fixed κ of the likelihood function $L(\omega)$. Since $\omega = (\kappa, \psi)$ we may in general write

$$\tilde{L}(\kappa) = L(\kappa, \hat{\psi}_\kappa), \tag{2.2}$$

where $\hat{\psi}_\kappa$ is the maximum likelihood estimate of ψ when κ is considered as known.

When the amount of information in the data x on the parameter ψ is large, the profile likelihood $\tilde{L}(\kappa)$ may be used as a pseudolikelihood function for κ. In cases where the amount of information on ψ is considerable, though not sufficient to warrant such use of $\tilde{L}(\kappa)$, it is in broad generality possible to adjust for the lack of information by multiplying $\tilde{L}(\kappa)$ by a certain factor. The resulting pseudolikelihood is termed the modified profile likelihood for κ and is defined by

$$\overset{\dagger}{L}(\kappa) = \left|\frac{\partial^2 l}{\partial\psi\,\partial\hat{\psi}}(\kappa, \hat{\psi}_\kappa; \hat{\kappa}, \hat{\psi}, a)\right|^{-1} \left|j_{\psi\psi}(\kappa, \hat{\psi}_\kappa)\right|^{1/2} \tilde{L}(\kappa), \tag{2.3}$$

where a is an ancillary statistic such that $(\hat{\kappa}, \hat{\psi}, a)$ is minimal sufficient. The rationale for and properties of this modified profile likelihood are discussed in Barndorff-Nielsen (1983, 1985).

Now suppose that there exists an L-sufficient statistic u for κ (cf. Section 1). The corresponding marginal likelihood $\check{L}(\kappa)$ for κ is then defined as the likelihood function for κ based on the marginal distribution of u, i.e., $\check{L}(\kappa) = L(\kappa; u)$.

In Section 2.2 we review the statistical geometries based on $L(\omega)$. Similar geometries can be constructed using $\check{L}(\kappa)$. In Section 4 we consider geometries based on the pseudolikelihoods $\tilde{L}(\kappa)$ and $\overset{\dagger}{L}(\kappa)$.

2.2. *Statistical geometries.* Following a suggestion of Lauritzen (1987) we shall speak of a differentiable manifold equipped with a pair (ϕ, S), where ϕ

is a metric tensor and S is a skewness tensor, as a *statistical manifold*. Here, by a skewness tensor we mean a symmetric covariant tensor of rank 3. Furthermore, we shall refer to such a pair (ϕ, S) as a *kit*. With each statistical manifold is associated a one-parameter family of affine connections—the α-connections—given by $\{\overset{\phi}{\Gamma} - \frac{1}{2}\alpha S^{\#}\colon -\infty < \alpha < +\infty\}$, where $\overset{\phi}{\Gamma}$ denotes the Riemannian connection determined by ϕ and the $(1,2)$-tensor $S^{\#}$ is defined by $\phi(S^{\#}(X, Y), Z) = S(X, Y, Z)$.

The parameter space Ω of a parametric statistical model $(\mathscr{X}, p(x; \omega), \Omega)$ can be usefully set up as a statistical manifold in two ways, at least. We denote the dimension of the parameter ω by d and we indicate the coordinates as $\omega = (\omega^1, \omega^2, \dots, \omega^d)$. Furthermore, with $l = l(\omega)$ as the log likelihood function, we write $l_i = \partial l/\partial \omega^i$, $l_{ij} = \partial^2 l/\partial \omega^i \partial \omega^j$, etc.

The *expected kit* (i, D) consists of the expected (or Fisher) information metric given by

$$i(\omega) = -\left[El_{ij}\right]$$

and of the expected skewness tensor whose (i, j, k)-entry is

$$D_{ijk}(\omega) = E_\omega\left(T_{ijk}\right),$$

where

$$T_{ijk} = -\left\{l_{ijk} + l_{ij}l_k + l_{jk}l_i + l_{ki}l_j\right\}.$$

The associated connections are the α-connections $\overset{\alpha}{\nabla}$ of Chentsov (1972) and Amari (1982a, b, 1987). To make this obvious we note that the tensor D may be reexpressed as

$$D_{ijk} = E_\omega\{l_i l_j l_k\}$$

as may be seen by differentiating the equation $El_i = 0$ twice.

Now, let $\hat{\omega}$ be the maximum likelihood estimator of ω and let a be an auxiliary statistic such that $(\hat{\omega}, a)$ is sufficient for ω. In applications a will be ancillary, i.e., in addition to $(\hat{\omega}, a)$ being sufficient a is distribution constant, either exactly or to a sufficient degree of approximation. By the sufficiency of $(\hat{\omega}, a)$ we may think of the log-likelihood function l, in its dependence on the data x, as being a function of $(\hat{\omega}, a)$ and this we indicate by writing $l(\omega; \hat{\omega}, a)$. Let $\partial_i = \partial/\partial \omega^i$ and $\delta_i = \partial/\partial \hat{\omega}^i$, and write

$$l_{i;\,j} = \partial_i \delta_j l, \qquad l_{ij;\,k} = \partial_i \partial_j \delta_k l,$$

etc. Furthermore, for any symbol indicating a function of ω and $(\hat{\omega}, a)$ we use a slash (/) through the symbol to indicate the substitution $\hat{\omega} \mapsto \omega$, turning the function into a function of ω and a. Thus

$$\not{l} = l(\omega; \omega, a), \qquad \not{l}_{ij} = l_{ij}(\omega; \omega, a), \qquad \not{l}_{ij;\,k} = l_{ij;\,k}(\omega; \omega, a),$$

etc. For any fixed value of a the *observed kit* $(\not{j}, \not{T})$ is given by $\not{j} = j(\omega; \omega, a)$, where $j = j(\omega; \hat{\omega}, a)$ is the observed information

$$j = -\frac{\partial^2 l}{\partial \omega\, \partial \omega} = -\left[l_{ij}\right],$$

and by the observed skewness

$$\not{T}_{ijk} = -\{\not{l}_{ijk} + \not{l}_{ij;\,k} + \not{l}_{jk;\,i} + \not{l}_{ki;\,j}\}.$$

The proof that this is a tensor, given in Barndorff-Nielsen (1987), rests on the relation

$$\not{l}_{i;\,j} = \not{j}_{ij}$$

which is obtained from the likelihood equation $l_i(\hat{\omega};\, \hat{\omega}, a) = 0$ by differentiation and substitution of ω for $\hat{\omega}$. The α-connections associated with $(\not{j}, \not{T})$ are denoted by $\overset{\alpha}{\not\nabla}$. Note that, though this is suppressed in the notation, the kit $(\not{j}, \not{T})$ and the connections $\overset{\alpha}{\not\nabla}$ depend on the value of a which is considered fixed, in accordance with the conditionality principle.

For some applications of the expected and observed kits (i, D) and $(\not{j}, \not{T})$, see Amari (1982a, b, 1987), Amari and Kumon (1983), Lauritzen (1987) and Barndorff-Nielsen (1986, 1987).

3. Transfer of geometries. In this section we consider a purely differential-geometric construction which we shall later apply in our statistical context. Let $\pi\colon \Omega \to \mathscr{K}$ be a differentiable function from one smooth manifold to another. We shall suppose that π is a *submersion*, i.e., at each point ω of Ω the tangent map π_* of π maps the tangent space $T\Omega_\omega$ *onto* the tangent space $T\mathscr{K}_{\pi(\omega)}$, or in terms of local coordinates $\omega^1, \dots, \omega^{p+q}$ on Ω and $\kappa^1, \dots, \kappa^p$ on $\mathscr{K}$, the Jacobian matrix $[\partial\kappa^i/\partial\omega^j]$ has rank p. It follows [see, e.g., Lang (1972), page 28 or Boothby (1975), pages 47 and 79] that each fibre $\pi^{-1}(\kappa)$ is a q-dimensional submanifold of Ω. Further, if Ψ denotes one of these fibres, then small portions of Ω look like small portions of $\mathscr{K} \times \Psi$ and π can be identified locally with projection of $\mathscr{K} \times \Psi$ onto $\mathscr{K}$. In our statistical context Ω will be the parameter space while $\mathscr{K}$ and Ψ will be the manifolds of the interest and incidental parameters, respectively. We shall consider ways in which geometric objects such as Riemannian metrics, tensors and affine connections on Ω can be transferred to corresponding geometric objects on $\mathscr{K}$.

This transfer of geometry along a submersion should be contrasted with the usual inheritance of geometry by submanifolds. If $\iota\colon \Theta \to \Omega$ is the inclusion map of the submanifold Θ in Ω (or, more generally, if ι is an immersion), then a Riemannian metric ϕ on Ω gives rise to a Riemannian metric $\iota^*\phi$ on Θ defined by $\iota^*\phi(X, Y) = \phi(\iota_*(X), \iota_*(Y))$ for $X, Y \in T\Theta_\theta$, or in terms of local coordinates $\theta^1, \dots, \theta^p$ on Θ and $\omega^1, \dots, \omega^{p+q}$ on Ω, $\iota^*\phi$ has components $(\partial\omega^i/\partial\theta^r)\phi_{ij}(\partial\omega^j/\partial\theta^s)$ (using the summation convention). More generally, ι induces a linear map $\iota^*\colon \mathscr{T}_s^{\,0}(\Omega) \to \mathscr{T}_s^{\,0}(\Theta)$ of "covariant" tensor fields of order s. Furthermore, given a Riemannian metric ϕ on Ω, the corresponding orthogonal projection of $T\Omega_{\iota(\theta)}$ onto $T\Theta_\theta$ enables tensor fields on Ω to give rise to corresponding tensor fields on Θ. Thus ι induces a linear map $\iota^*\colon \mathscr{T}_s^{\,r}(\Omega) \to \mathscr{T}_s^{\,r}(\Theta)$ of (r, s)-tensor fields, i.e., tensor fields of "contravariant" order r and "covariant" order s. Similarly, affine connections on Ω give rise to affine connections on Θ by means of the *Gauss decomposition* which is described for Riemannian connections in Kobayashi and Nomizu [(1969), Chapter VII]. The statistical

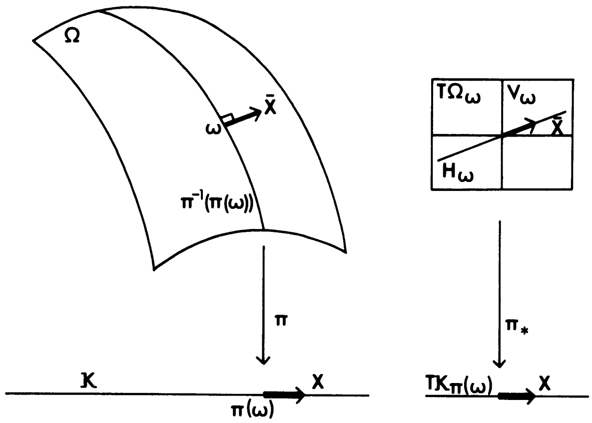

FIG. 1. *Illustration of the concepts of submersion π of one differentiable manifold Ω into another $\mathscr{K}$, and of lifting a tangent vector X of $\mathscr{K}$ to a horizontal tangent vector $\bar{X}$ of Ω.*

importance of this construction is its use in defining the curvature terms in the formulas of Efron (1975), Reeds (1975), Skovgaard (1984) and Amari (1982a, b).

For a submersion π: $\Omega \to \mathscr{K}$, a Riemannian metric ϕ on Ω gives rise to a useful decomposition of each tangent space of Ω. In the tangent space $T\Omega_\omega$ to Ω at ω, the *vertical subspace* V_ω is defined by $V_\omega = \{X \in T\Omega_\omega: \pi_*(X) = 0\}$. Thus, in terms of local coordinates $\psi^1, \ldots, \psi^q$ on the manifold Ψ already considered, V_ω can be identified with the span of $\{\partial/\partial\psi^i: i = 1, \ldots, q\}$. Using the Riemannian metric ϕ on Ω, we define the *horizontal subspace* H_ω as the orthogonal complement of V_ω in $T\Omega_\omega$. Thus ϕ decomposes $T\Omega_\omega$ as the orthogonal direct sum

$$T\Omega_\omega = V_\omega \oplus H_\omega \tag{3.1}$$

and this decomposition varies smoothly with ω. So given any tangent vector X to $\mathscr{K}$ at $\pi(\omega)$ we can define the *horizontal lift* $\bar{X}$ of X at ω as the unique element $\bar{X}$ of H_ω satisfying $\pi_*(\bar{X}) = X$. See Figure 1. A mapping which ascribes to each point ω of Ω a complementary subspace H_ω to V_ω in $T\Omega_\omega$ is called an (*Ehresmann*) *connection* on the submersion π. See Hermann [(1975), pages 71–86] for an equivalent definition. Any such connection can be used to construct lifts to $T\Omega_\omega$ of vectors in $T\mathscr{K}_{\pi(\omega)}$ and so may give rise to transfer of tensors from Ω to $\mathscr{K}$ as will be described. For simplicity we shall consider only those connections on π defined as before, by a Riemannian metric.

Horizontal lifts can now be used to transfer geometry along π to $\mathscr{K}$. An inner-product $\pi_\omega\phi$ on $T\mathscr{K}_{\pi(\omega)}$ is defined by

$$\pi_\omega\phi(X, Y) = \phi(\bar{X}, \bar{Y}), \tag{3.2}$$

where $\bar{X}, \bar{Y}$ are the horizontal lifts at ω of X, Y in $T\mathscr{K}_{\pi(\omega)}$. Let

$$\begin{bmatrix} \phi_{\kappa\kappa} & \phi_{\kappa\psi} \\ \phi_{\psi\kappa} & \phi_{\psi\psi} \end{bmatrix}$$

be the partitioned matrix of the expression for ϕ in terms of the local coordinates

$\kappa^1, \ldots, \kappa^p, \psi^1, \ldots, \psi^q$ on Ω. Then it is readily shown that the corresponding matrix for $\pi_\omega \phi$ is

$$\phi_{\kappa\kappa\cdot\psi} = \phi_{\kappa\kappa} - \phi_{\kappa\psi}\phi_{\psi\psi}^{-1}\phi_{\psi\kappa}. \tag{3.3}$$

In general, the inner-product $\pi_\omega \phi$ depends on the point ω to which the vectors X and Y are lifted. In the case when $\pi_\omega \phi$ depends only on $\pi(\omega)$ we obtain a well-defined Riemannian metric on $\mathscr{K}$.

DEFINITION 3.1. If $\pi_{\omega_1}\phi = \pi_{\omega_2}\phi$ whenever $\pi(\omega_1) = \pi(\omega_2)$, then ϕ *transfers* to $\mathscr{K}$ and the *transfer* $\pi_! \phi$ of ϕ is defined by

$$\pi_! \phi(\kappa) = \pi_\omega \phi \quad \text{for any } \omega \in \pi^{-1}(\kappa).$$

If ϕ transfers to $\mathscr{K}$, then π is a Riemannian submersion from (Ω, ϕ) to $(\mathscr{K}, \pi_! \phi)$ [see O'Neill (1966)].

One can also consider transfer of tensor fields from Ω to $\mathscr{K}$. For A in $\mathscr{T}_s{}^r(\Omega)$ and ω in Ω, the corresponding tensor $\pi_\omega A$ at $\pi(\omega)$ is defined by

$$(\pi_\omega A)(X_1, \ldots, X_s) = \left(\otimes^r \pi_*\right)\left(A(\overline{X}_1, \ldots, \overline{X}_s)\right), \tag{3.4}$$

where $X_i \in T\mathscr{K}_{\pi(\omega)}$, $\overline{X}_i$ is the horizontal lift to ω of X_i for $i = 1, \ldots, s$ and $\otimes^r \pi_*$: $\otimes^r T\Omega_\omega \to \otimes^r T\mathscr{K}_{\pi(\omega)}$ is the r-fold tensor product of the tangent map of π.

For example, if A is a $(0,3)$-tensor field on Ω, then calculation shows that

$$\pi_\omega A = A_{\kappa\kappa\kappa} - [3]\phi_{\kappa\psi}\phi_{\psi\psi}^{-1}A_{\psi\kappa\kappa} + [3]\left(\phi_{\kappa\psi}\phi_{\psi\psi}^{-1}\right)^2 A_{\psi\psi\kappa} - \left(\phi_{\kappa\psi}\phi_{\psi\psi}^{-1}\right)^3 A_{\psi\psi\psi},$$

where the notation is similar to that explained after equation (4.3). This case together with that of (3.3) are the ones of primary interest in this paper.

DEFINITION 3.2. If $\pi_{\omega_1} A = \pi_{\omega_2} A$ whenever $\pi(\omega_1) = \pi(\omega_2)$, then A *transfers* to $\mathscr{K}$ and the *transfer* $\pi_! A$ of A is defined by

$$\pi_! A(\kappa) = \pi_\omega A \quad \text{for any } \omega \in \pi^{-1}(\kappa).$$

Note that if $\pi_! A$ is defined it is an (r, s)-tensor field on $\mathscr{K}$.

The transfer takes a particularly simple form if the local coordinates ψ are orthogonal to κ. Here is a case of some statistical interest.

THEOREM 3.1. *Let ϕ be a Riemannian metric and S a $(0,3)$-tensor on Ω. Suppose that the local coordinates κ and ψ are orthogonal, i.e., $\phi_{\kappa\psi}(\omega) = 0$ for $\omega \in \Omega$. Suppose also that $\phi_{\kappa\kappa}(\omega)$ and $S_{\kappa\kappa\kappa}(\omega)$ depend only on $\pi(\omega)$. Then ϕ and S transfer to $\mathscr{K}$ and*

$$\pi_! \phi(\kappa) = \phi_{\kappa\kappa}(\omega), \qquad \pi_! S(\kappa) = S_{\kappa\kappa\kappa}(\omega) \quad \text{for } \omega \in \pi^{-1}(\kappa).$$

The proof is immediate.

There is an alternative construction of the transfer which is of interest. It is based on the following identification of tangent vectors of Ω with cotangent vectors of Ω by means of the Riemannian metric ϕ. A tangent vector X to Ω at ω can be identified with the cotangent vector (one-form) $X^\flat$ defined by

$$X^\flat(Y) = \phi(X, Y), \qquad Y \in T\Omega_\omega.$$

Thus there is an invertible linear mapping $\flat$: $T_0^1(\Omega)_\omega \to T_1^0(\Omega)_\omega$. In terms of local coordinates, the tangent vector with components a^i is mapped by $\flat$ into the cotangent vector with components $\phi_{ij}a^j$, i.e., the mapping $\flat$ is the traditional "lowering of indices." The inverse of $\flat$ is $\sharp$: $T_1^0(\Omega)_\omega \to T_0^1(\Omega)_\omega$, corresponding to "raising of indices" in local coordinates. More generally, ϕ gives rise to the invertible linear transformation $\flat$: $T_0^{r+s}(\Omega)_\omega \to T_s^r(\Omega)_\omega$ defined by

$$\begin{aligned}(X_1 \otimes \cdots \otimes X_r \otimes Y_1 \otimes \cdots \otimes Y_s)^\flat(Z_1 \otimes \cdots \otimes Z_s) \\ = \left\{\prod_{i=1}^{s} \phi(Y_i, Z_i)\right\} X_1 \otimes \cdots \otimes X_r\end{aligned}$$

for $X_1, \dots, X_r, Y_1, \dots, Y_s, Z_1, \dots, Z_s \in T\Omega_\omega$. In local coordinates the $(r+s, 0)$-tensor with components $a^{i_1 \cdots i_r j_1 \cdots j_s}$ is mapped by $\flat$ to the (r, s)-tensor with components $\phi_{j_1 k_1} \cdots \phi_{j_s k_s} a^{i_1 \cdots i_r j_1 \cdots j_s}$. Note that as ω varies in Ω the transformations $\flat$: $T_0^{r+s}(\Omega)_\omega \to T_s^r(\Omega)_\omega$ combine to give an invertible linear mapping $\flat$: $\mathscr{T}_0^{r+s}(\Omega) \to \mathscr{T}_s^r(\Omega)$ of tensor fields. The inverse of $\flat$ is denoted by $\sharp$. Note also that any positive-definite symmetric (2, 0)-tensor field is dual to a Riemannian metric and so gives rise to raising and lowering operators. Now the tangent map π_* of π induces a linear mapping "in the direction of π," π_*: $T_0^r(\Omega)_\omega \to T_0^r(\mathscr{K})_{\pi(\omega)}$. [For example, if Ω is the parameter space of a statistical model and π: $\Omega \to \mathscr{K}$ is the mapping to the parameter of interest, then π_*: $T_0^2(\Omega)_\omega \to T_0^2(\mathscr{K})_{\pi(\omega)}$ sends the formation matrix of $\hat{\omega}$ to that of $\hat{\kappa}$.] In coordinate terms, $a^{i_1 \cdots i_r}$ is mapped to

$$\frac{\partial \omega^{i_1}}{\partial \kappa^{k_1}} \cdots \frac{\partial \omega^{i_r}}{\partial \kappa^{k_r}} a^{k_1 \cdots k_r}.$$

[Note that, unless π is injective, π_* does not give rise to a corresponding linear mapping $\mathscr{T}_0^r(\Omega) \to \mathscr{T}_0^r(\mathscr{K})$ of tensor fields.] We can now define a linear mapping from $T_s^r(\Omega)_\omega$ to $T_s^r(\mathscr{K})_{\pi(\omega)}$ by means of the diagram

$$\begin{tikzcd} T_s^r(\Omega)_\omega \arrow[r, "\sharp"] & T_0^{r+s}(\Omega)_\omega \arrow[d, "\pi_*"] \\ T_s^r(\mathscr{K})_{\pi(\omega)} & T_0^{r+s}(\mathscr{K})_{\pi(\omega)} \arrow[l, "\flat"'] \end{tikzcd}$$

where $\flat$ is defined using the (2, 0)-tensor $\pi_*(\phi(\omega)^\sharp)$. It can be shown that this mapping is in fact π_ω as defined by (3.4). For example, following the matrix of the Riemannian metric around the preceding diagram in the case $r = 0$, $s = 2$, we have

$$\begin{tikzcd} & \phi \arrow[r, mapsto] & \phi^{-1} \arrow[d, mapsto] \\ \phi_{\kappa\kappa\cdot\psi} = \left((\phi^{-1})_{\kappa\kappa}\right)^{-1} & & (\phi^{-1})_{\kappa\kappa} \arrow[ll, mapsto] \end{tikzcd}$$

as in (3.3).

Horizontal lifts can be used to transfer affine connections as well as tensors. Given an affine connection ∇ on Ω we can define $\pi_\omega \nabla$ by

$$(\pi_\omega \nabla)_X Y = \pi_*(\nabla_{\overline{X}} \overline{Y}(\omega)), \tag{3.5}$$

where X, Y are vector fields in a neighbourhood of κ and $\overline{X}, \overline{Y}$ are their horizontal lifts to a neighbourhood of $\omega \in \pi^{-1}(\kappa)$. If the right-hand side of (3.5) depends on ω only through κ, then ∇ *transfers* to $\mathscr{K}$ and the *transfer* $\pi_! \nabla = \pi_\omega \nabla$ is an affine connection on $\mathscr{K}$. Simple calculations prove the following result which is useful in the statistical context.

THEOREM 3.2. *Let ϕ and ∇ be, respectively, a Riemannian metric and an affine connection on Ω.*

(i) *If ∇ transfers to $\mathscr{K}$, then so does its torsion tensor and we have*

$$\mathrm{Tor}(\pi_! \nabla) = \pi_!(\mathrm{Tor}\, \nabla),$$

where Tor *denotes the torsion tensor of a connection.*

(ii) *If ϕ and ∇ transfer to $\mathscr{K}$, then*

$$(\pi_! \nabla)(\pi_! \phi) = \pi_!(\nabla \phi).$$

(iii) *If ∇ is the Riemannian (Levi-Civita) connection of ϕ and ϕ transfers, then ∇ also transfers and $\pi_! \nabla$ is the Riemannian connection of $\pi_! \phi$.*

The general relationship between the Riemannian curvature tensor $R(\pi_! \nabla)$ of $\pi_! \nabla$ and the transfer $\pi_! R(\nabla)$ of the curvature tensor of ∇ is not simple. In the case where ∇ is a Riemannian connection, formulas analogous to the Gauss–Codazzi equations of an immersion are given by O'Neill (1966). Moreover, these formulas yield the following result.

THEOREM 3.3. *Let ∇ be a Riemannian connection which transfers to $\mathscr{K}$. Then the sectional curvatures of $\pi_! \nabla$ are greater than or equal to the corresponding sectional curvatures of ∇. Also the following are equivalent*:

(i) *equality of sectional curvatures*;
(ii) $R(\pi_! \nabla) = \pi_! R(\nabla)$;
(iii) *the local coordinates κ, ψ can be chosen to be orthogonal.*

From the statistical viewpoint, the differential-geometric objects of interest are *statistical manifolds* as defined in Section 2, i.e., a statistical manifold is a triple (Ω, ϕ, S), where Ω is a smooth manifold, ϕ is a Riemannian metric and S is a symmetric $(0, 3)$-tensor on Ω. The kit (ϕ, S) then determines a one-parameter family $\overset{\alpha}{\nabla}$ of torsion-free affine connections on Ω by

$$\overset{\alpha}{\nabla} \phi = \alpha S, \qquad \alpha \in \mathbb{R}.$$

If the kit is defined as the expected geometry of a statistical model, then the connections $\overset{\alpha}{\nabla}$ are those introduced by Amari (1982a); cf. Section 2.

In the rest of this paper we shall consider the following statistical manifolds determined by a statistical model with parameter space Ω: (i) the expected statistical manifold (Ω, i, D) and (ii) the observed statistical manifold $(\Omega, \not{j}, \not{T})$. Given a manifold $\mathscr{K}$ of subparameters of Ω we shall investigate in the following sections statistical conditions under which (i, D) and $(\not{j}, \not{T})$ transfer to $\mathscr{K}$. It follows from Theorem 3.2 that if (i, D) transfers, then so do the α-connections $\overset{\alpha}{\nabla}$ and

$$\left(\pi_! \overset{\alpha}{\nabla}\right)(\pi_! i) = \alpha \pi_! D.$$

Similarly, if $(\not{j}, \not{T})$ transfers, then so do $\overset{\alpha}{\not{\nabla}}$ and

$$\left(\pi_! \overset{\alpha}{\not{\nabla}}\right)(\pi_! \not{j}) = \alpha \pi_! \not{T}.$$

We shall also compare these geometrically transferred statistical manifolds on $\mathscr{K}$ with some others determined by purely statistical considerations.

4. Profile geometries, L-sufficiency and the transfer. The expected and observed statistical manifolds determined by a statistical model $(\mathscr{X}, p(x;\omega), \Omega)$ are defined in terms of the likelihood $L(\omega) = L(\omega; x)$. Thus to define corresponding statistical manifolds on the manifold $\mathscr{K}$ of interest parameters, it is natural to consider the profile likelihood for the interest parameter κ. This is defined by (2.1).

Geometrical quantities on $\mathscr{K}$ can be constructed using $\tilde{l} = \log \tilde{L}$. The *observed profile information* on κ is the matrix

$$\tilde{j}(\kappa) = -\frac{\partial^2 \tilde{l}(\kappa)}{\partial \kappa \, \partial \kappa}$$

and the inverse of this is the *observed profile formation*. The *observed profile skewness* $\tilde{T}$ is defined by

$$\tilde{T}_{ijk}(\kappa) = -\left(\frac{\partial^3 \tilde{l}}{\partial \kappa^i \, \partial \kappa^j \, \partial \kappa^k} + \frac{\partial^2 \tilde{l}}{\partial \kappa^i \, \partial \kappa^j} \frac{\partial \tilde{l}}{\partial \kappa^k} + \frac{\partial^2 \tilde{l}}{\partial \kappa^j \, \partial \kappa^k} \frac{\partial \tilde{l}}{\partial \kappa^i} + \frac{\partial^2 \tilde{l}}{\partial \kappa^k \, \partial \kappa^i} \frac{\partial \tilde{l}}{\partial \kappa^j}\right).$$

Note that, in general, the observed profile objects $\tilde{j}$ and $\tilde{T}$ are not tensors. On the other hand, the observed objects $\not{j}$ and $\not{T}$ defined later in this section are tensors. As noted by Richards (1961) and Patefield (1977), the observed profile formation is equal to the corresponding part of the (full) observed formation, in symbols

$$\tilde{j}(\kappa)^{-1} = j^{\kappa\kappa}(\kappa, \hat{\psi}_\kappa), \tag{4.1}$$

where $j^{\kappa\kappa}$ is the $\kappa\kappa$-part of the observed formation j^{-1} and $\hat{\psi}_\kappa$ is defined by (2.2). Alternatively, (4.1) may be expressed as

$$\tilde{j}(\kappa) = -\left(\frac{\partial^2 l}{\partial \kappa^2} + [2]\frac{\partial \hat{\psi}_\kappa}{\partial \kappa} \frac{\partial^2 l}{\partial \psi \, \partial \kappa} + \left(\frac{\partial \hat{\psi}_\kappa}{\partial \kappa}\right)^2 \frac{\partial^2 l}{\partial \psi^2}\right)(\kappa, \hat{\psi}_\kappa). \tag{4.2}$$

Similarly, it can be shown that

$$\begin{aligned} \frac{\partial^3 \tilde{l}}{\partial \kappa^3} = \Bigg(&\frac{\partial^3 l}{\partial \kappa^3} + [3]\frac{\partial \hat{\psi}_\kappa}{\partial \kappa} \frac{\partial^3 l}{\partial \psi \, \partial \kappa^2} + [3]\left(\frac{\partial \hat{\psi}_\kappa}{\partial \kappa}\right)^2 \frac{\partial^3 l}{\partial \psi^2 \, \partial \kappa} \\ &+ \left(\frac{\partial \hat{\psi}_\kappa}{\partial \kappa}\right)^3 \frac{\partial^3 l}{\partial \psi^3}\Bigg)(\kappa, \hat{\psi}_\kappa), \end{aligned} \tag{4.3}$$

where in this compressed notation, e.g., $[3](\partial\hat{\psi}_\kappa/\partial\kappa)(\partial^3 l/\partial\psi\,\partial\kappa^2)$ has components

$$\left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)_{ir}\frac{\partial^3 l}{\partial\psi^r\,\partial\kappa^j\,\partial\kappa^k}+\left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)_{jr}\frac{\partial^3 l}{\partial\psi^r\,\partial\kappa^k\,\partial\kappa^i}+\left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)_{kr}\frac{\partial^3 l}{\partial\psi^r\,\partial\kappa^i\,\partial\kappa^j}.$$

Define $\hat{\omega}_\kappa\colon \mathscr{K}\to\Omega$ by $\hat{\omega}_\kappa(\kappa)=(\kappa,\hat{\psi}_\kappa)$. Then the tangent map $\hat{\omega}_{\kappa*}$ of $\hat{\omega}_\kappa$ can be used to lift tangent vectors from $\mathscr{K}$ to Ω. Indeed, we can regard $(\partial/\partial\kappa^i)+(\partial\hat{\psi}_\kappa/\partial\kappa)_{ir}(\partial/\partial\psi^r)$ as the "horizontal lift" at $(\kappa,\hat{\psi}_\kappa)$ of $\partial/\partial\kappa^i$. (Note that this construction does not make use of any connection on π or of any metric tensor on Ω.) Then we can summarize (4.2) and (4.3) as follows: For $r=2,3$, the rth derivative of the profile log-likelihood is equal to the rth derivative of the full log-likelihood at the partial maximum likelihood estimate acting on horizontal lifts. This property does not hold for $r=4$.

The "infinite-sample" analogue of log-likelihood, the expected log-likelihood, is closely related to the discrimination information (or Kullback–Leibler distance)

$$I(\omega_1,\omega_2)=\int\log\left\{\frac{p(x;\omega_1)}{p(x;\omega_2)}\right\}p(x;\omega_1)\,d\mu(x).$$

The function I determines the expected kit (i,D) by

$$i(\omega)=\left.\frac{\partial^2 I(\omega,\omega')}{\partial\omega'\,\partial\omega'}\right|_{\omega'\mapsto\omega} \tag{4.4}$$

and

$$D(\omega)=\left[\frac{\partial^3 I(\omega,\omega')}{\partial\omega\,\partial\omega'\,\partial\omega'}-\frac{\partial^3 I(\omega,\omega')}{\partial\omega'\,\partial\omega\,\partial\omega}\right]_{\omega'\mapsto\omega'}, \tag{4.5}$$

where in this compressed notation, e.g., $\partial^3 I(\omega,\omega')/\partial\omega\,\partial\omega'\,\partial\omega'$ indicates a three-dimensional array with components $\partial^3 I(\omega,\omega')/\partial\omega^i\,\partial\omega'^j\,\partial\omega'^k$. This suggests the definition of the *profile discrimination information* $\tilde{I}$ by

$$\tilde{I}(\kappa_1,\kappa_2)=\inf_{(\omega_1,\omega_2)|(\kappa_1,\kappa_2)} I(\omega_1,\omega_2). \tag{4.6}$$

It is shown in Theorem 7.2 that for composite transformation models the $\tilde{I}$ analogues of (4.4) and (4.5) equal the transfer $\pi_!(i,D)$ of (i,D).

In order to define geometries (i.e., statistical manifolds) on $\mathscr{K}$ based on profile likelihood, it is useful to have the concept of L-sufficiency.

DEFINITION 4.1. A statistic $u\colon \mathscr{X}\to\mathscr{Y}$ is *L-sufficient* for κ if both

(4.7) $\quad \dfrac{\partial\tilde{l}}{\partial\kappa}(\kappa;x)$ depends on the observation x only through u

and

(4.8) $\quad$ the distribution of u depends only κ.

If (4.7) holds, then u is *weakly L-sufficient* for κ.

Weak L-sufficiency was introduced by Rémon (1984) under the name of L-sufficiency. The preceding definition of L-sufficiency fits into the nomenclature of sufficiency established in Barndorff-Nielsen (1978a). Indeed, Rémon showed that L-sufficiency is a generalization both of S-sufficiency and of G-sufficiency. Whether an L-sufficient statistic can always be said to contain the entire available information on the interest parameter is a matter for further investigation, but we shall not pursue this question in the present paper.

Given an L-sufficient statistic for κ we can define the expected profile statistical manifold $(\mathscr{K}, \tilde{i}, \tilde{D})$. This is determined by the *expected profile information*

$$\tilde{i}(\kappa) = V_\omega\left(\frac{\partial \tilde{l}}{\partial \kappa}\right), \qquad \omega \in \pi^{-1}(\kappa),$$

and the *expected profile skewness tensor* $\tilde{D}$ with components

$$D_{ijk} = E_\omega\left\{\left(\frac{\partial \tilde{l}}{\partial \kappa^i} - \nu_i\right)\left(\frac{\partial \tilde{l}}{\partial \kappa^j} - \nu_j\right)\left(\frac{\partial \tilde{l}}{\partial \kappa^k} - \nu_k\right)\right\}, \qquad \omega \in \pi^{-1}(\kappa),$$

where $\nu = E_\omega(\partial \tilde{l}/\partial \kappa)$. Note that (4.8) ensures that $\tilde{i}(\kappa)$ and $\tilde{D}(\kappa)$ depend only on κ.

Profile likelihood gives rise also to observed geometries on $\mathscr{K}$. Let u: $\mathscr{X} \to \mathscr{Y}$ be minimal weakly L-sufficient for κ and let b: $\mathscr{Y} \to B$ be an auxiliary statistic such that $(b, \hat{\kappa})$: $\mathscr{Y} \to B \times \mathscr{K}$ is bijective. Then by analogy with the definition of observed kits (cf. Section 2), we define the *observed profile kit* as the pair $(\tilde{\jmath}, \tilde{T})$ consisting of the *observed profile metric* $\tilde{\jmath}$ and the *observed profile skewness tensor* $\tilde{T}$ given by

$$\tilde{\jmath}_{ij} = \tilde{l}_{i;\,j}$$

and

$$\tilde{T}_{ijk} = -\left(\tilde{l}_{ijk} + \tilde{l}_{ij;\,k} + \tilde{l}_{jk;\,i} + \tilde{l}_{ki;\,j}\right).$$

Thus, given the observed value of the auxiliary b we obtain a statistical manifold structure $(\tilde{\jmath}, \tilde{T})$ on $\mathscr{K}$. In the presence of weak L-sufficiency these observed profile geometries are the transfers of observed geometries on Ω.

THEOREM 4.1. *If u: $\mathscr{X} \to \mathscr{Y}$ is weakly L-sufficient for κ and a: $\mathscr{X} \to A$, b: $\mathscr{Y} \to B$ are such that*

(i) *$(a, \hat{\omega})$: $\mathscr{X} \to A \times \Omega$ is bijective and sufficient for ω,*
(ii) *$(b, \hat{\kappa})$: $\mathscr{Y} \to B \times \mathscr{K}$ is bijective,*
(iii) *$b(u)$ depends on x only through a,*

then $(\jmath, T)$ transfers and $\pi_!(\jmath, T) = (\tilde{\jmath}, \tilde{T})$.

PROOF. Differentiation of $\tilde{l}_i = 0$ gives

$$\tilde{l}_{ij} + \tilde{l}_{i;\,j} = 0.$$

Using (4.2) we obtain

$$\tilde{\jmath}_{ij}(\kappa) = (\jmath_{ij} - \jmath_{ir}\jmath^{rs}\jmath_{sj})(\kappa, \hat{\psi}_\kappa),$$

where $[\jmath^{rs}]$ denotes the inverse of the $\psi\psi$-part of $[\jmath_{ij}]$ and so, as this expression depends on $\hat{\psi}_\kappa$ only through u, $\jmath$ transfers and

$$\tilde{\jmath}(\kappa) = (\pi_!\jmath)(\kappa).$$

Also, differentiation with respect to $\hat{\kappa}$ of

$$l_{\kappa\kappa} + \frac{\partial\hat{\psi}_\kappa}{\partial\kappa} l_{\psi\kappa} = \tilde{l}_{\kappa\kappa}$$

[which is a version of (4.2)] and

$$l_{\kappa\psi} + \frac{\partial\hat{\psi}_\kappa}{\partial\kappa} l_{\psi\psi} = 0$$

together with some manipulation leads to

$$\tilde{\jmath}_{\kappa\kappa;\,\kappa} = \jmath_{\kappa\kappa;\,\kappa} + \frac{\partial\hat{\psi}_\kappa}{\partial\kappa}([2]\jmath_{\psi\kappa;\,\kappa} + \jmath_{\kappa\kappa;\,\psi})$$
$$+ \left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)^2 (\jmath_{\psi\psi;\,\kappa} + [2]\jmath_{\psi\kappa;\,\psi}) + \left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)^3 \jmath_{\psi\psi;\,\psi}.$$

Using this and (4.3) we find that $\mathcal{T}$ transfers and

$$\tilde{\mathcal{T}} = \pi_!\mathcal{T}. \qquad \square$$

This theorem leads to the conjecture that if u is an L-sufficient statistic, then (i, D) transfers. At present we do not know whether or not this holds in general. The inverse Gaussian example discussed after Theorem 5.2 shows that we do not always have $\pi_! D = \tilde{D}$, in contrast to the preceding results for the observed kit. As indicated briefly in the following text and described in greater detail in Theorem 7.2, for composite transformation models (i, D) does indeed transfer to $\mathscr{K}$. The most general results we have on L-sufficiency and the transfer of expected geometry involve repeated sampling. We shall use a subscript n to denote quantities based on a random sample of size n from the distribution parameterized by ω. For example $\tilde{j}_n$ and $\tilde{T}_n$ denote observed profile information and skewness for such a sample. The next theorem shows that if the transferred expected geometry exists, then it is the large-sample limit of the average observed profile geometry.

Theorem 4.2. *If (i, D) transfers to $\mathscr{K}$, then with probability* 1,

$$n^{-1}(\tilde{j}_n(\kappa), \tilde{T}_n(\kappa)) \to \pi_!(i, D)(\kappa)$$

as $n \to \infty$.

Proof. For random samples of size n from the distribution parameterized by ω we have from (4.2),

$$n^{-1}\tilde{j}_n(\kappa) = \left((n^{-1}j_n)_{\kappa\kappa} - (n^{-1}j_n)_{\kappa\psi}(n^{-1}j_n)^{\psi\psi}(n^{-1}j_n)_{\psi\kappa}\right)(\kappa, \hat{\psi}_\kappa)$$
$$\to \left(i_{\kappa\kappa} - i_{\kappa\psi}i^{\psi\psi}i_{\psi\kappa}\right)(\omega) = \pi_! i(\kappa),$$

where $\kappa = \pi(\omega)$. Similarly, (4.3) yields

$$n^{-1}\tilde{T}_n(\kappa)$$
$$= n^{-1}\left\{\left(T_{\kappa\kappa\kappa} + [3]\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}T_{\psi\kappa\kappa} + [3]\left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)^2 T_{\psi\psi\kappa} + \left(\frac{\partial\hat{\psi}_\kappa}{\partial\kappa}\right)^3 T_{\psi\psi\psi}\right)(\kappa, \hat{\psi}_\kappa)\right\}_n$$
$$\to \left(D_{\kappa\kappa\kappa} + [3](-i_{\kappa\psi}i^{\psi\psi})D_{\psi\kappa\kappa} + [3](-i_{\kappa\psi}i^{\psi\psi})^2 D_{\psi\psi\kappa} + (-i_{\kappa\psi}i^{\psi\psi})^3 D_{\psi\psi\psi}\right)(\omega)$$
$$= \pi_! D(\kappa). \qquad \square$$

We shall now, further, show that if there is L-sufficiency for all sample sizes then the expected geometry given by (i, D) transfers to $\mathscr{K}$.

Theorem 4.3. *Suppose that for all sample sizes n there exists a statistic u_n: $\mathscr{X}^n \to \mathscr{Y}_n$ which is L-sufficient for κ. Then (i, D) transfers to $\mathscr{K}$.*

Proof. The equation

$$n^{-1/2}\frac{\partial l_n}{\partial\omega} = n^{1/2}(\hat{\omega} - \omega)\{i(\omega) + o_p(1)\}$$

leads to the well-known result

$$i(\omega) = \left\{\lim_{n\to\infty} nV_\omega(\hat{\omega})\right\}^\flat$$

and to

$$D(\omega) = \left\{\lim_{n\to\infty} n^2 E_\omega\left[(\hat{\omega} - \omega)^3\right]\right\}^\flat,$$

where $\flat$ denotes "lowering of indices" as described in Section 3 and $E_\omega[(\hat{\omega} - \omega)^3]$ denotes the three-dimensional array whose (i, j, k)-entry is $E_\omega[(\hat{\omega} - \omega)^i(\hat{\omega} - \omega)^j(\hat{\omega} - \omega)^k]$. Using the alternative construction of the transfer given in Section 3 we see that $\pi_! i$ and $\pi_! D$ are given by

$$\pi_! i(\kappa) = \left\{\lim_{n\to\infty} nV_\omega(\hat{\kappa})\right\}^\flat, \tag{4.9}$$

$$\pi_! D(\kappa) = \left\{\lim_{n\to\infty} n^2 E_\omega\left[(\hat{\kappa} - \kappa)^3\right]\right\}^\flat, \tag{4.10}$$

provided that the right-hand sides depend only on κ. L-sufficiency of u_n ensures by (4.7) that $\hat{\kappa}$ depends on $(x_1, \dots, x_n)$ only through $u_n(x_1, \dots, x_n)$ and by (4.8) that the distribution of $\hat{\kappa}$ depends only on κ. Thus $\pi_!(i, D)$ is defined. $\square$

In particular, for a composite transformation model with acting group G the maximal invariants u_n: $\mathscr{X}^n \to \mathscr{X}^n/G$ are L-sufficient for the maximal invariant parameter κ and so (i, D) transfers to $\mathscr{K} = \Omega/G$.

The transferred expected geometry $\pi_!(i, D)$ has been defined in purely differential-geometric terms. We do not have a general purely statistical description of it—apart from that given by (4.9) and (4.10). However, it is shown in Theorem 7.2 that for composite transformation models $\pi_!(i, D)$ yields the same statistical manifold on Ω/G as that derived from $\tilde{I}$, which is also the large-sample limit of the marginal geometry.

We have in this section considered four types of kits on the interest parameter manifold $\mathscr{K}$, namely the transferred kits $\pi_!(i, D)$ and $\pi_!(j, \not{T})$ and the profile kits $(\tilde{i}, \tilde{D})$ and $(\tilde{j}, \tilde{\not{T}})$, and we have discussed the relevance of the concept of L-sufficiency in the present context. Suppose a minimal L-sufficient statistic u for κ exists and let $\check{l}(\kappa)$ be the marginal log-likelihood function for κ based on u, i.e., the log-likelihood function of the marginal model for u. In general, on $\mathscr{K}$ one would ideally wish to work with the geometrical structures determined by the expected kit and the observed kit given in terms of the marginal model for u; we shall denote those kits by $(\check{i}, \check{D})$ and $(\check{j}, \check{\not{T}})$, respectively. However, these marginal structures are often intractable both analytically and numerically; then the transferred structures and the profile structures come into the picture as possible approximations to the marginal structures. From this viewpoint it would be natural also to study *modified profile geometries*, defined in analogy with the profile geometries but from the modified profile likelihood $\overset{\dagger}{L}(\kappa)$ [introduced in Barndorff-Nielsen (1983)] rather than from $\tilde{L}(\kappa)$. In considerable generality it is, in fact, possible to define modified profile kits $(\overset{\dagger}{i}, \overset{\dagger}{D})$ and $(\overset{\dagger}{j}, \overset{\dagger}{\not{T}})$ and these will as a rule approximate $(\check{i}, \check{D})$ and $(\check{j}, \check{\not{T}})$ to a higher degree of accuracy than the transferred or profile kits; cf. Barndorff-Nielsen (1983, 1985). In Section 7 the relations between these various kits are studied in some detail for composite transformation models; cf. also Section 5.4. To help the overview the kits are set out in Table 1.

Note that if u is any statistic satisfying (4.8), then, since margining does not increase expected information, we have

$$\pi_! i \geq \check{i}, \tag{4.11}$$

i.e., $\pi_! i - \check{i}$ is positive semidefinite.

TABLE 1

Kit = (metric tensor, skewness tensor) specifications, each (when well defined) turning the interest parameter manifold into a statistical manifold.

	Transferred	Profile	Modified Profile	Marginal
Observed	$(\pi_! j, \pi_! \not{T})$	$(\tilde{j}, \tilde{\not{T}})$	$(\overset{\dagger}{j}, \overset{\dagger}{\not{T}})$	$(\check{j}, \check{\not{T}})$
Expected	$(\pi_! i, \pi_! D)$	$(\tilde{i}, \tilde{D})$	$(\overset{\dagger}{i}, \overset{\dagger}{D})$	$(\check{i}, \check{D})$

Kits and their transfers occur also in the wider framework of contrast functions. Let ρ be any smooth contrast function, i.e., ρ: $\Omega \times \Omega \to \mathbb{R}$ satisfies $\rho(\omega, \omega') \geq 0$ with equality if and only if $\omega = \omega'$. Then, as shown by Eguchi (1983), ρ defines a kit (i^{ρ}, D^{ρ}) by

$$i^{\rho}(\omega) = \left. \frac{\partial^2 \rho(\omega, \omega')}{\partial \omega' \, \partial \omega'} \right|_{\omega' \mapsto \omega}$$

and

$$D^{\rho}(\omega) = \left[\frac{\partial^3 \rho(\omega, \omega')}{\partial \omega \, \partial \omega' \, \partial \omega'} - \frac{\partial^3 \rho(\omega, \omega')}{\partial \omega' \, \partial \omega \, \partial \omega} \right]_{\omega' \mapsto \omega}.$$

A "profile" version $\tilde{\rho}$ of ρ is defined by

$$\tilde{\rho}(\kappa, \kappa') = \inf_{(\omega, \omega')|(\kappa, \kappa')} \rho(\omega, \omega').$$

In general, $\tilde{\rho}$ is a smooth contrast function on $\mathscr{K}$ and so gives rise to a kit $(i^{\tilde{\rho}}, D^{\tilde{\rho}})$.

It is possible in this context to define a form of sufficiency which ensures that (i^{ρ}, D^{ρ}) transfers to $\mathscr{K}$ and that the transferred kit is equal to $(i^{\tilde{\rho}}, D^{\tilde{\rho}})$. We omit the details.

5. Some special cases. We now illustrate the general theory discussed in the foregoing sections by considering four special cases. Each case concerns a general class of models, these classes being:

1. models with cuts, including exponential models with θ-parallel foliations;
2. exponential models with τ-parallel foliations, including reproductive exponential models;
3. extended generalized linear models;
4. composite transformation models.

5.1. *Cuts.* Perhaps the simplest examples of L-sufficient statistics are cuts. Recall [Barndorff-Nielsen (1978a)] that u: $\mathscr{X} \to \mathscr{Y}$ is a cut if $\Omega = \mathscr{K} \times \Psi$ and

$$l(\kappa, \psi; x) = \check{l}(\kappa; u) + l(\psi; x|u)$$

with the marginal likelihood $\check{l}(\kappa; u)$ depending on ω only through κ and the conditional likelihood $l(\psi; x|u)$ depending on ω only through ψ. In this case we have $(\partial^r l/\partial \kappa^r)(\kappa, \psi; x) = (\partial^r \check{l}/\partial \kappa^r)(\kappa; u)$ and so for the expected geometry on Ω:

(i) κ, ψ are orthogonal.
(ii) $\pi_! i = \tilde{i} = \check{i} = i_{\kappa\kappa}$, $\pi_! D = \tilde{D} = \check{D} = D_{\kappa\kappa\kappa}$.

There is a similar result for observed geometries. Using Theorem 4.1 we have that if u is a cut, then:

(i) κ, ψ are orthogonal for $\not{j}$.
(ii) $\pi_! \not{j} = \tilde{\not{j}} = \check{\not{j}} = \not{j}_{\kappa\kappa}$, $\pi_! \not{T} = \tilde{\not{T}} = \check{\not{T}} = \not{T}_{\kappa\kappa\kappa}$.

For full steep exponential models there is an intimate connection between cuts and the existence of transferred geometries. The precise statement is given in Theorem 5.1.

Consider a full exponential model on a sample space $\mathscr{X}$ and with minimal exponential representation

$$p(x;\theta) = a(\theta)b(x)e^{\theta\cdot t(x)}, \tag{5.1}$$

where θ and $t = t(x)$ are vectors of dimension k. We denote the domain of variation for the canonical parameter θ by Θ and the closed convex hull of the marginal distribution of the canonical statistic t by C. Furthermore, for $\theta \in \operatorname{int}\Theta$ (the interior of Θ) we let $\tau = \tau(\theta) = E_\theta t$, i.e., τ is the mean value parameter, and we use the notation $\mathscr{T}$ for the set of mean values $\tau(\operatorname{int}\Theta)$. The model is assumed to be steep which is equivalent to $\mathscr{T} = \operatorname{int} C$; cf. Barndorff-Nielsen (1978a), Theorem 9.2. This is the case, in particular, if the canonical parameter domain Θ of (5.1) is open.

Let (t_1, t_2) be a partition of t and let (θ_1, θ_2) and (τ_1, τ_2) be the corresponding partitions of θ and τ. We denote the common dimension of t_1, θ_1 and τ_1 by d. For $\tau \in \mathscr{T}$ we may express τ as a function of the mixed parameter (θ_1, τ_2). If this expression is of the form

$$\tau_1 = -\tau_2 h(\theta_1) + k(\theta_1) \tag{5.2}$$

for some $(k - d)\times d$ matrix function h and some $1 \times d$ vector function k, then, for reasons given in Barndorff-Nielsen and Blæsild (1983b), the model (5.1) is said to have a θ-parallel foliation. We now characterize those models (5.1) for which the geometries on $\mathscr{T}$ transfer by τ_2.

Theorem 5.1. *The following are equivalent:*

(i) t_2 *is a cut.*

(ii) *The model has a θ-parallel foliation of the form* (5.2).

(iii) *The expected and observed geometries on $\mathscr{T}$ (which coincide) transfer to* $\tau_2(\mathscr{T})$.

If any of the preceding hold, then

$$\pi_! i(\tau_2) = \Sigma_{22}^{-1},$$

$$\pi_! D_{ijk}(\tau_2) = \Sigma_{22}^{ip}\Sigma_{22}^{jq}\Sigma_{22}^{kr}\mu_{pqr},$$

where Σ_{22} is the variance matrix of t_2 and Σ_{22}^{ij} and μ_{pqr} denote the components of Σ_{22}^{-1} and the third central moment of t_2, respectively.

Proof. This follows from Theorems 4.1 and 4.2 of Barndorff-Nielsen and Blæsild (1983b). Transfer of D is ensured by their equation (4.6). □

5.2. *Exponential models with τ-parallel foliations.* A more interesting class of exponential models for which geometries transfer is provided by the exponential models with τ-parallel foliations. These are the models (5.1) for which we can write

$$\theta_2 = -\theta_1 h(\tau_2) + k(\tau_2) \tag{5.3}$$

for some $d \times (k - d)$ matrix function h and some $1 \times (k - d)$ vector function k. As shown in Barndorff-Nielsen and Blæsild (1983b), there exist a $1 \times d$ vector function $H(\tau_2)$ and real functions $K(\tau_2)$ and $M(\theta_1)$ such that $h(\tau_2) = \partial H^T(\tau_2)/\partial \tau_2$ and $k(\tau_2) = \partial K(\tau_2)/\partial \tau_2$ and such that (5.1) can be rewritten as

$$\begin{aligned} p(x; \theta_1, \tau_2) & \\ = b(x)\exp(-M(\theta_1))\exp\big(\theta_1 \cdot \{t_1 - (t_2 - \tau_2)h^T(\tau_2) - H(\tau_2)\} & \\ +(t_2 - \tau_2) \cdot k(\tau_2) + K(\tau_2)\big). & \end{aligned} \tag{5.4}$$

We shall be concerned with those cases where in addition to (5.3) we have that t_2 follows an exponential model with $(H(t_2), t_2)$ as canonical statistic and θ as the corresponding canonical parameter, in symbols

$$t_2 \sim EM\big((H(t_2), t_2), \theta\big). \tag{5.5}$$

This holds in particular if $t_2(x) = x$.

Examples of models satisfying (5.3) and (5.5) and discussions of related issues may be found in Barndorff-Nielsen and Blæsild (1983b, c, 1988a, b).

THEOREM 5.2. *For the exponential model* (5.1) *the following are equivalent*:

(i) *The expected and observed geometries on* int Θ (*which coincide*) *transfer to* θ_1(int Θ).

(ii) *Condition* (5.3) *is satisfied.*

(iii) *There is a continuous* $1 \times d$ *vector function* $H(t_2)$ *such that* $t_1 - H(t_2)$ *is weakly L-sufficient for* θ_1.

If any of these hold, then

$$(\pi_! i)(\theta_1) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

where $\Sigma = V_\theta t$, *the variance matrix of* t. *If in addition* (5.5) *is satisfied, then* $t_1 - H(t_2)$ *is L-sufficient for* θ_1.

PROOF. Equivalence of (i) and (ii) is a consequence of Theorem 5.2 and Lemma 3.3 of Barndorff-Nielsen and Blæsild (1983b) and of Theorem 3.1 and formula (5.4). That (ii) implies (iii) follows from formula (5.4). For the converse, note that (5.1) implies that

$$\tilde{l}(\theta_1; x) = \log b(x) + \theta_1 \cdot t_1 + \gamma(\theta_1; t_2)$$

for some function γ. If $t_1 - H(t_2)$ is weakly L-sufficient for θ_1, then it follows that

$$\frac{\partial \tilde{l}}{\partial \theta_1} = t_1 - H(t_2) - m(\theta_1)$$

and so

$$\tilde{l}(\theta_1; x) = \log b(x) + \theta_1 \cdot \{t_1 - H(t_2)\} - M(\theta_1) + K(t_2)$$

for suitable functions m, M and K. Substituting this equation in $\tilde{l}(\theta_1; x) = l(\theta; x)$, where $t(x) = \tau(\theta)$, and differentiating with respect to τ_2 we obtain (5.3). Finally, if both (5.3) and (5.5) hold, then it is a conclusion of Barndorff-Nielsen and Blæsild (1983c) that the distribution of $t_1 - H(t_2)$ depends only on θ_1. □

In connection with assertion (iii) it should be noted that $t_1 - H(t_2)$ is generally neither S-sufficient nor G-sufficient. For instance, if $x_1, \dots, x_n$ is a sample from the inverse Gaussian distribution

$$p(x; \kappa, \lambda) = (2\pi)^{-1/2} \kappa e^{\kappa\lambda} x^{-3/2} e^{-\{\kappa^2 x^{-1} + \lambda^2 x\}/2}, \tag{5.6}$$

then the assumptions of Theorem 5.2 are satisfied and $\Sigma(x_i^{-1} - \bar{x}^{-1})$ is L-sufficient, but not S- or G-sufficient, with respect to κ.

5.3. *An extended class of generalized linear models.* Model functions of the form

$$p(x; \kappa, \psi) = b(x; \kappa)\exp\{\alpha(\kappa)\varphi(x; \psi)\} \tag{5.7}$$

provide an extension of the class of generalized linear models and have been considered by Barndorff-Nielsen (1983) and by Jørgensen (1983, 1986, 1987). A feature of these models is that κ and ψ are orthogonal under the expected information i.

If $b(x; \kappa)$ in (5.7) factorises so that

$$p(x; \kappa, \psi) = a(x)c(\kappa)\exp\{\alpha(\kappa)\varphi(x; \psi)\}, \tag{5.8}$$

where $a(x)$ and $c(\kappa)$ are functions, respectively, of x alone and of κ alone, then [if $\alpha(\kappa)$ has constant sign] it can be shown that $u(x) = \varphi(x; \hat{\psi})$ is weakly L-sufficient for κ. It follows from Theorem 4.1 that, for suitable auxiliary statistics, observed geometries transfer to $\mathscr{K}$. Although it is not known whether or not u is always L-sufficient for κ, a calculation shows that, for models given by (5.8), $i_{\kappa\kappa}$ and $D_{\kappa\kappa\kappa}$ depend on ω only through κ. Thus the expected kit transfers and

$$\pi_! i = i_{\kappa\kappa}, \qquad \pi_! D = D_{\kappa\kappa\kappa}.$$

5.4. *Composite transformation models.* A composite transformation model is a parameterized statistical model $\mathscr{M} = (\mathscr{X}, p, \Omega)$ in which a group G acts on $\mathscr{X}$ so that Ω is mapped into itself by the induced action of G on the set of all probability measures on $\mathscr{X}$.

We adopt here the definitions and terminology of composite transformation models as discussed in Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982) and Barndorff-Nielsen (1983, 1987); the reader is referred to these papers for details and additional results.

The group G acts on both the sample space $\mathscr{X}$ and the parameter space Ω. We suppose that κ is a maximal invariant under the latter action and that the model function has the property that

$$p(x; \kappa, g\psi) = \chi(g, x)p(g^{-1}x; \kappa, \psi) \tag{5.9}$$

for every $g \in G$ and some function $\chi(g, x)$. Note that $\chi(g, x)$ is the multiplier of μ, in the terminology of Barndorff-Nielsen (1987). All actions, mappings, etc., occurring in the following are assumed to be sufficiently smooth to justify the differentiations, etc., performed. A model satisfying (5.9) is said to be a composite transformation model with index parameter κ. The submodels obtained by fixing κ are (pure) transformation models.

In most cases of interest the model has the following additional structure, after minimal sufficient reduction. There exists a left factorization $G = HK$ of the group G, i.e., there are two subsets H and K of G such that K is a subgroup of G and such that each element g of G can be factorized uniquely as $g = hk$ where $h \in H$ and $k \in K$. The two associated mappings sending g to h and g to k will be denoted by η and ζ. Moreover, it is possible to select on each orbit Gx of $\mathscr{X}$ an orbit representative $u \in Gx$ such that K is the isotropy group of u, i.e., $K = G_u = \{g \in G\colon\ gu = u\}$. Further, the domain of variation for (κ, ψ) is a product set, and the domain Ψ of ψ contains a point ψ_e such that K is also the isotropy group of ψ_e, $K = G_{\psi_e}$. Then, in effect, we may make the identifications $\Psi \equiv H \equiv G/K$ and $\psi_e \equiv e \equiv K$, where e indicates the unit element of G (and also of K). Accordingly, the natural action of G on $G/K = \{gK\colon\ g \in G\}$, i.e., the action given by

$$\begin{gathered} G \times G/K \to G/K, \\ (g', gK) \mapsto g'gK, \end{gathered}$$

may be viewed as an action of G on Ψ. We shall denote this latter action by $\hat{\gamma}$. It is identical to the action of G on Ψ induced from the action of G on $\mathscr{X}$ by $\hat{\psi}$, the maximum likelihood estimator of ψ. If the measure μ on $\mathscr{X}$ is invariant, then the mapping $\hat{\psi}$ also induces an invariant measure ν on Ψ from μ. Finally, the subgroup K is compact.

The composite transformation models to be considered in this paper possess all of the abovementioned properties.

To see that for a composite transformation model (5.9) the maximal invariant u is L-sufficient for κ, one may argue as follows. From (5.9) we find for any $g \in G$,

$$l(\kappa, g\psi; x) = l(\kappa, \psi; g^{-1}x) + \log \chi(g, x) \tag{5.10}$$

and hence

$$\begin{aligned} \tilde{l}(\kappa; x) &= \sup_{\psi|\kappa} l(\kappa, \psi; x) \\ &= \sup_{\psi|\kappa} l(\kappa, g\psi; x) \\ &= \sup_{\psi|\kappa} l(\kappa, \psi; g^{-1}x) + \log \chi(g, x) \\ &= \tilde{l}(\kappa; g^{-1}x) + \log \chi(g, x), \end{aligned}$$

from which (4.7) follows; that (4.8) is fulfilled is well known. [For $\chi(g, x) = 1$ this result was proved by Rémon (1984).]

For very many composite transformation models the action of the group ensures that geometries on Ω transfer nicely to the orbit parameter space $\mathscr{K} = \Omega/G$. This follows from the results of Section 4 together with the L-sufficiency of the maximal invariant u. [For any open submodel of a composite transformation model the geometry transfers. A simple example, considered from a different viewpoint by Rémon (1984), is that of the submodel of the set of $N(\mu, \sigma^2)$ distributions with $\mu > 0$, where the parameter of interest is σ^2 and the data consist of a sample $x_1, \dots, x_n$ from $N(\mu, \sigma^2)$.]

The transferred geometry of composite transformation models is considered at greater length in Section 7.

We conclude the present discussion with a look at coefficients of variation for location-scale models.

EXAMPLE 5.1. Let f be a known probability density function on $\mathbb{R}$ and consider the corresponding location-scale model for samples $x_1, \dots, x_n$ of size n. Then we have

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

Under the action of the multiplicative group of nonnegative real numbers, i.e., under changes of scale, this is a composite transformation model with invariant (interest) parameter $\kappa = \sigma^{-1}\mu$ and group parameter $\psi = \sigma^{-1}$ (say). The configuration of the sample $(x_1, \dots, x_n)$ is

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

Using Example 6.1 of Barndorff-Nielsen (1987) together with the tensorial nature of $\jmath$ and $\mathcal{T}$ we obtain the following results for the observed geometries conditional on the ancillary a:

$$\jmath_{\kappa\kappa}(\kappa, \psi) = \sum_{i=1}^{n} g''(a_i),$$

$$\jmath_{\kappa\psi}(\kappa, \psi) = -\psi^{-1} \sum_{i=1}^{n} (a_i + \kappa) g''(a_i),$$

$$\jmath_{\psi\psi}(\kappa, \psi) = \psi^{-2}\left[n + \sum_{i=1}^{n} (a_i + \kappa)^2 g''(a_i)\right],$$

where $g(x) = -\log f(x)$ and where $g''(x)$ is assumed to be positive, i.e., f is log concave. Thus, using (3.3), we have

$$\pi_! \jmath = \sum_{i=1}^{n} g''(a_i) - \frac{\{\Sigma_{i=1}^{n}(a_i + \kappa) g''(a_i)\}^2}{n + \Sigma_{i=1}^{n}(a_i + \kappa)^2 g''(a_i)}.$$

We also have

$$\pi_! \mathcal{T} = \frac{6n\kappa(1 + V(a))\{n^{-1}\Sigma_{i=1}^{n}(a_i + \kappa) g''(a_i)\}^2}{\{1 + n^{-1}\Sigma_{i=1}^{n}(a_i + \kappa)^2 g''(a_i)\}^3},$$

where $V(a) = n^{-1}\Sigma_{i=1}^{n}(a_i - \tilde{a})^2 g''(a_i)$ and $\tilde{a} = n^{-1}\Sigma_{i=1}^{n} a_i g''(a_i)$. The maximal invariant can be taken as u, where $u(x_1, \dots, x_n) = (s^{-1}x_1, \dots, s^{-1}x_n)$, s^2 denoting the sample variance. Defining the statistic b by $b(u(x)) = a$, we see that the conditions of Theorem 4.1 are satisfied. It follows that $\pi_!(\not{j}, \not{T}) = (\tilde{\not{j}}, \tilde{\not{T}})$. This equation holds for general composite transformation models as shown in Theorem 7.1.

Similarly, for the transferred expected geometries we have

$$\pi_! i = m_0 - \frac{m_1^2}{1 + m_2},$$

$$\pi_! D = \frac{E\left(\{m_1 + g'(z)[1 + m_2 - m_1(z + \kappa)]\}^3\right)}{(1 + m_2)^3},$$

where

$$m_r = E\left((z + \kappa)^r g''(z)\right) \quad \text{for } r = 0, 1, 2$$

and where E indicates mean value over z with z having probability density function f.

6. The relation between marginal likelihood and modified profile likelihood for composite transformation models. In this section we consider composite transformation models with properties as described in Section 5.4. We shall assume that the dominating measure μ is invariant, so that $\chi(g, x) = 1$ in (5.9). Then the marginal likelihood for the index parameter κ based on the maximal invariant statistic u may be expressed as

$$\check{L}(\kappa) = \int L(\kappa, h; x)\Delta(h)^{-1}\, d\nu(h), \tag{6.1}$$

where ν is an invariant measure on $H \equiv \Psi \equiv G/K$ and Δ is the modular function of G [cf. Barndorff-Nielsen (1983), Theorem 5.1].

For later use we wish to assess the relation between this marginal likelihood on the one hand and the profile likelihood and modified profile likelihood for κ on the other, when the original likelihood $L(\kappa, h) = L(\kappa, h; x)$ is based on a sample $y_1, \dots, y_n$ of size n from a composite transformation model with model function $p_0(y; \kappa, h)$. In this case $x = (y_1, \dots, y_n)$ and

$$p(x; \kappa, h) = \prod_{i=1}^{n} p_0(y_i; \kappa, h).$$

Any version of the likelihood function for (κ, h) based on x may be inserted on the right-hand side of (6.1). To determine the behaviour of the integral (6.1) it is convenient to take

$$L(\kappa, h) = \frac{p(x; \kappa, h)}{p(x; \kappa_0, h_0)},$$

where, in the present context, (κ_0, h_0) denotes the actual (or "true") value of the parameter (κ, h). With $l = \ln L$, formula (6.1) may then be rewritten as

$$\check{L}(\kappa) = \int e^{-nI_n(\kappa_0, h_0; \kappa, h)} \Delta(h)^{-1} \, d\nu(h), \tag{6.2}$$

where

$$I_n(\kappa_0, h_0; \kappa, h) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{p_0(y_i; \kappa, h)}{p_0(y_i; \kappa_0, h_0)}.$$

For $n \to \infty$ we have, with probability 1,

$$I_n(\kappa_0, h_0; \kappa, h) \to I(\kappa_0, h_0, \kappa, h), \tag{6.3}$$

where $I(\kappa_0, h_0; \kappa, h)$ is the discrimination (or Kullback–Leibler) information

$$I(\kappa_0, h_0; \kappa, h) = -\int \log \frac{p_0(y; \kappa, h)}{p_0(y; \kappa_0, h_0)} p_0(y; \kappa_0, h_0) \, d\mu_0(y),$$

μ_0 denoting the dominating measure for the probability functions $p_0(\cdot; \kappa, h)$. In view of the convergence (6.3) we may expect Laplace's method to be applicable for the asymptotic evaluation of (6.2). Before this method can be applied, it is, however, necessary to reexpress (6.2) as an integral with respect to Lebesgue measure λ on H, rather than the invariant measure ν. Denote the maximum likelihood estimator of h when κ is fixed by $\hat{h}_\kappa$ and let $u_\kappa = \hat{h}_\kappa^{-1} x$ be the corresponding maximal invariant. Then, after rescaling ν by multiplication by a constant, we have, by formulas (4.12) and (4.13) of Barndorff-Nielsen (1983),

$$d\nu(h) = |j_{hh}(\kappa, h; hu_\kappa)|^{1/2} |j_{hh}(\kappa, e; u_\kappa)|^{-1/2} \, d\lambda(h)$$

so that

$$\check{L}(\kappa) = \int L(\kappa, h; x) \Delta(h)^{-1} |j_{hh}(\kappa, h; hu_\kappa)|^{1/2} |j_{hh}(\kappa, e; u_\kappa)|^{-1/2} \, d\lambda(h).$$

By formal application of the multivariate version of Laplace's formula [cf. Glynn (1980)], we now obtain

$$\check{L}(\kappa) \sim (2\pi)^{d/2} n^{-d/2} \Delta(\hat{h}_\kappa)^{-1} |j_{hh}(\kappa, e; u_\kappa)|^{-1/2} \tilde{L}(\kappa), \tag{6.4}$$

and, furthermore, we may expect the relative error of this asymptotic formula to be $O(n^{-1})$; cf. Barndorff-Nielsen and Cox (1979) and Barndorff-Nielsen (1983). Let

$$R(\kappa) = \Delta(\hat{h}_\kappa)^{-1} |j_{hh}(\kappa, e; u_\kappa)|^{-1/2} \tag{6.5}$$

and $r(\kappa) = \ln R(\kappa)$. Then, taking logarithms in (6.4) and dropping an additive term depending on n only, we find that, in considerable generality,

$$\check{l}(\kappa) = \tilde{l}(\kappa) + r(\kappa) + O_p(n^{-1}) \tag{6.6}$$

with $\tilde{l}(\kappa)$ of order $O_p(n)$ and $r(\kappa)$ of order $O_p(1)$.

As we will show, the quantity $R(\kappa)$ is—except for a multiplicative term depending on the observations only—equal up to a factor of $1 + O_p(n^{-1/2})$ to the adjustment factor which, when applied to $\tilde{L}(\kappa)$, yields the modified profile likelihood $\overset{\dagger}{\tilde{L}}(\kappa)$ as defined, for general models, by (2.3). That is, we have

$$\overset{\dagger}{\tilde{l}}(\kappa) = \tilde{l}(\kappa) + r(\kappa) + O_p(n^{-1/2}), \tag{6.7}$$

and hence, by (6.6),

$$\check{l}(\kappa) = \overset{\dagger}{\tilde{l}}(\kappa) + O_p(n^{-1/2}). \tag{6.8}$$

Formulated otherwise, (6.7) states that, in the present case of repeated sampling from a composite transformation model, the modified profile likelihood for the index parameter κ is equal up to a factor of $1 + O_p(n^{-1/2})$ to the approximation to the marginal likelihood for κ obtained by applying Laplace's method to the integral representation (6.1).

In some important classes of cases, stronger versions of (6.7) and (6.8) hold. If $\hat{h}_\kappa = \hat{h}$ or if G acts freely on Ω, then we have

$$\overset{\dagger}{\tilde{l}}(\kappa) = \tilde{l}(\kappa) + r(\kappa) \tag{6.9}$$

and so

$$\check{l}(\kappa) = \overset{\dagger}{\tilde{l}}(\kappa) + O_p(n^{-1}). \tag{6.10}$$

(It may be added that G acting freely on Ω is equivalent to K consisting only of the identity element.) Equation (6.10) holds also if κ and h are orthogonal (i.e., $i_{\kappa h} = 0$). This too will be shown.

For any action γ of the group G on a space $\mathscr{Y}$ we let $J_{\gamma(g)}(y)$ denote the Jacobian of the transformation $\gamma(g)$ of $\mathscr{Y}$, evaluated at $y \in \mathscr{Y}$. To derive (6.9) and (6.7), let δ and ε denote, respectively, the action of G on itself from the left and from the right, i.e., $\delta(g)\acute{g} = g\acute{g}$ and $\varepsilon(g)\acute{g} = \acute{g}g^{-1}$. Comparing (2.3) with (6.5) and using formula (4.13) of Barndorff-Nielsen (1983), bearing in mind that $\psi \equiv h$, we find that (6.9) is equivalent [up to addition to one side of (6.9) of a function of the observations alone] to

$$\begin{aligned} &\left|j_{hh}(\kappa, \hat{h}_\kappa)\right| \\ &\quad = \left|\frac{\partial^2 l}{\partial h\, \partial \hat{h}}(\kappa, \hat{h}_\kappa; \hat{\kappa}, \hat{h}, a)\right| \Delta(\hat{h}_\kappa)^{-1} J_{\acute{\gamma}(\hat{h}_\kappa)}(e)^{-1} \Delta(\hat{h}) J_{\delta(\hat{h})}(e), \end{aligned} \tag{6.11}$$

where the action $\acute{\gamma}$ has been defined in Section 5.4. Similarly, (6.7) is equivalent to (6.11) holding except for an error of order $O_p(n^{-1/2})$.

Now, the log-likelihood function may be written as

$$l(\kappa, h; \hat{\kappa}, \hat{h}, a) = l(\kappa, e; \hat{\kappa}, \eta(h^{-1}\hat{h}), a),$$

where η is defined in Section 5.4. On differentiating this we find

$$\begin{aligned} &j_{hh}(\kappa, \hat{h}_\kappa) \\ &\quad = -\left\{\frac{\partial \eta(h^{-1}\hat{h})}{\partial h^T}\right\}_{h \mapsto \hat{h}_\kappa} \frac{\partial^2 l}{\partial \hat{h}\, \partial \hat{h}}(\kappa, e; \hat{\kappa}, \eta(\hat{h}_\kappa^{-1}\hat{h}), a) \left\{\frac{\partial \eta(h^{-1}\hat{h})^T}{\partial h}\right\}_{h \mapsto \hat{h}_\kappa} \end{aligned}$$

and

$$\left\{\frac{\partial^2 l}{\partial h^T \partial \hat{h}}\right\}_{h \mapsto \hat{h}_\kappa}$$

$$= \left\{\frac{\partial \eta(h^{-1}\hat{h})}{\partial h^T}\right\}_{h \mapsto \hat{h}_\kappa} \frac{\partial^2 l}{\partial \hat{h}\, \partial \hat{h}}\left(\kappa, e; \hat{\kappa}, \eta\left(\hat{h}_\kappa^{-1}\hat{h}\right), a\right)\left\{\frac{\partial \eta(h^{-1}\hat{h})^T}{\partial \hat{h}}\right\}_{h \mapsto \hat{h}_\kappa}.$$

Thus (6.11) is equivalent to

$$\left\{\left|\frac{\partial \eta(h^{-1}\hat{h})^T}{\partial h}\right|\right\}_{h \mapsto \hat{h}_\kappa} = \left|\frac{\partial \eta\left(\hat{h}_\kappa^{-1}\hat{h}\right)^T}{\partial \hat{h}}\right| \Delta\left(\hat{h}_\kappa\right)^{-1} J_{\dot{\gamma}(\hat{h}_\kappa)}(e)^{-1} \Delta(\hat{h}) J_{\delta(\hat{h})}(e). \tag{6.12}$$

To prove the validity of (6.12), either exactly (for the cases indicated) or up to an error of $O_p(n^{-1/2})$, we need the following two lemmas.

LEMMA 6.1. *Let $G = HK$ be a left factorization of the group G. For any fixed element h' of H we have*

$$\left|\frac{\partial \eta(h^{-1}h')}{\partial h}\right|_{h \mapsto h'} = \left|\frac{\partial \eta(h'^{-1}h)}{\partial h}\right|_{h \mapsto h'}. \tag{6.13}$$

PROOF. Define α: $H \to G$, β: $H \to H$ and ι: $G \to G$ by $\alpha(h) = h'^{-1}h$, $\beta(h) = \eta(h^{-1}h')$ and $\iota(g) = g^{-1}$. Then the following diagram of mappings commutes:

$$\begin{tikzcd} G \arrow[r, "\iota"] & G \arrow[d, "\eta"] \\ H \arrow[u, "\alpha"] \arrow[r, "\beta"] & H \end{tikzcd}$$

The "chain rule" of differentiation yields a corresponding commutative diagram of differential mappings between tangent spaces:

$$\begin{tikzcd} TG_{h'^{-1}h} \arrow[r, "\iota_*"] & TG_{h^{-1}h'} \arrow[d, "\eta_*"] \\ TH_h \arrow[u, "\alpha_*"] \arrow[r, "\beta_*"] & TH_{\eta(h^{-1}h^*)} \end{tikzcd}$$

If $h = h'$, then ι_*: $TG_{h'^{-1}h} \to TG_{h^{-1}h'}$ is $-I$, where I denotes the identity mapping. Formula (6.13) can now be read off from the diagram of derivatives. □

LEMMA 6.2. *Let $G = HK$ be a left factorization of G. Then $J_{\dot{\gamma}(h)}(e) = J_{\delta(h)}(e)$ for any $h \in H$.*

PROOF. First note that if h and h' are any two elements of H, then

$$\eta(h'h) = \dot{\gamma}(h')h. \tag{6.14}$$

Next let $g = hk$ denote an arbitrary element of G. Writing g symbolically as (h, k) and employing the mappings η and ζ defined by

$$\eta: g \mapsto h, \qquad \zeta: g \mapsto k$$

(cf. Section 5.4), we have, for any $h' \in H$,

$$\delta(h')g = \delta(h')(h, k) = (\eta(h'h), \zeta(h'hk))$$

and hence the differential of $\delta(h')$ at g is

$$D\delta(h')(g) = \begin{bmatrix} \dfrac{\partial \eta(h'h)^T}{\partial h} & 0 \\ \dfrac{\partial \zeta(h'hk)^T}{\partial h} & \dfrac{\partial \zeta(h'hk)^T}{\partial k} \end{bmatrix},$$

from which we find, using (6.14) and $\zeta(h'k) = k$,

$$\begin{aligned} J_{\delta(h')}(e) &= J_{\dot{\gamma}(h')}(e) \left| \frac{\partial \zeta(h'k)^T}{\partial k} \right|_{k \mapsto e} \\ &= J_{\dot{\gamma}(h')}(e). \end{aligned}$$

□

Formula (6.13) and Lemma 6.2 together show that (6.12) holds exactly provided $\hat{h} = \hat{h}_\kappa$. When $\hat{h}$ and $\hat{h}_\kappa$ are not identical, (6.12) is still correct to order $O_p(n^{-1/2})$ as may be seen by expanding $\hat{h}_\kappa$ in κ around $\hat{\kappa}$, the first term in this expansion being $\hat{h}$. In case the parameters h and κ are orthogonal, i.e., $i_{h\kappa} = 0$, we have that $\hat{h}_\kappa - \hat{h}$ is of order $O_p(n^{-1})$ rather than $O_p(n^{-1/2})$ and consequently (6.12) is then valid to order $O_p(n^{-1})$.

Now suppose that G acts freely on Ω. Then $H = G$ and so we have $\eta(h^{-1}\hat{h}) = h^{-1}\hat{h}$ and

$$(6.15) \qquad \begin{aligned} \frac{\partial \eta(h^{-1}\hat{h})}{\partial \hat{h}} &= J_{\delta(h^{-1})}(\hat{h}) \\ &= J_{\delta(h^{-1}\hat{h})}(e)/J_{\delta(\hat{h})}(e). \end{aligned}$$

We next need the facts that

$$(6.16) \qquad \left| \frac{\partial h^{-1}}{\partial h} \right| = J_{\varepsilon(h)}(e)/J_{\delta(h)}(e)$$

and

$$(6.17) \qquad \Delta(h) = J_{\varepsilon(h^{-1})}(e)/J_{\delta(h)}(e).$$

[See, for instance, Fraser (1979), page 148.] Using the chain rule of differentiation together with formulas (6.15), (6.16) and (6.17) we obtain

$$\begin{aligned} \left| \frac{\partial \eta(h^{-1}\hat{h})}{\partial h} \right| &= \left| \frac{\partial \eta(\hat{h}^{-1}h)}{\partial h} \right| \left| \frac{\partial \iota_H(g)}{\partial g} \right|_{g \mapsto \hat{h}^{-1}h} \\ &= \Delta(h^{-1}\hat{h}) J_{\delta(h^{-1}\hat{h})}(e)/J_{\delta(h)}(e). \end{aligned}$$

Formula (6.12) now follows from Lemma 6.2 in conjunction with (6.15) and the fact that the modular function is a multiplier, i.e., $\Delta(hh') = \Delta(h)\Delta(h')$.

We have now established (6.7) and (6.8) and we have shown the validity of (6.10) subject to either h and κ being orthogonal (which is the case, in particular, if $\hat{h}_\kappa = \hat{h}$) or G acting freely on Ω.

Using (6.6) we can relate the marginal maximum likelihood estimate $\hat{\hat{\kappa}}$ of κ to the maximum likelihood estimate $\hat{\kappa}$ from the full likelihood. In considerable generality, we have

$$\hat{\hat{\kappa}} = \hat{\kappa} + a_1 n^{-1} + O_p(n^{-2})$$

for some coefficient $a_1 = a_1(\kappa_0, h_0)$. Differentiating (6.6) with respect to κ and inserting $\hat{\hat{\kappa}}$ for κ, we obtain

$$\begin{aligned} 0 &= (D\tilde{l})(\hat{\hat{\kappa}}) + (Dr)(\hat{\hat{\kappa}}) + O_p(n^{-1}) \\ &= -\hat{\tilde{\not{j}}} a_1 n^{-1} + (Dr)(\hat{\kappa}) + O_p(n^{-1/2}) \\ &= -\hat{\tilde{\not{j}}}_0 a_1 + (Dr)(\hat{\kappa}) + O_p(n^{-1/2}), \end{aligned}$$

where $\hat{\tilde{\not{j}}}_0 = n^{-1}\hat{\tilde{\not{j}}} = n^{-1}(D^2\tilde{l})(\hat{\kappa})$ is the observed profile information on κ per observation. Thus, by Theorem 4.2, we may take

$$a_1(\kappa, h) = \{\pi_! i_0(\kappa)\}^{-1} Dr(\kappa)$$

with i_0 denoting the expected information per observation.

7. Geometries of distributional-shape manifolds. In composite transformation models it is often appropriate to think of the index parameter κ as determining distributional shape while the group parameter ψ may be considered as a generalized location parameter. We are interested here in various statistical geometries on the manifold $\mathscr{K}$, the range space of the parameter κ, and we shall refer to $\mathscr{K}$ as the manifold of distributional shapes.

As indicated in Table 1, there are several ways in which the distributional-shape manifold can be set up as a statistical manifold, i.e., can be equipped with a kit. [Recall from Section 2.2 that a kit is a pair (ϕ, S) consisting of a metric tensor ϕ and a skewness tensor S.] The kits in Table 1 are defined in Section 4.

That $\pi_!(i, D) = (\pi_! i, \pi_! D)$ is well defined is a consequence of Theorem 4.3 and of the fact that the property of being a composite transformation model is preserved under repeated sampling.

Sufficient conditions for $\pi_!(\not{j}, \not{T}) = (\pi_! \not{j}, \pi_! \not{T})$ to be well defined are contained in Theorem 7.1, which also relates $\pi_!(\not{j}, \not{T})$ to the profile object $(\tilde{\not{j}}, \tilde{\not{T}})$.

The profile kits $(\tilde{\not{j}}, \tilde{\not{T}})$ and $(\tilde{i}, \tilde{D})$ are defined for general parametric models, whether of the composite transformation type or not, provided there is an L-sufficient statistic for the parameter κ; cf. Section 4.

The definitions of the modified profile kits $(\overset{+}{\not{j}}, \overset{+}{\not{T}})$ and $(\overset{+}{i}, \overset{+}{D})$ are analogous to those of the profile kits $(\tilde{\not{j}}, \tilde{\not{T}})$ and $(\tilde{i}, \tilde{D})$. The proof that $(\overset{+}{\not{j}}, \overset{+}{\not{T}})$ is a kit is similar to that for $(\not{j}, \not{T})$ and uses the fact that $\overset{+}{\not{j}}$ and $\overset{+}{\not{T}}$ are obtained by evaluating higher mixed derivatives of $\overset{+}{l}$ at points where $\overset{+}{l}_i = 0$. That $(\overset{+}{i}, \overset{+}{D})$ is well defined for models with cuts, exponential models with τ-parallel foliations and composite transformation models follows from Barndorff-Nielsen (1985).

The marginal kits $(\check{\not{j}}, \check{\not{T}})$ and $(\check{i}, \check{D})$ are simply the observed and expected kits, defined like $(\not{j}, \not{T})$ and (i, D) but from the marginal model for u.

We emphasize again that throughout the following the statistical model $(\mathscr{X}, p, \Omega)$ is assumed to be a composite transformation model as described in Section 5.4.

THEOREM 7.1. *Let a be a G-invariant statistic such that $(\hat{\omega}, a)$: $\mathscr{X} \to \Omega \times A$ is bijective and sufficient for ω and let b be the statistic defined by $a(x) = b(u(x))$, where u is the orbit representative. With $(\not{j}, \not{T})$ and $(\tilde{\not{j}}, \tilde{\not{T}})$ defined on the basis of the ancillaries a and b, respectively, we have that $\not{j}$ and $\not{T}$ transfer to $\mathscr{K}$, by $\pi_!$ and that*

$$\pi_!(\not{j}, \not{T}) = (\tilde{\not{j}}, \tilde{\not{T}}). \tag{7.1}$$

PROOF. Due to the L-sufficiency of u (cf. Section 5.4), the theorem is a corollary of Theorem 4.1. □

We have not found any general conditions which would ensure that, in analogy with (7.1), $\pi_!(i, D) = (\tilde{i}, \tilde{D})$. However, a related result may be established on the basis of the concept of profile discrimination information, as defined in (4.6).

The following theorem is a "profile" version of the result that the discrimination information I determines both the expected information i and the expected skewness tensor D on Ω [by (4.4) and (4.5)].

THEOREM 7.2. *For a composite transformation model with a quasiinvariant dominating measure we have*

$$\pi_! i(\kappa) = \left. \frac{\partial^2 \tilde{I}(\kappa, \kappa')}{\partial \kappa' \, \partial \kappa'} \right|_{\kappa' \mapsto \kappa} \tag{7.2}$$

and

$$\pi_! D(\kappa) = \left[\frac{\partial^3 \tilde{I}(\kappa, \kappa')}{\partial \kappa \, \partial \kappa' \, \partial \kappa'} - \frac{\partial^3 \tilde{I}(\kappa, \kappa')}{\partial \kappa' \, \partial \kappa \, \partial \kappa} \right]_{\kappa' \mapsto \kappa}. \tag{7.3}$$

PROOF. Choose ω in $\pi^{-1}(\kappa)$. For κ' near κ define $\hat{\psi}_{\kappa'}$ by

$$I\big(\omega, (\kappa', \hat{\psi}_{\kappa'})\big) = \min_{\psi} I(\omega, (\kappa', \psi)).$$

It follows from (5.10) that $I(g\omega, g\omega') = I(\omega, \omega')$ and so

$$\tilde{I}(\kappa, \kappa') = I\big(\omega, (\kappa', \hat{\psi}_{\kappa'})\big).$$

A simple analogue of the proof of Theorem 4.2, using the distribution corresponding to ω instead of the empirical distribution, gives the required results. □

More generally, in suitable cases, kits obtained from contrast functions (as indicated at the end of Section 4) transfer to $\mathscr{K}$, as in (7.2) and (7.3).

From the inferential point of view one should ideally work with the geometries $(\check{j}, \check{T})$ and $(\check{i}, \check{D})$ when κ is the parameter of interest. Often, however, these quantities are analytically and numerically intractable and approximations are needed. Under wide conditions the corresponding profile and modified profile geometries, as set out in Table 1, afford such approximations, with modified profile likelihood yielding accuracy of higher order than profile likelihood; cf. Section 6. We shall not attempt a detailed comparison here, but we wish to make a couple of observations.

Suppose, as in Section 6, that the composite transformation model considered is for a sample $y_1, \dots, y_n$ of size n from an underlying composite transformation model which is indicated by lower index 0. From Theorem 4.2, we find that as $n \to \infty$, with probability 1,

$$n^{-1}\tilde{j} \to \pi_! i_0$$

or, otherwise expressed,

$$n^{-1}\left(\tilde{j} - \pi_! i\right) \to 0.$$

Furthermore, by differentiating (6.6) twice and three times we obtain

$$\check{j} = \tilde{j} + O_p(1)$$

and

$$\check{T} = \tilde{T} + O_p(1).$$

Consequently, we find that, with probability 1,

$$n^{-1}\left(\check{i}(\kappa) - \pi_! i(\kappa)\right) \to 0$$

and

$$n^{-1}\left(\check{D}(\kappa) - \pi_! D(\kappa)\right) \to 0.$$

8. Examples concerning composite transformation models. We now illustrate the results of Section 7 by comparing some of the diverse kits for a variety of examples.

EXAMPLE 8.1. *Normal model under scale.* The univariate normal model with probability density function

$$p(x; \mu, \sigma) = (2\pi)^{-1/2}\sigma^{-1}\exp\left\{-(x-\mu)^2/2\sigma^2\right\}$$

and with parameter space $\{(\mu, \sigma) \in \mathbb{R}^2 \colon \sigma > 0\}$ is a composite transformation model under the multiplicative (scale) action of $G = \mathbb{R}^+$ on $\mathscr{X} = \mathbb{R}$. The coefficient of variation $\kappa = \sigma^{-1}\mu$ is an invariant parameter and the incidental parameter can be taken as $\psi = \sigma^{-1}$. Then by direct calculation or from the result for general location-scale models given in Section 5.4 we have

$$i_{\kappa\kappa} = 1, \qquad i_{\kappa\psi} = -\psi^{-1}\kappa, \qquad i_{\psi\psi} = \psi^{-2}(2 + \kappa^2),$$

$$D_{\kappa\kappa\kappa} = 0, \qquad D_{\kappa\kappa\psi} = -2\psi^{-1}, \qquad D_{\kappa\psi\psi} = 4\psi^{-2}\kappa, \qquad D_{\psi\psi\psi} = -2\psi^{-3}(4 + 3\kappa^2)$$

and so

$$\pi_! i = 2(2 + \kappa^2)^{-1},$$

$$\pi_! D = -8\kappa(3 + \kappa^2)(2 + \kappa^2)^{-3}.$$

EXAMPLE 8.2. *Inverse Gaussian model under scale.* The inverse Gaussian model $N^-(\chi, \psi)$ has sample space $(0, \infty)$ and model function

$$p(x; \chi, \psi) = \frac{\sqrt{\chi}}{\sqrt{2\pi}} \exp(\sqrt{\chi\psi}) x^{-3/2} \exp\left\{-\frac{1}{2}(\chi x^{-1} + \psi x)\right\}, \tag{8.1}$$

where $\chi > 0$ and $\psi \geq 0$. It is a composite transformation model under the multiplicative action of $\mathbb{R}^+$ on $(0, \infty)$. On the submodel with $\psi > 0$ we can use the invariant parameter $\kappa = \sqrt{\chi\psi}$ and the incidental parameter $\mu = \sqrt{\chi\psi^{-1}}$. Calculation yields

$$i_{\kappa\kappa} = \tfrac{1}{2}\kappa^{-2}, \qquad i_{\kappa\mu} = \tfrac{1}{2}\mu^{-1}\kappa^{-1}, \qquad i_{\mu\mu} = \tfrac{1}{2}\mu^{-2}(1 + 2\kappa)$$

so that

$$\pi_! i = \kappa^{-1}(1 + 2\kappa)^{-1}.$$

We find also that

$$D_{\kappa\kappa\kappa} = -\kappa^{-3}, \qquad D_{\kappa\kappa\mu} = -\mu^{-1}\kappa^{-2},$$

$$D_{\kappa\mu\mu} = -\mu^{-2}\kappa^{-1}(1 + \kappa), \quad D_{\mu\mu\mu} = -\mu^{-3}$$

and so

$$\pi_! D = -\kappa^{-3}(1 + 2\kappa)^{-3}(8\kappa^3 + 6\kappa^2 + 3\kappa).$$

EXAMPLE 8.3. *von Mises–Fisher model.* Consider the $(d-1)$-dimensional von Mises–Fisher distribution

$$p(x; \kappa, \psi) = a_d(\kappa)\exp(\kappa\psi \cdot x), \tag{8.2}$$

where x and ψ are unit vectors in $\mathbb{R}^d$, $\kappa \geq 0$ and

$$a_d(\kappa) = \kappa^{d/2-1} / \left\{(2\pi)^{d/2} I_{d/2-1}(\kappa)\right\}.$$

The sample space is S^{d-1}, the unit sphere in $\mathbb{R}^d$. The action of the rotation group $SO(d)$ on S^{d-1} gives this model the structure of a composite transformation model. After removing the origin [which is the fixed-point set of the action of $SO(d)$ on the parameter space] from $\mathbb{R}^d$ we have a submersion $(\kappa, \psi) \to \kappa$ of $\Omega = \mathbb{R}^d \setminus \{0\}$ onto $\mathscr{K} = \mathbb{R}^+$. Simple calculations show that the parameters κ and ψ are orthogonal and

$$\pi_! i_0 = i_{0\kappa\kappa} = -\frac{d^2}{d\kappa^2}\log a_d(\kappa),$$

$$\pi_! D_0 = D_{0\kappa\kappa\kappa} = -\frac{d^3}{d\kappa^3}\log a_d(\kappa).$$

Let r denote the resultant length of a random sample of size n from (8.2). Then

$$\tilde{l}(\kappa) = n \log a_d(\kappa) + \kappa r$$

and calculation shows that

$$(\tilde{i}, \tilde{D}) = \pi_!(i, D) + O(1).$$

Since the logarithm of the modified profile likelihood is

$$\tilde{\tilde{l}}(\kappa) = \tilde{l}(\kappa) + \frac{(1-d)}{2}\log(\kappa r),$$

we have

$$(\tilde{\tilde{i}}, \tilde{\tilde{D}}) = (\tilde{i}, \tilde{D}).$$

The marginal log-likelihood of κ based on r is

$$\check{l}(\kappa; r) = n \log a_d(\kappa) - \log a_d(r\kappa)$$

so that

$$n^{-1}(\check{i} - \pi_! i) = -n^{-1}E(r^2 A'(\kappa r)),$$

$$n^{-1}(\check{D} - \pi_! D) = n^{-1}\{-E(r^3 A^{(2)}(\kappa r)) + 3nA(\kappa)E(r^2 A'(\kappa r))$$
$$+3nA'(\kappa)E(r\{A(\kappa r) - 1\}) + 3E(r^3\{A(\kappa r) - 1\}A'(\kappa r))\},$$

where $A(\kappa) = -(d/d\kappa)\log a_d(\kappa)$. It follows from standard theory of exponential families that $A'(\kappa) > 0$, and so $\check{i} < \pi_! i$, in agreement with (4.11).

Example 8.4. *Hyperboloid model.* The hyperboloid model of exponential order 3 may be defined as the joint distribution of two random variables u and v with probability density function

$$p(u, v; \chi, \varphi, \lambda) = (2\pi)^{-1}\lambda e^{\lambda} \sinh u \exp(-\lambda\{\cosh \chi \cosh u - \sinh \chi \sinh u \cos(v - \varphi)\}), \tag{8.3}$$

where $0 \le u$, $0 \le v < 2\pi$ and where the parameter $\omega = (\chi, \varphi, \lambda)$ satisfies $0 \le \chi < \infty$, $0 \le \varphi < 2\pi$ and $0 < \lambda < \infty$. This can be viewed [Barndorff-Nielsen (1978b) and Jensen (1981)] as a model for observations on the positive unit hyperboloid in $\mathbb{R}^3$ with properties similar to those of the von Mises–Fisher distribution on the unit sphere. For a sample $(u_1, v_1), \dots, (u_n, v_n)$ from (8.3) the maximum likelihood estimator $(\hat{\chi}, \hat{\varphi}, \hat{\lambda})$ is minimal sufficient and $\hat{\lambda} = ((r/n) - 1)^{-1}$, where

$$r = \left\{\left(\sum \cosh u_i\right)^2 - \left(\sum \sinh u_i \cos v_i\right)^2 - \left(\sum \sinh u_i \sin v_i\right)^2\right\}^{1/2}$$

is the analogue of the resultant length of von Mises–Fisher observations. The statistic $r' = r - n$ follows the gamma distribution

$$\frac{\lambda^{n-1}}{\Gamma(n-1)} r'^{n-2} e^{-\lambda r'}$$

and the joint distribution of $\hat{\chi}$ and $\hat{\varphi}$ conditional on r (or, equivalently, conditional on $\hat{\lambda}$) is that of the hyperboloid model (8.3) but with λ replaced by $r\lambda$. For fixed λ, (8.3) is a transformation model, the group action being that of the special pseudoorthogonal group $SO^+(2,1)$. [For proofs and details, see Jensen (1981).]

The model for $(r, \hat{\chi}, \hat{\varphi})$ can be considered as a composite transformation model with index parameter $\kappa = (\chi, \lambda)$, relative to the obvious action of $SO(2)$. The profile, modified profile and marginal log-likelihoods of κ are

$$\tilde{l}(\kappa) = n\{\ln\lambda + \lambda\} - r\lambda\cosh(\hat{\chi} - \chi), \tag{8.4}$$

$$\tilde{\tilde{l}}(\kappa) = \tilde{l}(\kappa) + \tfrac{1}{2}\ln(r\lambda\sinh\hat{\chi}\sinh\chi), \tag{8.5}$$

$$\begin{aligned} \check{l}(\kappa) &= l(\kappa; \hat{\chi}, r) \\ &= \tilde{l}(\kappa) - r\lambda\sinh\hat{\chi}\sinh\chi + \log I_0(r\lambda\sinh\hat{\chi}\sinh\chi) \qquad (8.6)\\ &= \tilde{\tilde{l}}(\kappa) + \log\{I_0/I_0^\dagger\}. \end{aligned}$$

In the latter expression and the text to follow, I_ν indicates one of the Bessel functions, in the standard notation. Furthermore, $I_\nu^\dagger$ is the leading term of the asymptotic expansion of $I_\nu(x)$ for $x \to \infty$.

Expected and observed geometries on Ω are identical since (8.3) is a full exponential model. From

$$l(\chi, \varphi, \lambda) = \tilde{l}(\chi, \lambda) + r\lambda\sinh\hat{\chi}\sinh\chi\{\cos(\hat{\varphi} - \varphi) - 1\}$$

we find $j_{\kappa\varphi} = 0$ and, using (7.1) and $r = n(\hat{\lambda}^{-1} + 1)$ and writing $\lambda^\dagger$ for $\lambda^{-1} + 1$,

$$(\pi_! j)_{\chi\chi} = n\lambda\lambda^\dagger, \qquad (\pi_! j)_{\chi\lambda} = 0, \qquad (\pi_! j)_{\lambda\lambda} = n\lambda^{-2}$$

and

$$\begin{aligned} (\pi_! T)_{\chi\chi\chi} &= 3n\lambda\lambda^\dagger, & (\pi_! T)_{\chi\chi\lambda} &= 0, \\ (\pi_! T)_{\chi\lambda\lambda} &= 0, & (\pi_! T)_{\lambda\lambda\lambda} &= -n\lambda^{-3}. \end{aligned}$$

On the other hand, as the modified profile likelihood $\tilde{\tilde{l}}(\kappa)$ depends on $(\hat{\chi}, \hat{\varphi}, \hat{\lambda})$ through $(\hat{\chi}, \hat{\lambda})$ only and since $(\hat{\chi}, \hat{\lambda})$ is minimally L-sufficient for κ, modified profile geometries on $\mathscr{K}$ are defined from (8.5) by

$$\tilde{\tilde{j}}_{\chi\chi} = n\lambda\lambda^\dagger + \tfrac{1}{2}\sinh^{-2}\chi, \qquad \tilde{\tilde{j}}_{\chi\lambda} = 0, \qquad \tilde{\tilde{j}}_{\lambda\lambda} = (n + \tfrac{1}{2})\lambda^{-2}$$

and

$$\begin{aligned} \tilde{\tilde{T}}_{\chi\chi\chi} &= 3n\lambda\lambda^\dagger + \cosh\chi\sinh^{-3}\chi, & \tilde{\tilde{T}}_{\chi\chi\lambda} &= 0, \\ \tilde{\tilde{T}}_{\chi\lambda\lambda} &= 0, & \tilde{\tilde{T}}_{\lambda\lambda\lambda} &= -(n-1)\lambda^{-3}. \end{aligned}$$

These latter geometries approximate the marginal geometries induced by $(\hat{\chi}, r)$ to order $O(n^{-1})$ since the asymptotic expansion of $I_0(x)/I_0^\dagger(x)$ descends in powers of x. This also follows from (6.10).

We may further use the hyperboloid model to illustrate the induction of geometries from profile likelihood or modified profile likelihood in the presence of an ancillary statistic. For this purpose, suppose λ is known and take χ as the index parameter. Then

$$\tilde{l}(\chi) = -r\lambda \cosh(\hat{\chi} - \chi),$$

which depends on the observations through $(\hat{\chi}, r)$ only. Now, r is exactly ancillary and treating $\tilde{l}(\chi)$ as if it were a genuine log-likelihood function with associated ancillary r we find $\tilde{l}_{\chi\chi\chi} = \tilde{l}_{\chi\chi;\chi} = 0$, so that the induced skewness tensor $\tilde{T}$ is 0. It is easily checked that $\pi_! T = \tilde{T}$.

Note that in the present case the observed and expected skewness tensors as determined from the conditional model for $\hat{\chi}$ given r are complicated expressions involving the Bessel function I_0 and its first three derivatives.

EXAMPLE 8.5. *Hyperbolic distribution.* The hyperbolic distribution

$$(8.7)\quad \begin{aligned} p(x; \zeta, \pi, \mu, \delta) &= \left\{2\delta\sqrt{1+\pi^2}K_1(\zeta)\right\}^{-1} \\ &\quad \times \exp\left\{-\zeta\left[\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2} - \pi\left(\frac{x-\mu}{\delta}\right)\right]\right\} \end{aligned}$$

has been found to fit empirical distributions from a wide variety of contexts. Examples are given in Barndorff-Nielsen and Blæsild (1981, 1983a) and Barndorff-Nielsen, Blæsild, Jensen and Sørensen (1985). It is of some practical interest to be able to compare the "shapes" of the various hyperbolic distributions which occur, irrespective of the origin and scale used for each set of observations. In other words, it is useful to study the shape manifold of the composite transformation model (8.7).

This hyperbolic shape manifold can be parameterized by the invariant parameters (ζ, π) or by the parameters (ξ, χ) defined by $\xi = (1 + \zeta)^{-1/2}$ and $\chi = \xi\pi(1 + \pi^2)^{-1/2}$. The domain of variation of (ξ, χ) is an isosceles triangle and this shape triangle has proved useful in the comparison of hyperbolic shapes; see Barndorff-Nielsen, Blæsild, Jensen and Sørensen (1985) and Barndorff-Nielsen and Christiansen (1985, 1988). For more detailed comparisons of shapes, it would be useful to have some knowledge of the various geometries on the shape triangle. Expressions in closed form for $\pi_!(i, D)$, $(\tilde{i}, \tilde{D})$ and (i, D) seem, however, not to be available and considerable numerical calculations are required to obtain such knowledge.

A number of further examples are discussed in a report available from the authors.

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. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statist.* **28.** Springer, Berlin.

AMARI, S.-I. (1987). Differential geometrical theory of statistics. In *Differential Geometry in Statistical Inference* 19–94. IMS, Hayward, Calif.

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. (1978a). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

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

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

BARNDORFF-NIELSEN, O. E. (1985). Properties of modified profile likelihood. In *Contributions to Probability and Statistics in Honour of Gunnar Blom* (J. Lanke and G. Lindgren, eds.) 25–38. Dept. Mathematical Statistics, Lund Univ.

BARNDORFF-NIELSEN, O. E. (1986). Likelihood and observed geometries. *Ann. Statist.* **14** 856–873.

BARNDORFF-NIELSEN, O. E. (1987). Differential and integral geometry in statistical inference. In *Differential Geometry in Statistical Inference* 95–161. IMS, Hayward, Calif.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and applications. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. A. Baldessari, eds.) 4 19–44. Reidel, Dordrecht.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1983a). Hyperbolic distributions. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **3** 700–707. Wiley, New York.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1983b). Exponential models with affine dual foliations. *Ann. Statist.* **11** 753–769.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1983c). Reproductive exponential families. *Ann. Statist.* **11** 770–782.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1988a). Combination of reproductive models. *Ann. Statist.* **16** 323–341.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1988b). Reproductive models. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **8**. Wiley, New York.

BARNDORFF-NIELSEN, O. E., BLÆSILD, P., JENSEN, J. L. and JØRGENSEN, B. (1982). Exponential transformation models. *Proc. Roy. Soc. London Ser. A* **379** 41–65.

BARNDORFF-NIELSEN, O. E., BLÆSILD, P., JENSEN, J. L. and SØRENSEN, M. (1985). The fascination of sand. In *A Celebration of Statistics, The ISI Centenary Volume* (A. C. Atkinson and S. E. Fienberg, eds.) 57–87. Springer, Berlin.

BARNDORFF-NIELSEN, O. E. and CHRISTIANSEN, C. (1985). The hyperbolic shape triangle and classification of sand sediments. In *Proc. of International Workshop on the Physics of Blown Sand* (O. E. Barndorff-Nielsen, J. T. Møller, K. Rømer Rasmussen and B. B. Willetts, eds.) 649–676. Dept. Theoretical Statistics, Aarhus Univ.

BARNDORFF-NIELSEN, O. E. and CHRISTIANSEN, C. (1988). Erosion, deposition and size distributions of sand. *Proc. Roy. Soc. London Ser. A.* To appear.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 279–312.

BOOTHBY, W. M. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic, New York.

CHENTSOV, N. N. (1972). *Statistical Decision Rules and Optimal Conclusions* (in Russian). Nauka, Moscow. English translation (1982), Amer. Math. Soc., Providence, R.I.

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

EGUCHI, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.* **11** 793–803.

FRASER, D. A. S. (1979). *Inference and Linear Models.* McGraw-Hill, New York.

GLYNN, W. J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the population parameters have arbitrary multiplicity. *Ann. Statist.* **8** 958–976.

HERMANN, R. (1975). *Gauge Fields and Cartan–Ehresmann Connections, Part A. Interdisciplinary Mathematics* **10**. Mathematical Sciences Press, Brookline, Mass.

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

JØRGENSEN, B. (1983). Maximum likelihood estimation and large-sample inference for generalized and nonlinear regression models. *Biometrika* **70** 19–28.

JØRGENSEN, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.* **13** 187–197.

JØRGENSEN, B. (1987). Exponential dispersion models (with discussion). *J. Roy. Statist. Soc. Ser. B* **49** 127–162.

KOBAYASHI, S. and NOMIZU, K. (1969). *Foundations of Differential Geometry* **2**. Interscience, New York.

LANG, S. (1972). *Differential Manifolds*. Addison-Wesley, Reading, Mass.

LAURITZEN, S. L. (1987). Statistical manifolds. In *Differential Geometry in Statistical Inference* 163–216. IMS, Hayward, Calif.

O'NEILL, B. (1966). The fundamental equations of a submersion. *Michigan Math. J.* **13** 459–469.

PATEFIELD, W. M. (1977). On the maximized likelihood function. *Sankhyā Ser. B* **39** 92–96.

REEDS, J. (1975). Discussion of "Defining the curvature of a statistical problem (with applications to second order efficiency)" by B. Efron. *Ann. Statist.* **3** 1234–1238.

RÉMON, M. (1984). On a concept of partial sufficiency: *L*-sufficiency. *Internat. Statist. Rev.* **52** 127–135.

RICHARDS, F. S. G. (1961). A method of maximum likelihood estimation. *J. Roy. Statist. Soc. Ser. B* **23** 469–475.

SKOVGAARD, L. T. (1984). A Riemannian geometry of the multivariate normal model. *Scand. J. Statist.* **11** 211–223.

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

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF ST. ANDREWS
NORTH HAUGH
ST. ANDREWS KY16 9SS
UNITED KINGDOM