The Annals of Statistics
1993, Vol. 21, No. 2, 1018–1039

ORTHOGEODESIC MODELS

By O. E. Barndorff-Nielsen and P. Blæsild

Aarhus University

A variety of exponential models with affine dual foliations have been noted to possess certain rather similar statistical properties. To give a precise meaning to what has been conceived as "similar," we here propose a set of five conditions, of a differential geometric/statistical nature, that specify the class of what we term orthogeodesic models. It is discussed how these conditions capture the properties in question, and it is shown that some important nonexponential models turn out to satisfy the conditions, too. The conditions imply, in particular, a higher-order asymptotic independence result. A complete characterization of the structure of exponential orthogeodesic models is derived.

1. Introduction. Some composite transformation models have properties similar to certain of the properties of exponential models with affine dual foliations. The properties in question are of a geometrical nature. The present paper represents an attempt to give a unified delineation of those properties, in differential geometric terms. More specifically, we define, by purely differential geometric conditions, a class of parametric statistical models which we call *orthogeodesic models* and which comprises the exponential models with affine dual foliations as well as all the composite transformation models that we had noted for their similarity with such exponential models. Except for general smoothness assumptions, we use five conditions to define an orthogeodesic model. These are stated in Section 3, which also contains a number of examples. Section 4 consists of a discussion of the implications of the four defining conditions; in particular, it contains a result on higher-order asymptotic independence. In Section 5 we study what further properties can be inferred if the model is assumed to be exponential. Section 2 reviews some basic concepts from statistical differential geometry and establishes the notation used throughout the paper.

2. Preliminaries. In this section we introduce some notation and review those concepts from the theory of statistical manifolds [as described, for instance, in Lauritzen (1987)] which are needed throughout the rest of the paper.

We consider a *statistical model* $\mathscr{M}$, that is, a set of probability measures, on a *sample space* $\mathscr{X}$ and assume that $\mathscr{M}$ is a d-dimensional *differentiable*

Received February 1991; revised December 1991.

AMS 1991 *subject classifications*. Primary 62E15, secondary 62F99.

Key words and phrases. Affine connections, α-connections, expected information, exponential models, exponential transformation models, flat submanifolds, geodesic submanifolds, higher-order asymptotic independence, location-scale models, parallel submanifolds, pivot, proper dispersion models, statistical-differential geometry, Student distribution, τ-parallel models, θ-parallel models.

manifold. For simplicity of exposition we suppose, moreover, that $\mathscr{M}$ may be covered by a single chart or, equivalently, that $\mathscr{M} = \{P_\omega: \omega \in \Omega\}$, where Ω is an open subset of R^d and where the mapping φ taking ω into P_ω is smooth and one-to-one; this assumption is not essential and our theoretical results hold without it. The set Ω is the *parameter space* and we use suffixes $r, s, t, \dots$ to denote generic components of the *parameter* $\omega = (\omega^1, \dots, \omega^d)$. Furthermore, we assume that $\mathscr{M}$ is dominated by some σ-finite measure μ and we let p denote the *model function*, that is,

$$p(x;\omega) = \frac{dP_\omega}{d\mu}(x), \qquad x \in \mathscr{X}, \qquad \omega \in \Omega,$$

which is assumed to be positive.

We let $\partial_{r\omega}$ or simply ∂_r denote the coordinate frame $\partial/\partial\omega^r$ at ω, and for an arbitrary real-valued function f defined on Ω we write $f_{/r} = \partial_r f(\omega)$, $f_{/rs} = \partial_r \partial_s f(\omega)$, and so on. Thus, in particular, $l_{/r}$ and $l_{/rs}$ are the two first derivatives of the *log-likelihood function*

$$l(\omega) = l(\omega; x) = \log p(x;\omega).$$

With a slight abuse of notation we later on let ∂_{rm} or simply ∂_r denote the element in $T_m\mathscr{M}$, the tangent space to $\mathscr{M}$ at $m = P_\omega$, given by

$$\partial_r f = \frac{\partial}{\partial\omega^r} f(\varphi(\omega))$$

for any smooth real-valued function f defined in a neighborhood $U_m \subset \mathscr{M}$ of m.

We impose the standard conditions that $E_\omega\{l_{/r}\} = 0$ and $E_\omega\{l_{/r} l_{/s}\} = E_\omega\{-l_{/rs}\}$, where E_ω indicates mean value under the probability measure P_ω, and that the *expected* (or Fisher) *information matrix* $i(\omega) = \{i_{rs}(\omega)\}$ is positive definite, and we use $\{i^{rs}(\omega)\}$ to denote the inverse $i^{-1}(\omega)$. Under these assumptions $\mathscr{M}$ is a Riemannian manifold, the metric on $\mathscr{M}$ being the *expected information metric* which in the local coordinates ω is given by $i(\omega)$.

Any *affine connection* ∇ on the tangent bundle $T\mathscr{M}$ may be characterized in the local coordinate system ω by the *upper Christoffel symbols* Γ^r_{st}, the relation being

$$\nabla_{\partial_s} \partial_t = \Gamma^r_{st} \partial_r, \tag{2.1}$$

or, equivalently, using the expected information metric, by the *lower Christoffel symbols* given by

$$\Gamma_{stu} = \Gamma^r_{st} i_{ru}. \tag{2.2}$$

In (2.1) and (2.2) summation over repeated indices is assumed, that is, here, as throughout the paper, we adopt the Einstein summation convention.

If θ is an alternative set of local coordinates for which generic components are indicated by the letters $\rho, \sigma, \tau \dots$, the upper Christoffel symbols satisfy the *transformation law*

$$\Gamma^r_{st} = \{\Gamma^\rho_{\sigma\tau}\theta^\sigma_{/s}\theta^\tau_{/t} + \theta^\rho_{/st}\}\omega^r_{/\rho}. \tag{2.3}$$

Here and later on we suppress notationally the dependence of a quantity on a particular coordinate system since the indices indicate the dependence.

We now review the definition of the *expected α-connections* on $\mathscr{M}$ introduced by Chentsov (1972) and Amari (1980, 1982); see also Amari (1985, 1987). The expected 0-connection $\overset{0}{\nabla}$ is the Riemannian connection or the Levi–Civita connection, given by the lower Christoffel symbols

$$\overset{0}{\Gamma}_{rst} = \tfrac{1}{2}\{i_{rt/s} + i_{st/r} - i_{rs/t}\}, \tag{2.4}$$

and for $\alpha \in R$ the expected α-connection $\overset{\alpha}{\nabla}$ is defined by

$$\overset{\alpha}{\Gamma}_{rst} = \overset{0}{\Gamma}_{rst} - \frac{\alpha}{2}T_{rst}, \tag{2.5}$$

where T_{rst} is the *expected skewness tensor*, that is,

$$T_{rst} = E_\omega\{l_{/r}l_{/s}l_{/t}\}. \tag{2.6}$$

For later use note that, as seen from (2.5) and (2.6),

$$T_{rst} = 2\left(\overset{0}{\Gamma}_{rst} - \overset{1}{\Gamma}_{rst}\right) \tag{2.7}$$

and

$$\overset{\alpha}{\Gamma}_{rst} = (1-\alpha)\overset{0}{\Gamma}_{rst} + \alpha\overset{1}{\Gamma}_{rst}. \tag{2.8}$$

The *α-Riemannian curvature tensor* $\overset{\alpha}{R}$ has components

$$\overset{\alpha}{R}_{rstu} = \overset{\alpha}{R}{}^v_{rst}i_{uv}, \tag{2.9}$$

where $\overset{\alpha}{R}{}^v_{rst}$ is determined by

$$\overset{\alpha}{\nabla}_{\partial_r}\overset{\alpha}{\nabla}_{\partial_s}\partial_t - \overset{\alpha}{\nabla}_{\partial_s}\overset{\alpha}{\nabla}_{\partial_r}\partial_t = R^v_{rst}\partial_v.$$

The manifold $\mathscr{M}$ is said to be *α-flat* if $\overset{\alpha}{R} \equiv 0$. If $\mathscr{M}$ is α-flat there exists a coordinate system in which the Christoffel symbols of the connection $\overset{\alpha}{\nabla}$ are all 0. The parameter ω of the model $\mathscr{M}$ is said to be *α-affine* if $\overset{\alpha}{\Gamma}_{rst} = 0$ (or $\overset{\alpha}{\Gamma}{}^r_{st} = 0$). The 1-connection $\overset{1}{\nabla}$ is often referred to as the *exponential connection* due to the fact that the canonical parameter of an exponential model is 1-affine (cf. Sections 3 and 5). We also note that the mean value parameter of $\mathscr{M}$ is -1-affine. Thus exponential models are ± 1-flat.

In this paper we are interested in the situation where there exists a parametrization of $\mathscr{M}$ of the form $\omega = (\chi, \psi)$, where χ and ψ are variation independent. The domains of variation of χ and ψ are denoted by X and Ψ, respectively, and the variation independence means that $\Omega = X \times \Psi$. The dimensions of the subparameters χ and ψ are called d_χ and d_ψ, respectively. Furthermore, generic coordinates of χ and ψ are indicated by $\chi^a, \chi^b, \chi^c, \dots$ and $\psi^i, \psi^j, \psi^k, \dots$, respectively. For fixed χ we use $\mathscr{M}_\chi$ to denote the submodel $\{P_{(\chi,\psi)}\colon \psi \in \Psi\}$ and similarly we let $\mathscr{M}_\psi = \{P_{(\chi,\psi)}\colon \chi \in X\}$. In these circumstances the tangent space to $\mathscr{M}$ at $m = P_{(\chi,\psi)}$ is the direct sum of the tangent spaces to $\mathscr{M}_\psi$ and $\mathscr{M}_\chi$, that is,

$$\begin{aligned} T_m\mathscr{M} &= T_m\mathscr{M}_\chi \oplus T_m\mathscr{M}_\psi \\ &= \operatorname{span}\{\partial_{am}\} \oplus \operatorname{span}\{\partial_{jm}\} \end{aligned} \tag{2.10}$$

and

$$\overset{\alpha}{\nabla}_{\partial_r}\partial_s = \overset{\alpha}{\Gamma}{}^a_{rs}\partial_a + \overset{\alpha}{\Gamma}{}^j_{rs}\partial_j. \tag{2.11}$$

The *induced* (or inherited) *α-connection* ${}_\chi\overset{\alpha}{\nabla}$ on $\mathscr{M}_\chi$ is obtained from $\overset{\alpha}{\nabla}$ by projection. More specifically, if p_χ denotes the orthogonal projection on $T\mathscr{M}_\chi$ with respect to the expected information metric i, one has

$${}_\chi\overset{\alpha}{\nabla}_Y Z = p_\chi\left(\overset{\alpha}{\nabla}_Y Z\right) \tag{2.12}$$

for any pair of smooth vector fields (Y, Z) in $T\mathscr{M}_\chi$. Furthermore, the difference between $\overset{\alpha}{\nabla}$ and ${}_\chi\overset{\alpha}{\nabla}$, that is,

$${}_\chi\overset{\alpha}{H}(Y, Z) = \overset{\alpha}{\nabla}_Y Z - {}_\chi\overset{\alpha}{\nabla}_Y Z \tag{2.13}$$

is the *α-embedding curvature* (or α-shape tensor or Euler–Schouten curvature) of the submanifold $\mathscr{M}_\chi$. In local coordinates formula (2.13), which is often referred to as the *Gauss formula* [cf. Vos (1989)], becomes

$${}_\chi\overset{\alpha}{H}(\partial_j, \partial_k) = \overset{\alpha}{\Gamma}{}^a_{jk}\left(\partial_a - p_\chi(\partial_a)\right). \tag{2.14}$$

The α-embedding curvature ${}_\chi\overset{\alpha}{H}$ is determined by the components $\overset{\alpha}{H}{}^r_{jk}$ given by

$${}_\chi\overset{\alpha}{H}(\partial_j, \partial_k) = \overset{\alpha}{H}{}^r_{jk}\partial_r \tag{2.15}$$

or, equivalently, by the quantities

$$\overset{\alpha}{H}_{jks} = \overset{\alpha}{H}{}^r_{jk} i_{rs}. \tag{2.16}$$

The submanifold $\mathscr{M}_\chi$ is said to be *α-geodesic* if $\overset{\alpha}{\nabla}_Y Z$ is a smooth vector field in $T\mathscr{M}_\chi$ for any pair of smooth vector fields (Y, Z) in $T\mathscr{M}_\chi$ or, equivalently, because of (2.12) and (2.13), if the α-embedding curvature ${}_\chi\overset{\alpha}{H}$ is identically 0. Furthermore, using (2.11) or (2.14), one has that $\mathscr{M}_\chi$ is α-geodesic if and only if

$$\overset{\alpha}{\Gamma}{}^a_{jk}(\chi, \psi) = 0, \quad \text{for all } \psi \in \Psi. \tag{2.17}$$

The submanifold $\mathscr{M}_\chi$ is called *geodesic* if $\mathscr{M}_\chi$ is α-geodesic for all $\alpha \in R$.

We conclude this section with some remarks concerning a special case which is of particular interest in the present paper, namely the case where the sum in (2.10) is a direct orthogonal sum with respect to the expected information metric i, that is,

$$i_{aj}(\chi, \psi) = 0. \tag{2.18}$$

In this case we say that the subparameters χ and ψ are *expected* (*or i-*) *orthogonal* and since $p_\chi(\partial_a) = 0$ it follows from (2.14) and (2.15) that

$$\overset{\alpha}{H}{}^a_{jk} = \overset{\alpha}{\Gamma}{}^a_{jk} \tag{2.19}$$

and

$$\overset{\alpha}{H}{}^i_{jk} = 0 \tag{2.20}$$

or, equivalently, using (2.2), (2.16) and (2.18), that

$$\overset{\alpha}{H}_{jkb} = \overset{\alpha}{\Gamma}_{jkb} \tag{2.21}$$

and

$$\overset{\alpha}{H}_{jkl} = 0. \tag{2.22}$$

The formulas for the submanifold $\mathscr{M}_\psi$ analogous to (2.19)–(2.22) are

$$\overset{\alpha}{H}{}^i_{bc} = \overset{\alpha}{\Gamma}{}^i_{bc}, \tag{2.23}$$

$$\overset{\alpha}{H}{}^a_{bc} = 0, \tag{2.24}$$

$$\overset{\alpha}{H}_{bcj} = \overset{\alpha}{\Gamma}_{bcj} \tag{2.25}$$

and

$$\overset{\alpha}{H}_{bcd} = 0. \tag{2.26}$$

For the α-Riemannian curvature tensor ${}_\chi\overset{\alpha}{R}$ of the induced connection ${}_\chi\overset{\alpha}{\nabla}$ on $\mathscr{M}_\chi$, one has the *Gauss equation* [cf. Vos (1989)]

$${}_\chi\overset{\alpha}{R}_{jklm} = \overset{\alpha}{R}_{jklm} + \left(\overset{-\alpha}{H}_{jma}\overset{\alpha}{H}_{klb} - \overset{\alpha}{H}_{jla}\overset{-\alpha}{H}_{kmb}\right)i^{ab}. \tag{2.27}$$

Note finally the convenient fact that if the subparameters χ and ψ are expected orthogonal, then the Christoffel symbols of the α-connection for the submodel $\mathscr{M}_\chi$, considered as a model on its own, are equal to those α-Christoffel symbols from the full model whose indices correspond to the submodel, that is, the α-connection on $\mathscr{M}_\chi$ is given by $\overset{\alpha}{\Gamma}{}^i_{jk}$.

3. Orthogeodesic models: definition and examples.

DEFINITION 3.1. The model $\mathscr{M}$ is said to be *orthogeodesic* if there exists a parametrization $\omega = (\chi, \psi)$ such that the following conditions are satisfied:

(o) χ and ψ are variation independent.

(i) χ and ψ are expected orthogonal, that is,

$$i_{bj} = 0.$$

(ii) The ψ-part of the information matrix i depends on ψ only, that is,

$$i_{jk}(\chi, \psi) = i_{jk}(\psi).$$

(iii) For every value of χ the submanifold $\mathscr{M}_\chi$ is expected 1-geodesic, that is,

$$\overset{1}{\Gamma}{}^{a}_{jk} = 0.$$

(iv) For every value of χ the submanifold $\mathscr{M}_\chi$ is expected 1-flat and ψ is a 1-affine coordinate, that is,

$$\overset{1}{\Gamma}{}^{i}_{jk} = 0.$$

In this case the parametrization (χ, ψ) is said to be *ortho-affine*.

Definition 3.1 is formulated in terms of local coordinates since this is the most convenient from an application point of view. However, the concept of an orthogeodesic model is a geometric concept. To emphasize this, we give in Definition 3.1′ an equivalent formulation of the concept without reference to local coordinates.

DEFINITION 3.1′. The model $\mathscr{M}$ is *orthogeodesic* if the following conditions are satisfied:

(o)′ $\mathscr{M}$ is a product manifold of the form $\mathscr{M} = X \times \Psi$, where X and Ψ are differentiable manifolds.

(i)′ The factorization of $\mathscr{M}$ is orthogonal with respect to the expected information metric i on $\mathscr{M}$.

(ii)′ For every value of χ the restriction of the metric i to the submanifold $\mathscr{M}_\chi$ does not depend on χ.

(iii)′ For every value of χ and for some value $\alpha \neq 0$, the submanifold $\mathscr{M}_\chi$ is expected α-geodesic, that is, the α-embedding curvature ${}_\chi\overset{\alpha}{H}$ vanishes identically.

(iv)′ For every value of χ the submanifold $\mathscr{M}_\chi$ is expected 1-flat, that is, the Riemannian curvature tensor ${}_\chi\overset{1}{R}$ vanishes identically.

The equivalence of the two definitions is established after the proof of Theorem 4.2 in Section 4 below.

REMARK 3.1. Since a one-dimensional manifold is automatically (1-)flat, condition (iv) [or (iv)′] is fulfilled if ψ is one dimensional. In the beginning of Section 4 we show that condition (iv) is also redundant if $\mathscr{M}$ is an exponential model.

EXAMPLE 3.1. In this example we consider a location-scale model $\mathscr{M}$ on R with model function of the form

$$p(x; \chi, \sigma) = \sigma^{-1} f((x - \chi)/\sigma), \tag{3.1}$$

with x and χ in R and with $\sigma > 0$. We assume that f is positive and symmetric around 0. Furthermore, if $g = \log f$ we suppose that g is at least twice continuously differentiable and satisfies the conditions

$$\int y g'(y) f(y)\, dy = -1 \tag{3.2}$$

and

$$I_j = I_j(f) = \int (1 + y g'(y))^j f(y)\, dy < \infty, \qquad j = 2, 3. \tag{3.3}$$

Since the variation independent subparameters χ and σ are both one dimensional, we use in this case the parameters themselves rather than indices of the parameter components to indicate differentiation, and so on. Setting

$$y = (x - \chi)/\sigma,$$

the log-likelihood function and its first derivatives become

$$l(\chi, \sigma) = -\log \sigma + g(y),$$
$$l_{/\chi} = -\sigma^{-1} g'(y)$$

and

$$l_{/\sigma} = -\sigma^{-1}(1 + y g'(y)). \tag{3.4}$$

The conditions stated previously ensure that $E_{(\chi,\sigma)}\{l_{/*}\} = 0$ and that $i_{**}(\chi, \sigma) = E_{(\chi,\sigma)}\{l_{/*} l_{/*}\} = E_{(\chi,\sigma)}\{-l_{/**}\}$. [In fact, (3.2) is just a special case of $E_{(\chi,\sigma)}\{l_{/\sigma}\} = 0$.] Noticing that $l_{/\chi}$ and $l_{/\sigma}$ are, respectively, odd and even as functions of y, it follows that

$$i_{\chi\sigma} = E_{(\chi,\sigma)}\{l_{/\chi} l_{/\sigma}\} = 0 \tag{3.5}$$

and

$$T_{\sigma\sigma\chi} = E_{(\chi,\sigma)}\{l_{/\sigma} l_{/\sigma} l_{/\chi}\} = 0. \tag{3.6}$$

Furthermore, using (3.3) and (3.4), we get that

$$i_{\sigma\sigma} = \sigma^{-2} I_2. \tag{3.7}$$

Consequently, as seen from (3.5) and (3.7), the parametrization (χ, σ) satisfies conditions (i) and (ii) in Definition 3.1.

Using (2.4), (3.5) and (3.7), we obtain

$$\overset{0}{\Gamma}_{\sigma\sigma\chi} = 0,$$

which together with (2.5) and (3.6) imply that

$$\overset{1}{\Gamma}_{\sigma\sigma\chi} = 0$$

and using (3.5) we find that

$$\overset{1}{\Gamma}{}^{\chi}_{\sigma\sigma} = 0,$$

that is, the submanifold $\mathscr{M}_\chi$ is expected 1-geodesic.

The formulas (2.4) and (3.7) imply that

$$\overset{0}{\Gamma}_{\sigma\sigma\sigma} = \tfrac{1}{2} i_{\sigma\sigma/\sigma} = -\sigma^{-3} I_2 \tag{3.8}$$

and from (2.6), (3.3) and (3.4) it follows that

$$T_{\sigma\sigma\sigma} = -\sigma^{-3} I_3. \tag{3.9}$$

Using (2.5), (3.8) and (3.9), we obtain

$$\overset{1}{\Gamma}_{\sigma\sigma\sigma} = -\sigma^{-3}\left[I_2 - \tfrac{1}{2} I_3\right]$$

from which we may conclude that the submanifold $\mathscr{M}_\chi$ is expected 1-flat with σ as 1-affine coordinate if and only if $2I_2 = I_3$ which is generally not the case. Since the submanifold $\mathscr{M}_\chi$ is one dimensional, it follows from Remark 3.1 that $\mathscr{M}_\chi$ is expected 1-flat. Consequently, if $2I_2 \neq I_3$ it is possible to find an alternative 1-affine parametrization of $\mathscr{M}_\chi$.

For the cases where $I_2 \neq 0$ and $I_3 \neq 0$, we now demonstrate that replacing the subparameter σ by ψ, given by

$$\psi(\sigma) = \sigma^{1/c}, \tag{3.10}$$

where

$$c = 2I_2(f)/I_3(f), \tag{3.11}$$

that is, a quantity depending on the model, we obtain a parametrization (χ, ψ) which is ortho-affine. Because of the facts stated previously it suffices to show that

$$\overset{1}{\Gamma}_{\psi\psi\psi} = 0. \tag{3.12}$$

Since $\sigma(\psi) = \psi^c$ one has $\sigma_{/\psi} = c\psi^{c-1}$ and formulas (3.7) and (3.9) imply that

$$i_{\psi\psi} = i_{\sigma\sigma}(\sigma_{/\psi})^2 = \psi^{-2} c^2 I_2$$

and

$$T_{\psi\psi\psi} = T_{\sigma\sigma\sigma}(\sigma_{/\psi})^3 = -\psi^{-3} c^3 I_3$$

from which we find, using (2.4) and (2.5), that

$$\begin{aligned}\overset{1}{\Gamma}_{\psi\psi\psi} &= \tfrac{1}{2} i_{\psi\psi/\psi} - \tfrac{1}{2} T_{\psi\psi\psi} \\ &= \tfrac{1}{2} c^2 \psi^{-3}(-2I_2 + cI_3),\end{aligned}$$

and formula (3.12) follows from (3.11).

If f denotes the probability density function of the standard normal distribution, that is,

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

it is easily seen that $I_2 = 2$ and $I_3 = -8$. From (3.10) and (3.11) we find that the location-scale model (3.1) with normal errors is orthogeodesic with ortho-affine parameter (χ, σ^{-2}).

As a final illustration suppose that f is the probability density function of the t distribution with $\nu(>0)$ degrees of freedom, that is,

$$f(y) = \frac{1}{\sqrt{\nu}\, B(1/2, 1/2\nu)} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}, \tag{3.13}$$

where B denotes the beta function. From formula 8.380.3 in Gradshteyn and Ryzhik (1965) it follows that

$$\int_{-\infty}^{\infty} \left(1 + \frac{y^2}{\nu}\right)^{-\beta} y^{2\alpha}\, dy = \nu^{\alpha+1/2} B(\alpha + 1/2, \beta - (\alpha + 1/2)) \tag{3.14}$$

for $\alpha > -1/2$ and $\beta > \alpha + 1/2$. Using (3.14), one finds after some calculations that

$$I_2 = \frac{2\nu}{\nu + 3}$$

and

$$I_3 = \frac{-8\nu(\nu - 1)}{(\nu + 5)(\nu + 3)}.$$

Consequently, unless $\nu = 1$, which corresponds to the Cauchy distribution, the location-scale model based on the distribution (3.13) is orthogeodesic with ortho-affine parameter $(\chi, \sigma^{1/c})$, where

$$c = \frac{-(\nu + 5)}{2(\nu - 1)}.$$

The location-scale model corresponding to the Cauchy distribution is easily seen to be orthogeodesic with ortho-affine parameter $(\chi, \log \sigma)$.

Finally, note that the results for the normal location-scale model may be obtained as limiting cases of those for the t distribution by letting $\nu \to \infty$.

The rest of the examples we will consider here are concerned with exponential models. The model function (w.r.t. some dominating measure) for such a model $\mathscr{M}$ is of the form

$$\exp\{\theta^\rho t_\rho(x) - \kappa(\theta) - \varphi(x)\}. \tag{3.15}$$

We assume that the order of $\mathscr{M}$ is d and, as indicated in (3.15), we use the letters $\rho, \sigma, \ldots$ to denote generic coordinates of the canonical parameter $\theta = (\theta^1, \ldots, \theta^d)$ and of the canonical statistic $t(x) = (t_1(x), \ldots, t_d(x))$. Furthermore, in the terminology of Barndorff-Nielsen (1988), we suppose that $\mathscr{M}$ is a *core exponential model*, that is, the parameter domain of $\mathscr{M}$ equals int Θ, the interior of the canonical parameter domain Θ of the full exponential model generated by $\mathscr{M}$, and that the full model is steep. The mean value mapping defined on int Θ is denoted by τ, that is, $\tau = (\tau_1, \ldots, \tau_d) = (E_\theta\{t_1\}, \ldots, E_\theta\{t_d\}) = E_\theta\{t\}$, and we let $\mathscr{T} = \tau(\text{int } \Theta)$.

In Section 5 we prove a theorem characterizing the structure of exponential orthogeodesic models and using this theorem it is easily established, as shown in Section 5, that the models in Examples 3.2 and 3.3 are indeed orthogeodesic.

EXAMPLE 3.2 (τ-parallel models and θ-parallel models). These two types of models were introduced and studied in Barndorff-Nielsen and Blæsild (1983a, b). In both cases a partition $(\theta_{(1)}, \theta_{(2)})$ of the canonical parameter θ is considered, the dimensions of $\theta_{(1)}$ and $\theta_{(2)}$ being denoted by $d_{(1)}$ and $d_{(2)}$, respectively. Letting $(\tau^{(1)}, \tau^{(2)})$ denote a similar partition of the mean value τ, it is shown in Barndorff-Nielsen and Blæsild (1983a) that the components of the mixed parameter $(\tau^{(1)}, \theta_{(2)})$ are variation independent and similarly for the other mixed parameter $(\theta_{(1)}, \tau^{(2)})$. These mixed parameters play for the two model types the role of the parameter (χ, ψ) in Definition 3.1. Consequently, we use the letters $a, b, c, \dots$ and $i, j, k, \dots$ to denote generic coordinates of, respectively, the first and second components in the mixed parameters. With $\mathscr{T}^{(1)}$ we denote the domain of variation of $\tau^{(1)}$ and for fixed $\tau_0^{(1)} \in \mathscr{T}^{(1)}$ we let $\mathscr{T}_{\tau_0^{(1)}}$ denote the set of τ values whose first component is τ_0, that is, $\mathscr{T}_{\tau_0^{(1)}} = \{\tau \colon \tau^{(1)} = \tau_0^{(1)}\}$. In a similar way we define $\Theta_{(1)}$ and $(\operatorname{int} \Theta)_{\theta_{(1)}}$.

An exponential model $\mathscr{M}$ possessing a τ-parallel foliation or briefly a *τ-parallel model* is orthogeodesic with the mixed parameter $(\tau^{(1)}, \theta_{(2)})$ being ortho-affine and the model (3.15) is τ-parallel if and only if $\theta_{(1)}(\tau^{(1)}, \theta_{(2)})$ is of the form

$$\theta_{(1)}^{a}\big(\tau^{(1)}, \theta_{(2)}\big) = -\theta_{(2)}^{i} h_i^{a}(\tau^{(1)}) + k^{a}(\tau^{(1)}) \tag{3.16}$$

or, equivalently, if and only if $\tau^{(2)}(\tau^{(1)}, \theta_{(2)})$ is of the form

$$\tau_i^{(2)}\big(\tau^{(1)}, \theta_{(2)}\big) = H_i(\tau^{(1)}) + m_i\big(\theta_{(2)}\big) \tag{3.17}$$

for certain (vector) functions h, k, H and m. The quantity h in (3.16) is obtainable from H in (3.17) by differentiation, that is, $h_i^{a}(\tau^{(1)}) = \partial H_i / \partial \tau_a^{(1)}$. Thus for a τ-parallel model the foliation of $\mathscr{T}$ with parallel leaves $\{\mathscr{T}_{\tau^{(1)}} \colon \tau^{(1)} \in \mathscr{T}^{(1)}\}$ corresponds to a foliation of $\operatorname{int} \Theta$ into affine subspaces. The subclass of τ-parallel models for which $k^{a} = 0$ and the statistic $t(x)$ is of the form $(x, H(x))$ possesses some particularly nice properties. If $x_1, \dots, x_n$ is a sample from such a model and if $\bar{x} = n^{-1}(x_1 + \cdots + x_n)$, it is shown in Barndorff-Nielsen and Blæsild (1983a) that the components $\hat{\tau}^{(1)}$ and $\hat{\theta}_{(2)}$ of the maximum likelihood estimator of the mixed parameter $(\tau^{(1)}, \theta_{(2)})$ are stochastically independent, that is, $\hat{\tau}^{(1)} \perp\!\!\!\perp \hat{\theta}_{(2)}$, or, equivalently, that $\bar{x} \perp\!\!\!\perp (\overline{H} - H(\bar{x}))$, where $\overline{H} = n^{-1}(H(x_1) + \cdots + H(x_n))$. Furthermore, one has that the distribution of $\bar{x}$ is given by (3.15) with θ replaced by $n\theta$. If $d = 2$ and $d_{(1)} = 1$ it is proved in Blæsild and Jensen (1985) that the only such models are those corresponding to the normal distribution, the gamma distribution and the inverse Gaussian distribution, the function $H(x)$ being x^2, $\log x$ and x^{-1}, respectively. However, as explained in Barndorff-Nielsen and Blæsild (1988) these models may be combined according to a certain scheme to give models of the kind in question for higher-dimensional variates.

As a side remark, note that for the normal distribution with mean χ and variance σ^2 the mixed parameter $(\tau^{(1)}, \theta_{(2)})$ corresponding to the canonical statistic (x, x^2) is $(\chi, -1/2\sigma^{-2})$ in agreement with the results in Example 3.1.

The *θ-parallel models* are orthogeodesic with the mixed parameter $(\theta_{(1)}, \tau^{(2)})$ being ortho-affine and the model (3.15) is θ-parallel if and only if $\tau^{(1)}(\theta_{(1)}, \tau^{(2)})$ is of the form

$$\tau_a^{(1)}\left(\theta_{(1)}, \tau^{(2)}\right) = -\tau_j^{(2)} h_a^j(\theta_{(1)}) + k_a(\theta_{(1)}) \tag{3.18}$$

or, equivalently, if and only if $\theta_{(2)}(\theta_{(1)}, \tau^{(2)})$ is of the form

$$\theta_{(2)}^j\left(\theta_{(1)}, \tau^{(2)}\right) = H^j(\theta_{(1)}) + m^j(\tau^{(2)}), \tag{3.19}$$

where $h_a^j(\theta_{(1)}) = \partial H^j(\theta_{(1)})/\partial\theta_{(1)}^a$. Hence for a θ-parallel model the foliation of int Θ with parallel leaves $\{(\text{int }\Theta)_{\theta_{(1)}}: \theta_{(1)} \in \text{int }\Theta_{(1)}\}$ corresponds to a foliation of $\mathscr{T}$ into affine subspaces.

The existence of a θ-parallel foliation is equivalent to the existence of a proper cut in $\mathscr{M}$ and thus this concept is intimately related to the kind of likelihood independence known as S-ancillarity and S-sufficiency. [For details, see Barndorff-Nielsen (1978) and Barndorff-Nielsen and Blæsild (1983a).]

EXAMPLE 3.3. Suppose that the exponential model (3.15) is also a transformation model, that is, there exists a group G acting on the sample space $\mathscr{X}$ and $gP \in \mathscr{M}$ for every $P \in \mathscr{M}$ and every $g \in G$. Here gP denotes the measure P lifted by the transformation corresponding to g, that is, $(gP)(A) = P(g^{-1}A)$ for every measurable set A.

Assuming that the model function (3.15) is w.r.t. some invariant measure on $\mathscr{X}$, the group G induces, as discussed in Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982), an affine action on $\mathscr{T}$ as well as on int Θ. More specifically, considering t and θ as, respectively, a column vector and a row vector, there exist matrices $A(g)$, column vectors $B(g)$ and row vectors $D(g)$ such that the mappings from G into $GA(d)$, the general affine group, given by, respectively,

$$g \to [A(g), B(g)] \tag{3.20}$$

and

$$g \to [C(g), D(g)], \tag{3.21}$$

where

$$C(g) = A(g)^{-1} = A(g^{-1}) \tag{3.22}$$

are both representations of G. The affine actions on $\mathscr{T}$ and int Θ are given by, respectively,

$$\begin{gathered} G \times \mathscr{T} \to \mathscr{T}, \\ (g, t) \to gt = A(g)t + B(g) \end{gathered} \tag{3.23}$$

and

$$(3.24)\qquad \begin{aligned} &G \times \operatorname{int}\Theta \to \operatorname{int}\Theta,\\ &(g,\theta) \to g\theta = \theta C(g) + D(g). \end{aligned}$$

Furthermore, the model function (3.15) may be rewritten as

$$(3.25)\quad p(x;g\theta) = \exp\{(g\theta)t(x) - [\alpha(\theta) + \delta(g) + (g\theta)B(g)] - \varphi(x)\}$$

[cf. Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982)].

Here we consider *composite transformation models*, that is, models for which the action (3.24) is not transitive. We assume that $\mathscr{M}$ has *constant orbit type* or, equivalently, that there exist a subset Ψ of int Θ and a subgroup K of G such that

$$(3.26)\qquad G_\psi = \{g \in G | g\psi = \psi\} = K, \qquad \forall\, \psi \in \Psi,$$

that is, the isotropy group at ψ is K for all $\psi \in \Psi$. Furthermore, we suppose that $\mathscr{M}$ has *affine orbit representation*, that is, the set Ψ of orbit representatives is the intersection of int Θ and an affine subspace. Without loss of generality we may assume that Ψ is of the form

$$\Psi = \left\{\psi \in \operatorname{int}\Theta | \psi = \left(\tilde{\psi}, 0\right), \tilde{\psi} \in \tilde{\Psi}\right\}$$

for some open subset $\tilde{\psi}$ of R^{d_ψ} and from now on we identify the sets Ψ and $\tilde{\Psi}$. Letting gK denote the left coset $\{gk\colon k \in K\}$, it follows, using (3.24) and (3.26), that the equation

$$(3.27)\qquad \theta = \psi C(g) + D(g)$$

establishes a one-to-one correspondence between θ and (gK, ψ). Finally, assuming that the set of left cosets $\{gK\colon g \in G\}$ may be parametrized by a parameter χ varying in an open subset X of R^{d_χ}, where $d_\chi = d - d_\psi$, one has that χ and ψ are variation independent and, furthermore, that

$$(3.28)\qquad \theta^\rho(\chi,\psi) = \psi^j C_j^\rho(\chi) + D^\rho(\chi)$$

as seen from (3.27). Without loss of generality we suppose that $0 \in X$ and that 0 corresponds to the coset eK, where e denotes the identity element in G.

Inserting (3.28) in (3.25), it follows that

$$(3.29)\quad p(x;\chi,\psi) = \exp\left\{\left(\psi^j C_j^\rho(\chi) + D^\rho(x)\right)t_\rho(x) - \kappa(\chi,\psi) - \varphi(x)\right\},$$

where

$$(3.30)\qquad \kappa(\chi,\psi) = \alpha(\psi) + \gamma(\chi) + \psi^j C_j^\rho(\chi) B_\rho(\chi)$$

and

$$(3.31)\qquad \gamma(\chi) = \delta(\chi) + D^\rho(\chi) B_\rho(\chi).$$

The identity

$$E_{g\theta}\{t\} = E_\theta\{gt\}$$

implies that

$$\tau(\chi, \psi) = A(\chi)\tau(0, \psi) + B(\chi). \tag{3.32}$$

Let $(\theta_{(1)}, \theta_{(2)})$ be a partition of θ such that $\theta_{(1)}$ is d_ψ-dimensional and let $(\tau^{(1)}, \tau^{(2)})$ be a similar partition of τ. Since $C^\rho_\sigma(0) = \delta^\rho_\sigma = (A^\rho_\sigma(0))$, $B_\rho(0) = 0$ and $D^\rho(0) = 0$, it follows from (3.28) that $\theta_{(1)}(0, \psi) = \psi$ and, in addition, from (3.29) and (3.30) that

$$\tau^{(1)}_j(0, \psi) = \alpha_{/j}(\psi) \tag{3.33}$$

and

$$\dim \operatorname{aff}\{\tau^{(1)}(0, \psi) : \psi \in \Psi\} = d_\psi. \tag{3.34}$$

By Theorem 4.2 in Section 4 below, for an orthogeodesic model the submanifold $\mathscr{M}_\chi$ is geodesic and thus in particular -1-geodesic. A necessary condition for the model (3.29) to be orthogeodesic with (χ, ψ) as ortho-affine parameter is therefore that $\mathscr{T}_\chi = \{\tau(\chi, \psi) : \psi \in \Psi\}$ is an affine subspace of $\mathscr{T}$ of dimension d_ψ or, consequently, using (3.34), that $\tau^{(2)}(0, \psi)$ is a constant, which without loss of generality may be assumed to be 0. With this condition it follows from (3.32) and (3.33) that

$$\tau_\rho(\chi, \psi) = \alpha_{/j}(\psi) A^j_\rho(\chi) + B_\rho(\chi). \tag{3.35}$$

The formulas (3.28), (3.35) and (3.30) are identical to the formulas (5.7)–(5.9) in Theorem 5.1 of Section 5 below and formula (3.22) implies (5.2). Thus a transformation model possessing a dual affine foliation given by (3.28) and (3.35) is orthogeodesic if and only if the conditions (5.3)–(5.6) are fulfilled.

In most of the examples of such models, the leaves of the foliation are contained in linear rather than affine subspaces, that is, $B_\rho(\chi) = 0 = D^\rho(\chi)$. For such models it is shown in Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982) that

$$\delta(\chi) = 0. \tag{3.36}$$

From (3.24) it follows that $\gamma(\chi) = 0$ and so the conditions (5.4)–(5.6) are fulfilled, that is, the property of orthogeodesicity is equivalent to condition (5.3) for such models.

EXAMPLE 3.4. Suppose the exponential model (3.15) may be written in the form

$$\exp\{\psi^j R_j(x; \chi) - \alpha(\psi) - \varphi(x)\}, \tag{3.37}$$

where χ and ψ are variation independent and where

$$R_j(x; \chi) = r_j(\chi) + s^\rho_j(\chi) t_\rho(x). \tag{3.38}$$

From (3.37) it follows that

$$l_{/a} = \psi^j R_{j/a}(x; \chi), \tag{3.39}$$

$$l_{/j} = -\alpha_{/j}(\psi) + R_j(x; \chi) \tag{3.40}$$

and

$$l_{/aj} = R_{j/a}(x;\chi).$$

Thus the parameters χ and ψ are expected orthogonal if and only if

$$E_{(\chi,\psi)}\{R_{j/a}(x;\chi)\} = 0. \tag{3.41}$$

If ψ is one dimensional Barndorff-Nielsen and Jørgensen (1991) refer to the model (3.37) as a *proper exponential dispersion model* and in that case (3.41) follows from (3.39) and the condition $E_{(\chi,\psi)}\{l_{/a}\} = 0$. If $\dim \operatorname{aff} \Psi = d_\psi > 1$, formula (3.41) is equivalent, as seen from (3.39), to the condition that $E_{(\chi,\psi)}\{R_{j/a}(x;\chi)\}$ depends on χ only.

Formula (3.40) implies that

$$i_{jk}(\chi,\psi) = \alpha_{/jk}(\psi),$$

that is, condition (ii) of Definition 3.1 is fulfilled. It follows from (3.15), (3.37) and (3.38) that

$$\theta^\rho(\chi,\psi) = \psi^j s_j^\rho(\chi)$$

and, consequently, that

$$\theta^\rho_{/jk} = 0. \tag{3.42}$$

Using (2.3), (3.42) and the fact that $\mathscr{M}$ is 1-flat in θ, one finds that $\overset{1}{\Gamma}{}^r_{jk} = 0$.

Thus, in conclusion, the model (3.37) is orthogeodesic with (χ,ψ) as orthoaffine parameter if and only if the condition (3.41) is fulfilled and this is the case for a proper exponential dispersion model.

Finally, note that for an exponential transformation model with a dual linear foliation, that is, a transformation model of the kind considered in Example 3.3 for which $B_\rho(\chi) = 0 = D^\rho(\chi)$, the model function is of the form (3.37) with

$$r_j(\chi) = 0$$

and

$$s_j^\rho(\chi) = C_j^\rho(\chi),$$

as seen from (3.29)–(3.31) and (3.36). In this case one has, using (3.35), that

$$\begin{aligned} E_{(\chi,\psi)}\{R_{j/a}(x;\chi)\} &= C^\rho_{j/a}(\chi)\tau_\rho(\chi,\psi) \\ &= C^\rho_{j/a}(\chi)A^i_\rho(\chi)\tau_i(0,\psi) \end{aligned}$$

and since $\dim \operatorname{aff}\{\tau^{(1)}(0,\psi)\colon \psi \in \Psi\} = d_\psi$ it follows that the conditions (3.41) and (5.3) are equivalent.

4. Orthogeodesic models: properties. In this section we discuss various implications of the five conditions in Definition 3.1.

Some implications of condition (i) have already been considered at the end of Section 2. Recall that condition (i) implies that the Christoffel symbols for the induced α-connection ${}_\chi\overset{\alpha}{\nabla}$ on the submanifold $\mathscr{M}_\chi$ and the components of the α-embedding curvature ${}_\chi\overset{\alpha}{H}$ are both equal to those Christoffel symbols of the α-connection $\overset{\alpha}{\nabla}$ from the full model $\mathscr{M}$ whose indices are those corresponding to the submodel. Furthermore, one has the following result.

LEMMA 4.1. *Under condition* (i) *the Christoffel symbols of the Riemannian connection* $\overset{0}{\nabla}$ *corresponding to the expected information* i *are given by*

$$\overset{0}{\Gamma}_{bcd} = \tfrac{1}{2}\{i_{bd/c} + i_{cd/b} - i_{bc/d}\}, \tag{4.1}$$

$$\overset{0}{\Gamma}_{kbc} = -\overset{0}{\Gamma}_{bck} = \tfrac{1}{2} i_{bc/k}, \tag{4.2}$$

$$\overset{0}{\Gamma}_{cjk} = -\overset{0}{\Gamma}_{jkc} = \tfrac{1}{2} i_{jk/c}, \tag{4.3}$$

$$\overset{0}{\Gamma}_{jkl} = \tfrac{1}{2}\{i_{jl/k} + i_{kl/j} - i_{jk/l}\} \tag{4.4}$$

and

$$\begin{aligned} \overset{0}{\Gamma}{}^{a}_{tu} &= \overset{0}{\Gamma}_{tub} i^{ab}, \\ \overset{0}{\Gamma}{}^{i}_{tu} &= \overset{0}{\Gamma}_{tuj} i^{ij}. \end{aligned} \tag{4.5}$$

PROOF. Condition (i) implies that $i_{bj/t} = 0$ and the formulas (4.1)–(4.4) follow immediately from (2.4). The formulas (4.5) are obtained from (2.2) by noticing that for the expected formation $i^{-1} = \{i^{rs}\}$ one has $i^{aj} = 0$. □

The implications of the orthogonality of the subparameters χ and ψ concerning the induced connections and the embedding curvatures related to the 0-connection may be summarized as follows.

COROLLARY 4.1. *Under condition* (i) *the Christoffel symbols* ${}_{\chi}\overset{0}{\Gamma}$ *of the connection* ${}_{\chi}\overset{0}{\nabla}$ *on* $\mathscr{M}_{\chi}$ *induced by the* 0-*connection* $\overset{0}{\nabla}$ *are*

$${}_{\chi}\overset{0}{\Gamma}_{jkl} = \overset{0}{\Gamma}_{jkl}, \qquad {}_{\chi}\overset{0}{\Gamma}{}^{i}_{jk} = \overset{0}{\Gamma}{}^{i}_{jk} \tag{4.6}$$

and the components of the 0-*embedding curvature of* $\mathscr{M}_{\chi}$ *are*

$$\overset{0}{H}_{jkc} = \overset{0}{\Gamma}_{jkc} = -\overset{0}{\Gamma}_{cjk} = -\tfrac{1}{2} i_{jk/c}. \tag{4.7}$$

Similar formulas hold for the submanifold $\mathscr{M}_{\psi}$.

PROOF. Formula (4.7) is a consequence of (2.21) and (4.3), and (4.6) is a special case of the first remark in this section. □

THEOREM 4.1. *Suppose condition* (i) *is fulfilled. Then the following conditions are equivalent*:

(ii) $i_{jk}(\chi, \psi) = i_{jk}(\psi)$.

(iia) *For every value of* χ *the submanifold* $\mathscr{M}_{\chi}$ *is* 0-*geodesic* (*with respect to* i).

(iib) $\overset{0}{\Gamma}_{jkc} = 0$ *or, equivalently*, $\overset{0}{\Gamma}{}^{a}_{jk} = 0$.

Note that conditions (ii) and (iia) are concerned with the metric and the embedding curvature, which are geometric quantities often referred to as the first and second fundamental forms, respectively, whereas condition (iib) is expressed in terms of local coordinates and is important from the application point of view.

PROOF. The equivalences follow easily from Lemma 4.1 and formula (4.7). □

Condition (ii) is equivalent to the condition stating that $\{\mathscr{M}_\chi\}_{\chi \in X}$ is a collection of isometric manifolds, the isometry between $\mathscr{M}_\chi$ and $\mathscr{M}_{\chi'}$ being the map $(\chi, \psi) \to (\chi', \psi)$. Thus under (ii) one has, in particular, that a curve in $\mathscr{M}_\chi$ of the form $q(t) = (\chi, \gamma(t))$, $t \in [0, \tau]$, where γ is a curve in Ψ, has a length $L_{\gamma\chi}$ which does not depend on χ, and the submanifolds $\{\mathscr{M}_\psi\}_{\psi \in \Psi}$ are *parallel* in the sense that $L_{\gamma\chi} = L_{\gamma\chi'}$ for all values of χ and χ' and for all curves γ in Ψ. Conversely, it is easily seen that if the submanifolds $\{\mathscr{M}_\psi\}_{\psi \in \Psi}$ are parallel, then condition (ii) is fulfilled.

THEOREM 4.2. *Suppose that conditions* (i) *and* (ii) *are satisfied. Then the following conditions are equivalent*:

(iii) *The submanifolds* $\mathscr{M}_\chi$ *are* 1-*geodesic for every* $\chi \in X$.
(iiia) *The submanifolds* $\mathscr{M}_\chi$ *are geodesic for every* $\chi \in X$.
(iiib) $\overset{\alpha}{\Gamma}_{jkc} = 0$ *or, equivalently,* $\overset{\alpha}{\Gamma}{}^{a}_{jk} = 0$ *for all* $\alpha \in R$.
(iiic) $T_{jkc} = 0$.

PROOF. According to Lauritzen (1987), Proposition 3.12, a submanifold is geodesic if and only if it is α_i-geodesic, $i = 1, 2$, for some $\alpha_1, \alpha_2 \in R$ with $\alpha_1 \neq \alpha_2$. By Theorem 4.1, $\mathscr{M}_\chi$ is 0-geodesic and, consequently, the equivalence of (iii) and (iiia) is established. The rest of the theorem follows from the formulas (2.7) and (2.17). □

PROOF OF THE EQUIVALENCE OF DEFINITIONS 3.1 AND 3.1′. Conditions (o)–(iv) clearly imply (o)′–(ii)′. Furthermore, (i), (iii) and (iv) imply that $\overset{1}{\Gamma}_{jks} = 0$ and it follows, using (2.21) and the remark at the end of Section 2, that $\overset{1}{H}_{jkc} = 0$ and ${}_\chi\overset{1}{\Gamma}_{jkl} = 0$. Conditions (iii)′ and (iv)′ are now consequences of Theorem 4.2 and formula (2.9) (for ${}_\chi\overset{1}{R}$), respectively.

Suppose conversely that (o)′–(iv)′ are fulfilled. These conditions clearly imply (o)–(ii). Let χ be a fixed value in X. Since ${}_\chi\overset{1}{R} \equiv 0$ we may choose a local coordinate system ψ on Ψ such that $\overset{1}{\Gamma}_{jkl}(\chi, \psi) = 0$ from which (iv) follows. According to (iii)′, one has $\overset{\alpha}{H}_{jkc} = 0$ for some $\alpha \neq_{0} 0$ or, equivalently, by (2.21), $\overset{\alpha}{\Gamma}_{jkc_1} = 0$. Since (i) and (ii)′ [or (ii)] implies that $\Gamma_{jkc} = 0$, it follows from (2.8) that $\Gamma_{jkc} = 0$ which, because of (ii)′, is equivalent to (iii). □

COROLLARY 4.2. *Under ordinary repeated sampling from an orthogeodesic model with* (χ, ψ) *as ortho-affine parameter, not only are the score components* $l_{/a}$ *and* $l_{/j}$ *asymptotically independent to error* $O(n^{-1/2})$ [*as follows immediately from condition* (*i*)] *but we have the stronger result that to order* $O(n^{-1})$ *the score component* $l_{/a}$ *is asymptotically independent of*

$$l_{/j} - \tfrac{1}{2} T_{jab} i^{ac} i^{bd} (l_{/c} l_{/d} - i_{cd}). \tag{4.8}$$

PROOF. This follows from a general result concerning asymptotic independence, given in Barndorff-Nielsen and Blæsild (1992). □

EXAMPLE 4.1. As discussed in Example 3.1, the univariate normal distribution $N(\mu, \sigma^2)$ determines an orthogeodesic model, the ortho-affine coordinates being $\chi = \mu$, $\psi = \sigma^{-2}$. The score components $l_{/\mu}$ and $l_{/\sigma^{-2}}$ are orthogonal but, unlike $\hat{\mu}$ and $\hat{\sigma}^2$, not independent. However, in this case (4.8) turns out to be not only asymptotically independent of $l_{/\mu}$ to order $O(n^{-1})$ but in fact independent of $l_{/\mu}$.

It might be guessed that a similar complete independence result would hold for the inverse Gaussian distribution, but that is not the case.

For details, see Barndorff-Nielsen and Blæsild (1992).

5. Exponential orthogeodesic models. Throughout this section we consider a d-dimensional exponential model $\mathscr{M}$ with model function of the form

$$\exp\{\theta^\rho t_\rho(x) - \kappa(\theta) - \varphi(x)\}. \tag{5.1}$$

We assume that the exponential model is steep [in the terminology of Barndorff-Nielsen (1978)] and let Θ denote the canonical parameter domain. As in (5.1) we use the letters $\rho, \sigma, \dots$ to indicate arbitrary components of the canonical parameter θ as well as of the canonical statistic t, and we restrict the parameter domain of $\mathscr{M}$ to int Θ, the interior of Θ. The mean value mapping defined on int Θ will, as previously, be denoted by τ, that is, $\tau_\rho(\theta) = E_\theta\{t_\rho\}$.

For an exponential model we first note that condition (iv) in Definition 3.1, or, equivalently, condition (iv)′ in Definition 3.1′, is redundant. An exponential model is 1-flat (in the canonical parameter) so $\overset{1}{R} \equiv 0$. Thus if the submanifold $\mathscr{M}_\chi$ is 1-geodesic [conditions (iii) and (iii)′], that is, if the components $\overset{1}{H}_{jkc}$ of the 1-embedding curvature of $\mathscr{M}_\chi$ all vanish, it follows from (2.27) (with $\alpha = 1$) that ${}_\chi\overset{1}{R} \equiv 0$.

With the notation introduced previously we have the following theorem concerning the structure of an exponential orthogeodesic model.

THEOREM 5.1. *Let* $\omega = (\chi, \psi)$ *be a parametrization of the exponential model* (5.1) *such that* χ *and* ψ *are variation independent. Then the model* (5.1) *is orthogeodesic with* ω *as ortho-affine parameter if and only if there exist* *scalars* $\alpha(\psi)$ *and* $\gamma(\chi)$, *vectors* $B_\rho(\chi)$ *and* $D^\rho(\chi)$ *and matrices* $A^i_\rho(\chi)$ *and* $C^\rho_i(\chi)$ *satisfying the conditions*

$$A^i_\rho(\chi)C^\rho_j(\chi) = \delta^i_j, \tag{5.2}$$

$$A^i_\rho(\chi)C^\rho_{j/a}(\chi) = 0, \tag{5.3}$$

$$A^i_\rho(\chi)D^\rho_{/a}(\chi) = 0, \tag{5.4}$$

$$B_{\rho/a}(\chi)C^\rho_i(\chi) = 0 \tag{5.5}$$

and

$$\gamma_{/a}(\chi) = B_\rho(\chi)D^\rho_{/a}(\chi), \tag{5.6}$$

such that

$$\theta^\rho(\chi,\psi) = \psi^i C^\rho_i(\chi) + D^\rho(\chi), \tag{5.7}$$

$$\tau_\rho(\chi,\psi) = \alpha_{/j}(\psi)A^j_\rho(\chi) + B_\rho(\chi) \tag{5.8}$$

and

$$\kappa(\chi,\psi) = \alpha(\psi) + \psi^i C^\rho_i(\chi)B_\rho(\chi) + \gamma(\chi). \tag{5.9}$$

PROOF. Suppose that the model (5.1) is orthogeodesic with ω as orthoaffine parameter. Then (iii) and (iv) in Definition 3.1 imply that

$$0 = \overset{1}{\Gamma}{}^s_{jk} = \left\{\overset{1}{\Gamma}{}^\sigma_{\tau\nu}\theta^\tau_{/j}\theta^\nu_{/k} + \theta^\sigma_{/jk}\right\}\omega^s_{/\sigma}.$$

Since an exponential model is 1-flat in the canonical parameter θ, that is, $\overset{1}{\Gamma}{}^\sigma_{\tau\nu} = 0$, it follows that

$$0 = \theta^\sigma_{/jk}\omega^s_{/\sigma}$$

or, equivalently, that

$$0 = \theta^\sigma_{/jk}\omega^s_{/\sigma}\theta^\rho_{/s} = \theta^\sigma_{/jk}\delta^\rho_\sigma = \theta^\rho_{/jk}$$

from which we obtain that there exist a vector $D^\rho(\chi)$ and a matrix $C^\rho_i(\chi)$ such that (5.7) is fulfilled.

Inserting $\theta^\rho_{/jk} = 0$ into

$$\kappa_{/jk} = \kappa_{/\rho\sigma}\theta^\rho_{/j}\theta^\sigma_{/k} + \kappa_{/\rho}\theta^\rho_{/jk},$$

we find, using (ii) in Definition (3.1), that

$$\begin{aligned}\kappa_{/jk} &= \kappa_{/\rho\sigma}\theta^\rho_{/j}\theta^\sigma_{/k}\\ &= i_{\rho\sigma}(\theta)\theta^\rho_{/j}\theta^\sigma_{/k}\\ &= i_{jk}(\chi,\psi)\\ &= i_{jk}(\psi).\end{aligned} \tag{5.10}$$

Consequently, there exists a scalar $\alpha(\psi)$ satisfying

$$\alpha_{/jk}(\psi) = i_{jk}(\psi), \tag{5.11}$$

and a scalar γ and a vector β_i both depending on χ only such that

$$\kappa(\chi, \psi) = \alpha(\psi) + \psi^i \beta_i(\chi) + \gamma(\chi). \tag{5.12}$$

Formula (5.7) implies that $\theta^\rho_{/j} = C^\rho_j(\chi)$ so from the calculations in (5.10) one gets

$$i_{jk}(\psi) = i_{\rho\sigma}(\theta) C^\rho_j(\chi) C^\sigma_k(\chi). \tag{5.13}$$

Let $A^i_\rho(\chi)$ denote the matrix

$$A^i_\rho(\chi) = i_{\rho\sigma}(\theta) C^\sigma_k(\chi) i^{ik}(\psi) \tag{5.14}$$

and observe, using (5.13), that

$$\begin{aligned} A^i_\rho(\chi) C^\rho_j(\chi) &= i_{\rho\sigma}(\theta) C^\sigma_k(\chi) i^{ik}(\psi) C^\rho_j(\chi) \\ &= i_{jk}(\psi) i^{ik}(\psi) \\ &= \delta^i_j, \end{aligned}$$

which is (5.2). Applying the formulas (5.11) and (5.14), we find that

$$\begin{aligned} \tau_{\rho/k} &= \tau_{\rho/\sigma} \theta^\sigma_{/k} \\ &= \tau_{\rho/\sigma} C^\sigma_k(\chi) \\ &= i_{\rho\sigma}(\theta) C^\sigma_i(\chi) i^{ij}(\psi) i_{jk}(\psi) \\ &= \alpha_{/jk}(\psi) A^j_\rho(\chi) \end{aligned}$$

and it follows that there exists a vector $B_\rho(\chi)$ such that (5.8) is fulfilled.

It remains to prove (5.3)–(5.6) and (5.9). Inserting (5.7) and (5.12) into (5.1) the log-likelihood function becomes

$$l(\chi, \psi) = -\alpha(\psi) - \psi^i \beta_i(\chi) - \gamma(\chi) + \big(\psi^i C^\rho_i(\chi) + D^\rho(\chi)\big) t_\rho(\chi). \tag{5.15}$$

From (5.15) it is easily seen that the conditions

$$E_\theta\{-l_{/aj}\} = i_{aj} = 0,$$
$$E_\theta\{l_{/a}\} = 0$$

and

$$E_\theta\{l_{/j}\} = 0$$

imply

$$\beta_{j/a}(\chi) = C^\rho_{j/a}(\chi) \tau_\rho(\chi, \psi), \tag{5.16}$$

$$\gamma_{/a}(\chi) = D^\rho_{/a}(\chi) \tau_\rho(\chi, \psi) \tag{5.17}$$

and

$$\alpha_{/j}(\psi) + \beta_j(\chi) = C^\rho_j(\chi) \tau_\rho(\chi, \psi). \tag{5.18}$$

Differentiating (5.18) with respect to χ^a, one obtains, using (5.16), that

$$C^\rho_j(\chi) \tau_\rho(\chi, \psi)_{/a} = 0. \tag{5.19}$$

Inserting (5.8) into (5.16) and (5.17), respectively, we find

$$\beta_{j/a}(\chi) = C^{\rho}_{j/a}(\chi)\big(\alpha_{/i}(\psi)A^i_\rho(\chi) + B_\rho(\chi)\big) \tag{5.20}$$

and

$$\gamma_{/a}(\chi) = D^{\rho}_{/a}(\chi)\big(\alpha_{/i}(\psi)A^i_\rho(\chi) + B_\rho(\chi)\big). \tag{5.21}$$

Hence, using that χ and ψ are variation independent, one has for $\psi_0 \neq \psi$ that

$$0 = C^{\rho}_{j/a}(\chi)A^i_\rho(\chi)\big(\alpha_{/i}(\psi) - \alpha_{/i}(\psi_0)\big) \tag{5.22}$$

and

$$0 = D^{\rho}_{/a}(\chi)A^i_\rho(\chi)\big(\alpha_{/i}(\psi) - \alpha_{/i}(\psi_0)\big). \tag{5.23}$$

From (5.2) it is seen that the rank of the matrix $A^i_\rho(\chi)$ is d_ψ, the dimension of the subparameter ψ, and hence it follows from (5.8) that there exist $\psi_1, \dots, \psi_{d_\psi}$ such that $\{\alpha_{/i}(\psi_k) - \alpha_{/i}(\psi_0)\colon k = 1, \dots, d_\psi\}$ is a set of linearly independent vectors. The formulas (5.22) and (5.23) now imply that (5.3) and (5.4) are fulfilled and formula (5.6) follows from (5.4) and (5.21).

From (5.2), (5.3) and (5.20) we find that

$$C^{\rho}_j(\chi)A^j_{\rho/a}(\chi) = 0 \tag{5.24}$$

and

$$\beta_{j/a}(\chi) = C^{\rho}_{j/a}(\chi)B_\rho(\chi). \tag{5.25}$$

Formulas (5.8), (5.19) and (5.24) now imply (5.5). Using (5.5) and (5.25), it follows that there exists a constant k_i such that

$$\beta_i(\chi) = C^\rho_i(\chi)B_\rho(\chi) + k_i.$$

Rewriting (5.12), we get

$$\begin{aligned}\kappa(\chi,\psi) &= \alpha(\psi) + \psi^i\big(C^\rho_i(\chi)B_\rho(\chi) + k_i\big) + \gamma(\chi)\\ &= \big(\alpha(\psi) + \psi^i k_i\big) + \psi^i C^\rho_i(\chi)B_\rho(\chi) + \gamma(\chi)\\ &= \tilde{\alpha}(\psi) + \psi^i C^\rho_i(\chi)B_\rho(\chi) + \gamma(\chi),\end{aligned}$$

which is (5.9) and the proof of the necessity of the conditions (5.2)–(5.9) is complete.

Conversely, suppose that the conditions (5.2)–(5.9) are fulfilled. Using (5.1), (5.7) and (5.9), the log-likelihood function becomes

$$l(\chi,\psi) = -\alpha(\psi) - \psi^i C^\rho_i(\chi)B_\rho(\chi) - \gamma(\chi) + \big(\psi^i C^\rho_i(\chi) + D^\rho(\chi)\big)t_\rho(x)$$

and, using (5.2)–(5.6) and (5.8), it is easily seen that $E_\theta\{l_{/a}\} = 0 = E_\theta\{l_{/j}\}$.

Furthermore, one has

$$\begin{aligned}i_{bj}(\chi,\psi) &= E_\theta\{-l_{/bj}\}\\ &= \big(C^\rho_j(\chi)B_\rho(\chi)\big)_{/b} - C^\rho_{j/b}(\chi)\tau_\rho(\chi,\psi)\end{aligned}$$

and, using (5.3), (5.5) and (5.8), it is easily established that $i_{bj}(\chi,\psi) = 0$, that is, that condition (i) in Definition 3.1 is fulfilled. Condition (ii) is a consequence of the fact that $l_{/jk} = -\alpha_{/jk}(\psi)$. Finally, from (5.7) it follows that

$$\theta^{\rho}_{/jk} = 0$$

and since the model is 1-flat in θ this implies that conditions (iii) and (iv) in Definition 3.1 are fulfilled because

$$\begin{aligned} \overset{1}{\Gamma}{}^{r}_{jk} &= \left\{ \overset{1}{\Gamma}{}^{\sigma}_{\sigma\nu} \theta^{\rho}_{/j} \theta^{\nu}_{/k} + \theta^{\rho}_{/jk} \right\} \omega^{r}_{/\rho} \\ &= 0. \end{aligned}$$

□

COROLLARY 5.1. *If the model* (5.1) *is orthogeodesic, then the quantity* $P(x; \chi)$ *with components*

$$P_i(x; \chi) = C_i^{\rho}(\chi)\big(t_{\rho}(x) - B_{\rho}(\chi)\big)$$

has Laplace transform

$$E_{\theta}\{\exp(\zeta^i P_i)\} = \exp(\alpha(\psi + \zeta) - \alpha(\psi)). \tag{5.26}$$

Consequently, the distribution of P *depends on* ψ *only, that is,* P *is a pivot provided* ψ *is known.*

PROOF. Using (5.7) and (5.9), the density (5.1) may be rewritten as

$$\begin{aligned} &\exp\left\{-\alpha(\psi) - \psi^i C_i^{\rho}(\chi) B_{\rho}(\chi) - \gamma(\chi) + \big(\psi^i C_i^{\rho}(\chi) + D^{\rho}(\chi)\big) t_{\rho}(x)\right\} \\ &\quad = \exp\left\{-\alpha(\psi) - \gamma(\chi) + D^{\rho}(\chi) t_{\rho}(x) + \psi^i P_i(x; \chi)\right\} \end{aligned}$$

from which (5.26) follows. □

In conclusion it may be noted that, denoting the mean value of $P = \{P_i\}$ by $\pi = \{\pi_i\}$, we have, as may be shown by means of the results of this section, that for exponential models, as considered here, conditions (iii) and (iv) of Definition 3.1 are jointly equivalent to $\overset{-1}{\Gamma}{}^{\gamma}_{ij}(\chi, \pi) = 0$.

Acknowledgment. We are much indebted to an anonymous referee of this paper for detailed and constructive comments.

REFERENCES

AMARI, S.-I. (1980). Theory of information spaces—a differential geometrical foundation of statistics. POST RAAG Report 106.

AMARI, S.-I. (1982). Differential geometry of curved exponential families—curvature and information loss. *Ann. Statist.* **10** 357–387.

AMARI, S.-I. (1985). *Differential-Geometric Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, Berlin.

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

BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, Chichester.

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

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

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

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

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1992). A type of second order asymptotic independence. *J. Roy. Statist. Soc. Ser. B.* **54** 897–901.

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. and JØRGENSEN, B. (1991). Proper dispersion models. Research Report 200, Dept. Theoretical Statistics, Univ. Aarhus.

BLÆSILD, P. and JENSEN, J. L. (1985). Saddlepoint formulas for reproductive exponential families. *Scand. J. Statist.* **12** 193–202.

CHENTSOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference*. Nauka, Moscow. (In Russian). English translation (1982). *Transl. Math. Monographs* **53**. Amer. Math. Soc., Providence, RI.

GRADSHTEYN, I. S. and RYZHIK, I. M. (1965). *Table of Integrals, Series and Products*, 4th ed. Academic, New York.

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

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

MATEMATISK INSTITUT
AARHUS UNIVERSITET
DK-8000 AARHUS C
DENMARK