Ann. Inst. Statist. Math.
Vol. 46, No. 1, 43–55 (1994)

MAXIMUM LIKELIHOOD ESTIMATION IN EXPONENTIAL ORTHOGEODESIC MODELS

PREBEN BLÆSILD

Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, Ny Munkegade, DK-8000, Aarhus C, Denmark

(Received September 1, 1992; revised May 24, 1993)

Abstract. An orthogeodesic statistical model is defined in terms of five conditions of differential geometric nature. These conditions are reviewed together with a characterization theorem for exponential orthogeodesic models. Orthogonal projections, relevant for maximum likelihood estimation in exponential orthogeodesic models, are described in a simple way in terms of some of the quantities in the characterization theorem. A unified procedure for performing maximum likelihood estimation in exponential orthogeodesic models is given and the use of this procedure is illustrated for some of the most important models of this kind such as θ-parallel models, τ-parallel models and certain transformation models.

Key words and phrases: Affine α-connections, expected information, flat submanifolds, geodesic submanifolds, likelihood equations, orthogonal projections, pivot, transformation models, τ-parallel models, θ-parallel models.

1. Introduction

The concept of an orthogeodesic statistical model was introduced in Barndorff-Nielsen and Blæsild (1993). A statistical model M is orthogeodesic if it satisfies the following five conditions of differential geometric nature:

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

(ii) The factorization of M is orthogonal with respect to the expected (Fisher) information metric i on M.

(iii) For every value of $\chi \in X$ the restriction of the metric i to the submanifold $M_\chi = \{(\chi, \psi) : \psi \in \Psi\}$ does not depend on χ.

(iv) For every value of $\chi \in X$ and for some value $\alpha \neq 0$ the submanifold M_χ is expected α-geodesic, i.e. the shape tensor ${}_\chi\overset{\alpha}{H}$ of M_χ corresponding to the expected α-connection vanishes identically.

(v) For every value of $\chi \in X$ the submanifold M_χ is expected 1-flat, i.e.

the Riemannian curvature tensor $\overset{1}{{}_\chi R}$ of M_χ corresponding to the expected 1-connection vanishes identically.

A discussion of these conditions and of a set of equivalent conditions, formulated in terms of local coordinates on M, may be found in Barndorff-Nielsen and Blæsild (1993) in which also various examples of orthogeodesic statistical models are given. Apart from the one-dimensional location-scale model all the examples are concerned with exponential models such as τ- or θ-parallel models, certain exponential transformation models and proper exponential dispersion models and generalizations hereof. Furthermore, Barndorff-Nielsen and Blæsild (1993) give a complete characterization of the structure of exponential orthogeodesic models, reviewed in Section 2 below, but they do not discuss the implications of this structure on the statistical inference of such models. A first step in that direction is taken in the present paper which discusses maximum likelihood estimation in exponential orthogeodesic models. The discussion is based on some relevant orthogonal projections which are described in a simple way in terms of some of the quantities in the characterization theorem. Despite the fact that the statistical properties of the exponential models mentioned above are very different, the maximum likelihood estimation in these models follows a certain scheme. To be more specific, let $l(\chi, \psi)$ denote the log likelihood function for the orthogeodesic parameter (χ, ψ) and let $\hat{\chi}_\psi$ denote the maximum likelihood estimate of χ for fixed value of ψ. Under mild regularity conditions it is shown that $\hat{\chi}_\psi = \hat{\chi}$, where $\hat{\chi}$ denotes the maximum likelihood estimate of χ in the full model. Consequently, the maximum likelihood estimate $\hat{\psi}$ of ψ may be found by maximizing $l(\hat{\chi}, \psi)$. Thus maximum likelihood estimation in an exponential orthogeodesic model may be considered as consisting of two steps. The situation is similar to that of estimating the two parameters μ and σ^2 of the univariate normal distribution $N(\mu, \sigma^2)$ based on a sample $x_1, \ldots, x_n$. Here one has

$$\hat{\mu} = \hat{\mu}_{\sigma^2} = \bar{x} = (x_1 + \cdots + x_n)/n$$

and $\hat{\sigma}^2$ is found by maximizing

$$l(\bar{x}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_i (x_i - \bar{x})^2.$$

The normal model is considered as an exponential orthogeodesic model in Example 5.2.

Besides the characterization theorem of exponential orthogeodesic models, Section 2 introduces the necessary notation. Section 3 contains some further results concerning some of the quantities in the characterization theorem, which are of use in the discussion, in Section 4, of maximum likelihood estimation in exponential orthogeodesic models. Finally, the results are illustrated in Section 5.

2. Exponential orthogeodesic models

Throughout the paper we consider a d-dimensional exponential model M with model function of the form

$$\exp\{\theta^\rho t_\rho(x) - \kappa(\theta) - \phi(x)\}. \tag{2.1}$$

We assume that the exponential model is steep (in the terminology of Barndorff-Nielsen (1978)) and we let Θ denote the canonical parameter domain. Throughout the paper we use the summation convention and as in (2.1) we use the letters $\rho, \sigma, \ldots$ to indicate arbitrary components of the canonical parameter $\theta = (\theta^1, \ldots, \theta^d)$ as well as of the canonical statistic $t = (t_1, \ldots, t_d)$. Furthermore, we restrict the parameter domain of M to $\operatorname{int}\Theta$, the interior of Θ, i.e. we assume that M is a core exponential model (in the terminology of Barndorff-Nielsen (1988)). The mean value mapping defined on $\operatorname{int}\Theta$ will be denoted by τ, i.e. $\tau_\rho(\theta) = E_\theta\{t_\rho\}$.

We use $\omega = (\omega^1, \ldots, \omega^d)$ as an alternative parameter of the exponential model (2.1), i.e. $M = \{P_\omega : \omega \in \Omega\}$, where P_ω is the probability measure corresponding to ω and where the parameter space Ω is an open subset of R^d. Generic components of ω are indicated by $\omega^r, \omega^s, \omega^t, \ldots$. Moreover, we let $\{\partial_r\}$ denote the coordinate frame $\{\partial/\partial\omega^r\}$ at ω and for an arbitrary real-valued function f defined on Ω we write $f_{/r}(\omega) = \partial_r f(\omega)$, $f_{/rs}(\omega) = \partial_r\partial_s f(\omega)$, etc. Furthermore, we use $\{i_{rs}(\omega)\}$ to denote the expected information $i(\omega)$ and $\{i^{rs}(\omega)\}$ to denote the inverse $i^{-1}(\omega)$.

As indicated in the introduction we are particularly interested in the situation where there exists a parametrization of the model 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 $\Omega = X \times \Psi$. We use the suffices $a, b, c, \ldots$ and $i, j, k, \ldots$ to denote generic components of χ and ψ, respectively. For fixed $\chi \in X$ we use M_χ to denote the submodel $\{P_{(\chi,\psi)} : \psi \in \Psi\}$ and similarly we let $M_\psi = \{P_{(\chi,\psi)} : \chi \in X\}$ for $\psi \in \Psi$.

With this notation we have the following theorem concerning the structure of an exponential orthogeodesic model, the proof of which may be found in Barndorff-Nielsen and Blæsild (1993).

THEOREM 2.1. *Let $\omega = (\chi, \psi)$ be a parametrization of the exponential model* (2.1) *such that χ and ψ are variation independent. Then the model* (2.1) *is orthogeodesic relative to ω 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*

$$A^i_\rho(\chi)C^\rho_j(\chi) = \delta^i_j, \tag{2.2}$$

$$A^i_\rho(\chi)C^\rho_{j/a}(\chi) = 0, \tag{2.3}$$

$$A^i_\rho(\chi)D^\rho_{/a}(\chi) = 0, \tag{2.4}$$

$$B_{\rho/a}(\chi)C^\rho_i(\chi) = 0 \tag{2.5}$$

and

$$\gamma_{/a}(\chi) = B_\rho(\chi)D^\rho_{/a}(\chi), \tag{2.6}$$

such that

$$\theta^\rho(\chi,\psi) = \psi^i C_i^\rho(\chi) + D^\rho(\chi), \tag{2.7}$$

$$\tau_\rho(\chi,\psi) = \alpha_{/j}(\psi) A_\rho^j(\chi) + B_\rho(\chi) \tag{2.8}$$

and

$$\kappa(\chi,\psi) = \alpha(\psi) + \psi^i C_i^\rho(\chi) B_\rho(\chi) + \gamma(\chi). \tag{2.9}$$

The formulas (2.1), (2.7) and (2.9) imply that, disregarding a quantity which depends on x only, the log likelihood function in terms of (χ,ψ) is

$$l(\chi,\psi;x) = -\alpha(\psi) - \gamma(\chi) + \psi^i C_i^\rho(\chi)\{t_\rho(x) - B_\rho(\chi)\} + D^\rho(\chi) t_\rho(x) \tag{2.10}$$

from which it follows that

$$i_{jk}(\chi,\psi) = i_{jk}(\psi) = \alpha_{/jk}(\psi). \tag{2.11}$$

Furthermore, Barndorff-Nielsen and Blæsild (1993) proved the following

COROLLARY 2.1. *The quantity $P(x;\chi)$ with components*

$$P_i(x;\chi) = C_i^\rho(\chi)\{t_\rho(x) - B_\rho(\chi)\} \tag{2.12}$$

has Laplace transform

$$E_\theta\{\exp(\zeta^i P_i)\} = \exp(\alpha(\psi+\zeta) - \alpha(\psi)). \tag{2.13}$$

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

Note that the quantity P depends on x through $t(x)$ only, i.e. P may be considered as a function of $(t;\chi)$, which we will do from now on. Similarly, since t is minimal sufficient, the log likelihood function may be considered as a function of t and expressed in terms of P we have

$$l(\chi,\psi;t) = -\alpha(\psi) - \gamma(\chi) + \psi^i P_i(t;\chi) + D^\rho(\chi) t_\rho. \tag{2.14}$$

3. Further results

This section contains some further results about some of the quantities in Theorem 2.1. Some of the results are used in the discussion of maximum likelihood estimation in exponential orthogeodesic models in Section 4.

LEMMA 3.1. *The derivatives of $\omega = (\chi,\psi)$ with respect to the canonical parameter θ are given by*

$$\omega^r_{/\rho} = \tau_{\rho/t} i^{rt}. \tag{3.1}$$

In particular one has

$$\chi^a_{/\rho} = (\alpha_{/j}(\psi)A^j_{\rho/b}(\chi) + B_{\rho/b}(\chi))i^{ab} \tag{3.2}$$

and

$$\psi^i_{/\rho} = A^i_\rho(\chi). \tag{3.3}$$

PROOF. Multiplying the identity

$$i_{rs}\omega^r_{/\rho}\omega^s_{/\sigma} = i_{\rho\sigma}$$

by $\theta^\sigma_{/t}$ we obtain

$$i_{rt}\omega^r_{/\rho} = i_{\rho\sigma}\theta^\sigma_{/t} = \tau_{\rho/\sigma}\theta^\sigma_{/t} = \tau_{\rho/t}$$

from which (3.1) follows. Since the parameters χ and ψ are i-orthogonal one obtains that $i^{aj} = 0$. Using this fact together with the formulas (2.8) and (2.11) it is easily seen that (3.2) and (3.3) are consequences of (3.1). □

COROLLARY 3.1. *Expressed in terms of the parameter* (χ, ψ) *the expected information* $\{i_{\rho\sigma}\}$ *for the canonical parameter* θ (*the covariance matrix of* t) *and its inverse* $\{i^{\rho\sigma}\}$ *are, respectively,*

$$\begin{aligned} i_{\rho\sigma} = {} & (\alpha_{/j}(\psi)A^j_{\rho/a}(\chi) + B_{\rho/a}(\chi))(\alpha_{/k}(\psi)A^k_{\sigma/b}(\chi) + B_{\sigma/b}(\chi))i^{ab} \\ & + i_{jk}A^j_\rho(\chi)A^k_\sigma(\chi) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} i^{\rho\sigma} = {} & (\psi^j C^\rho_{j/a}(\chi) + D^\rho_{/a}(\chi))(\psi^k C^\sigma_{k/b}(\chi) + D^\sigma_{/b}(\chi))i^{ab} \\ & + C^\rho_j(\chi)C^\sigma_k(\chi)i^{jk}. \end{aligned} \tag{3.5}$$

PROOF. Formula (3.4) is a consequence of (3.2) and (3.3) and of the formula

$$\begin{aligned} i_{\rho\sigma} &= i_{rs}\omega^r_{/\rho}\omega^s_{/\sigma} \\ &= i_{ab}\chi^a_{/\rho}\chi^b_{/\sigma} + i_{jk}\psi^j_{/\rho}\psi^k_{/\sigma}. \end{aligned}$$

Formula (3.5) follows, using (2.7), from the formula

$$\begin{aligned} i^{\rho\sigma} &= i^{rs}\theta^\rho_{/r}\theta^\sigma_{/s} \\ &= i^{ab}\theta^\rho_{/a}\theta^\sigma_{/b} + i^{jk}\theta^\rho_{/j}\theta^\sigma_{/k}. \end{aligned}$$

□

LEMMA 3.2. *Let* m *be the point in* M *corresponding to* (χ, ψ) *and let* p_χ *denote the orthogonal projection* (*with respect to the information metric* i) *on* T_mM_χ, *the tangent space at* m *to the submanifold* M_χ. *Similarly, let* p_ψ *denote the orthogonal projection on* T_mM_χ. *The matrices corresponding to these linear mappings are, respectively,*

$$(p_\chi)^\sigma_\rho = A^k_\rho(\chi)C^\sigma_k(\chi) \tag{3.6}$$

and

$$(p_\psi)^\sigma_\rho = \delta^\sigma_\rho - A^k_\rho(\chi)C^\sigma_k(\chi) \\ = \{\psi^i C^\sigma_{i/a}(\chi) + D^\sigma_{/a}(\chi)\}\chi^a_{/\rho}. \tag{3.7}$$

PROOF. To prove (3.6) we use the following well-known result: If L is the q-dimensional linear subspace of R^d spanned by the rows in the $q \times d$ matrix Δ then the orthogonal projection on L with respect to a positive definite matrix Σ has the matrix

$$p_L = \Sigma\Delta^T(\Delta\Sigma\Delta^T)^{-1}\Delta, \tag{3.8}$$

where Δ^T denotes the transpose of Δ.

Considering the local coordinates corresponding to the canonical parameter θ we have that $\Sigma_{\rho\sigma} = i_{\rho\sigma}$ and $L = T_m M_\chi = \text{span}\{\partial_k\}$ where, using (2.7),

$$\partial_k = \theta^\rho_{/k}\partial_\rho = C^\rho_k(\chi)\partial_\rho.$$

Consequently, we may use (3.8) with Δ as the matrix given by

$$\Delta^\rho_j = C^\rho_j(\chi).$$

From (3.4) we obtain that

$$\begin{aligned}(\Delta\Sigma\Delta^T)_{jk} &= \Delta^\rho_j\Sigma_{\rho\sigma}\Delta^\sigma_k \\ &= C^\rho_j(\chi)\{(\alpha_{/l}(\psi)A^l_{\rho/a}(\chi) + B_{\rho/a}(\chi))(\alpha_{/n}(\chi)A^n_{\sigma/b}(\chi) + B_{\sigma/b}(\chi))i^{ab} \\ &\quad + i_{ln}A^l_\rho(\chi)A^n_\sigma(\chi)\}C^\sigma_k(\chi).\end{aligned}$$

The formulas (2.2) and (2.3) imply that

$$A^i_{\rho/a}(\chi)C^\rho_j(\chi) = 0 \tag{3.9}$$

and from (2.2), (2.5) and (3.9) we find

$$\begin{aligned}(\Delta\Sigma\Delta^T)_{jk} &= i_{ln}C^\rho_j(\chi)A^l_\rho(\chi)A^n_\sigma(\chi)C^\sigma_k(\chi) \\ &= i_{ln}\delta^l_j\delta^n_k \\ &= i_{jk}.\end{aligned}$$

Consequently, using (3.4), (3.9), (2.2) and (2.3), we get

$$\begin{aligned}(p_\chi)^\sigma_\rho &= \Sigma_{\rho\upsilon}\Delta^\upsilon_j i^{jk}\Delta^\sigma_k \\ &= \{(\alpha_{/l}(\psi)A^l_{\rho/a}(\chi) + B_{\rho/a}(\chi))(\alpha_{/n}(\psi)A^n_{\upsilon/b}(\chi) + B_{\upsilon/b}(\chi))i^{ab} \\ &\quad + i_{ln}A^l_\rho(\chi)A^n_\upsilon(\chi)\}C^\upsilon_j(\chi)i^{jk}C^\sigma_k(\chi) \\ &= i_{ln}A^l_\rho(\chi)A^n_\upsilon(\chi)C^\upsilon_j(\chi)i^{jk}C^\sigma_k(\chi) \\ &= i_{ln}A^l_\rho(\chi)\delta^n_j i^{jk}C^\sigma_k(\chi) \\ &= i_{lj}A^l_\rho(\chi)i^{jk}C^\sigma_k(\chi) \\ &= \delta^k_l A^l_\rho(\chi)C^\sigma_k(\chi) \\ &= A^k_\rho(\chi)C^\rho_k(\chi)\end{aligned}$$

and the proof of (3.6) is complete.

Since the tangent spaces $T_m M_\chi$ and $T_m M_\psi$ are i-orthogonal the first equality in (3.7) is obvious. To prove the second equality in (3.7) note that (2.7) and (3.3) imply

$$\begin{aligned}\delta^\sigma_\rho = \theta^\sigma_{/\rho} &= \theta^\sigma_{/a}\chi^a_{/\rho} + \theta^\sigma_{/k}\psi^k_{/\rho}\\ &= \{\psi^i C^\sigma_{i/a}(\chi) + D^\sigma_{/a}(\chi)\}\chi^a_{/\rho} + C^\sigma_k(\chi)A^k_\rho(\chi)\end{aligned}$$

which was to be proved. □

4. Maximum likelihood estimation

The maximum likelihood estimate $(\hat\chi, \hat\psi)$ of (χ, ψ) may be found directly from the log likelihood function (2.14). Here, however, we derive the formulas for $(\hat\chi, \hat\psi)$ by a geometric argument in line with the nature of the model.

THEOREM 4.1. *The maximum likelihood estimate* $(\hat\chi, \hat\psi)$ *of* (χ, ψ) *is the unique solution to the equations*

$$\{\delta^\sigma_\rho - A^k_\rho(\chi)C^\sigma_k(\chi)\}\{t_\sigma - B_\sigma(\chi)\} = 0 \tag{4.1}$$

and

$$C^\sigma_j(\chi)\{t_\sigma - B_\sigma(\chi)\} = \alpha_{/j}(\psi). \tag{4.2}$$

PROOF. The maximum likelihood estimate $\hat m$ of m is the unique point in M for which the differential of the log likelihood function vanishes, i.e.

$$d_{\hat m} l = 0.$$

Expressed in terms of the canonical parameter θ the differential of l is

$$\begin{aligned}d_m l &= l_{/\rho} d\theta^\rho\\ &= (t_\rho - \tau_\rho)d\theta^\rho.\end{aligned}$$

Considering the score vector $t - \tau$ as a tangent vector we have to find a point $\hat m = m(\hat\theta)$ such that

$$0 = t - \hat\tau \in T_{\hat m} M. \tag{4.3}$$

Clearly, with the notation in Lemma 3.2, formula (4.3) is equivalent to the formulas

$$p_\chi(t - \tau) = 0 \tag{4.4}$$

and

$$p_\psi(t - \tau) = 0 \tag{4.5}$$

or, equivalently, using (3.6), (3.7) and (2.8), to the equations

$$A^k_\rho(\chi)C^\sigma_k(\chi)\{t_\sigma - \alpha_{/j}(\psi)A^j_\sigma(\chi) - B_\sigma(\chi)\} = 0 \tag{4.6}$$

and

$$\{\psi^i C^\sigma_{i/a}(\chi) + D^\sigma_{/a}(\chi)\}\chi^a_{/\rho}\{t_\sigma - \alpha_{/j}(\psi)A^j_\sigma(\chi) - B_\sigma(\chi)\} = 0. \tag{4.7}$$

Using (2.2) formula (4.6) is seen to be equivalent to

$$A^k_\rho(\chi)C^\sigma_k(\chi)\{t_\sigma - B_\sigma(\chi)\} - A^k_\rho(\chi)\alpha_{/k}(\psi) = 0.$$

Multiplying by $C^\rho_l(\chi)$ and using (2.2) the equivalence between this formula and formula (4.2) is easily established.

Applying (2.3) and (2.4) formula (4.7) becomes

$$\{\psi^i C^\sigma_{i/a}(\chi) + D^\sigma_{/a}(\chi)\}\chi^a_{/\rho}\{t_\sigma - B_\sigma(\chi)\} = 0$$

which, using (3.7), is seen to be equivalent to (4.1). □

Let π denote the mean value of the quantity P defined in (2.12). Using (2.2) and (2.7) we find

$$\begin{aligned}\pi_i(\chi,\psi) &= E_{(\chi,\psi)}P_i(t;\chi) \\ &= C^\rho_i(\chi)\{\alpha_{/j}(\psi)A^j_\rho(\chi) + B_\rho(\chi) - B_\rho(\chi)\} \\ &= \delta^j_i\alpha_{/j}(\psi) \\ &= \alpha_{/j}(\psi).\end{aligned} \tag{4.8}$$

Thus π depends on (χ,ψ) through ψ only and (2.8) implies that the mapping $\psi \to \{\alpha_{/i}(\psi)\} = \pi(\psi)$ is in fact one-to-one. Denoting the inverse mapping by ψ, i.e. $\psi = \psi(\pi)$ we have the following.

COROLLARY 4.1. *Let $(\hat\chi,\hat\psi)$ denote the maximum likelihood estimate of (χ, ψ). Then $\hat\chi$ is the unique solution to the equations*

$$(\delta^\sigma_\rho - A^k_\rho(\chi)C^\sigma_k(\chi))(t_\sigma - B_\sigma(\chi)) = 0 \tag{4.9}$$

and $\hat\psi$ is given by

$$\hat\psi = \psi(P(t;\hat\chi)). \tag{4.10}$$

COROLLARY 4.2. *Suppose for every value of ψ that the maximum likelihood estimate $\hat\chi_\psi$ of χ in the submodel $M_\psi = \{P_{(\chi,\psi)} : \chi \in X\}$ is a unique solution to the likelihood equations. Then $\hat\chi_\psi = \hat\chi$, where $\hat\chi$ is the maximum likelihood estimate of χ in the full model, i.e. $\hat\chi_\psi$ does not depend on ψ.*

PROOF. From (2.14) it follows that the likelihood equations in the submodel M_ψ is

$$l_{/a} = \psi^i P_{i/a}(t;\chi) + D^\rho_{/a}(\chi)t_\rho - \gamma_{/a}(\chi) = 0 \tag{4.11}$$

and under the present assumptions it suffices to prove that $\hat{\chi}$ is a solution to (4.11). Using that

$$t_\rho = \tau_\rho(\hat{\chi}, \hat{\psi}) = \alpha_{/j}(\hat{\psi})A^j_\rho(\hat{\chi}) + B_\rho(\hat{\chi}) \tag{4.12}$$

it follows from (2.12), (2.3) and (2.5) that

$$P_{i/a}(t; \hat{\chi}) = \alpha_{/j}(\hat{\psi})A^j_\rho(\hat{\chi})C^\rho_{i/a}(\hat{\chi}) = 0.$$

Moreover (4.12), (2.4) and (2.6) imply that

$$D^\rho_{/a}(\hat{\chi})t_\rho - \gamma_{/a}(\hat{\chi}) = D^\rho_{/a}(\hat{\chi})\{\alpha_{/j}(\hat{\psi})A^j_\rho(\hat{\chi}) + B_\rho(\hat{\chi})\} - \gamma_{/a}(\hat{\chi}) = 0$$

and the proof is complete. □

In the submodel $M_\chi = \{P_{(\chi,\psi)} : \psi \in \Psi\}$ the maximum likelihood estimate $\hat{\psi}_\chi$ of ψ is the unique solution to the equation

$$\pi(\psi_\chi) = E_{(\chi,\psi_\chi)}\{P(t; \chi)\} = P(t; \chi),$$

i.e.

$$\hat{\psi}_\chi = \psi(P(t; \chi))$$

which may be seen from (2.14) and (4.8).

5. Examples

In this section we give three examples of the result in Corollary 4.1 concerning maximum likelihood estimation in exponential orthogeodesic models. For the sake of notational convenience we consider exponential models of order 2 only. In the terminology of Barndorff-Nielsen and Blæsild (1993) the models considered in the examples below are examples of, respectively, θ-parallel models, τ-parallel models and transformation models. However, due to the fact that these rather different models are all orthogeodesic, the formal way of performing maximum likelihood estimation in these models is the same.

Example 5.1. Suppose that $x_1, \ldots, x_n$ are independent random variables and that x_i is Poisson distributed with mean value $\exp(a + bt_i)$ where t_i is a regression parameter. The joint density of $x_1, \ldots, x_n$ is

$$e^{-e^a \sum_i e^{bt_i}} e^{b \sum_i t_i x_i + a \sum_i x_i} \prod_i \frac{1}{x_i!},$$

i.e. of the form (2.1) with $\theta = (b, a)$ and $t^T(x) = (\sum_i t_i x_i, \sum_i x_i)$. It is easily seen that

$$\tau^T = (\tau_1, \tau_2) = \left(e^a \sum_i t_i e^{bt_i}, e^a \sum_i e^{bt_i}\right) = \tau_2(-h(b), 1), \tag{5.1}$$

where

$$h(b) = -\sum_i t_i e^{bt_i} \Big/ \sum_i e^{bt_i}.$$

Furthermore, formula (5.1) implies that

$$a = \ln \tau_2 - \ln\left(\sum_i e^{bt_i}\right).$$

Setting $\psi = \ln \tau_2$ we have, according to Theorem 2.1, that the model is orthogeodesic relative to (b, ψ) with $A^T(b) = \{-h(b)\ 1\}$, $B(b) = 0$, $C(b) = \{0\ 1\}$, $D(b) = (b, -\ln(\sum_i e^{bt_i}))$ and $\alpha(\psi) = e^{\psi}$, since

$$\theta = (b, a) = \psi(0, 1) + \left(b, -\ln\left(\sum_i e^{bt_i}\right)\right),$$

$$\tau = \begin{Bmatrix} \tau_1 \\ \tau_2 \end{Bmatrix} = e^{\psi} \begin{Bmatrix} -h(b) \\ 1 \end{Bmatrix}$$

and

$$\kappa = e^{\psi}.$$

In this situation formula (4.9) takes the form

$$\begin{Bmatrix} 1 & h(b) \\ 0 & 0 \end{Bmatrix} \begin{Bmatrix} \sum_i t_i x_i \\ \sum_i x_i \end{Bmatrix} = 0.$$

Consequently, the maximum likelihood estimate $\hat{b}$ of b is determined from the equation

$$\frac{\sum_i t_i e^{bt_i}}{\sum_i e^{bt_i}} = \frac{\sum_i t_i x_i}{\sum_i x_i}.$$

From (2.12) we find that

$$P(t; b) = C(b)t = \{0\ 1\} \left(\sum_i t_i x_i, \sum_i x_i\right)^T = \sum x_i,$$

i.e. in this situation the quantity P does not depend on b. Since $\pi = E_{(b,\psi)} P = \tau_2$ the equation (4.10) is equivalent to

$$\hat{\tau}_2 = \sum_i x_i.$$

The model in this example, often referred to as a log-linear Poisson regression, has been considered by Barndorff-Nielsen (1978) who noted that the quantity $\sum x_i$

is a cut which is S-sufficient for τ_2. This is in accordance with the fact that the model is θ-parallel in the terminology of Barndorff-Nielsen and Blæsild (1983*a*).

Example 5.2. If $x_1, \dots, x_n$ are independent and identically $N(\chi, \sigma^2)$-distributed it is easily seen that the density function of $x_1, \dots, x_n$ is of the form (2.1) with $\theta = (\theta^1, \theta^2) = (\chi/\sigma^2, -1/(2\sigma^2))$ and $t^T = (\sum_i x_i, \sum_i x_i^2)$. Moreover, the model is orthogeodesic relative to (χ, θ^2) as may be seen from Theorem 2.1 and the following equations:

$$(\theta^1, \theta^2) = (\chi/\sigma^2, -1/(2\sigma^2)) = \theta^2\{-2\chi \ 1\},$$

$$\begin{Bmatrix} \tau_1 \\ \tau_2 \end{Bmatrix} = \begin{Bmatrix} n\chi \\ n(\chi^2 + \sigma^2) \end{Bmatrix} = -\frac{n}{2\theta^2}\begin{Bmatrix} 0 \\ 1 \end{Bmatrix} + \begin{Bmatrix} n\chi \\ n\chi^2 \end{Bmatrix}$$

and

$$\kappa = \frac{n}{2} ln(2\pi) - \frac{n}{2} ln(-2\theta^2) - n\chi^2\theta^2. \tag{5.2}$$

In this situation the quantity P is, using (2.12),

$$P(t;\chi) = C(\chi)\{t - B(\chi)\} = \{-2\chi \ 1\}\begin{Bmatrix} \sum_i x_i - n\chi \\ \sum_i x_i^2 - n\chi^2 \end{Bmatrix} = \sum_i (x_i - \chi)^2. \tag{5.3}$$

Since formula (4.9) takes the form

$$\begin{Bmatrix} 1 & 0 \\ 2\chi & 0 \end{Bmatrix}\begin{Bmatrix} \sum_i x_i - n\chi \\ \sum_i x_i^2 - n\chi^2 \end{Bmatrix} = \begin{Bmatrix} \sum_i x_i - n\chi \\ 2\chi\left(\sum_i x_i - n\chi\right) \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

it follows that

$$\hat{\chi} = \bar{x} = \frac{1}{n}\sum_i x_i. \tag{5.4}$$

Finally, the formulas (4.8), (4.9) and (5.2)–(5.4) imply that $\hat{\theta}^2$ is determined by the equation

$$-\frac{n}{2\hat{\theta}^2} = \sum_i (x_i - \bar{x})^2$$

or, equivalently, that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2. \tag{5.5}$$

The results (5.4) and (5.5) are of course well-known as is the stochastic independence of the quantities $\hat{\chi}$ and $\hat{\sigma}^2$, a result which is an immediate consequence of the fact that the model is a reproductive τ-parallel model, cf. Barndorff-Nielsen and Blæsild (1983*b*).

Example 5.3. Suppose that a direction on the unit circle is determined by an angle in the interval from 0 to 2π. Furthermore, suppose that $v_1, \ldots, v_n$ are independent observations from the transformation model based on the von Mises-distribution with concentration parameter ψ and mean direction determined by χ. The joint density of $v_1, \ldots, v_n$ is then

$$a(\psi)^n e^{\psi(\cos\chi \sum_i \cos v_i + \sin\chi \sum_i \sin v_i)}. \tag{5.6}$$

Here $a(\psi) = \{2\pi I_0(\psi)\}^{-1}$ where I_0 denotes the modified Bessel function of the first kind and of order 0.

Formula (5.6) is of the form (2.1) with $t^T = (\sum_i \cos v_i, \sum_i \sin v_i)$,

$$\theta = \psi(\cos\chi, \sin\chi) \tag{5.7}$$

and

$$\kappa = -n \ln a(\psi) = \alpha(\psi). \tag{5.8}$$

Furthermore, since

$$\begin{Bmatrix} \tau_1 \\ \tau_2 \end{Bmatrix} = \alpha'(\psi) \begin{Bmatrix} \cos\chi \\ \sin\chi \end{Bmatrix} \tag{5.9}$$

it follows from Theorem 2.1 that the model is orthogeodesic relative to (χ, ψ). Letting

$$t^T = \left(\sum_i \cos v_i, \sum_i \sin v_i \right) = (R\cos d, R\sin d)$$

the equation (4.9) for determining $\hat{\chi}$ takes the form

$$\begin{Bmatrix} R\sin\chi\sin(\chi - d) \\ -R\cos\chi\sin(\chi - d) \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

and, consequently, one has

$$\hat{\chi} = d. \tag{5.10}$$

Furthermore, using (2.12), (5.7), (5.9) and (5.10), we find

$$P(t; \chi) = R\cos(\chi - \hat{\chi}).$$

Thus $P(t; \hat{\chi}) = R$ and using (4.10) the maximum likelihood estimate is determined from the equation

$$\alpha'(\psi) = R.$$

References

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

Barndorff-Nielsen, O. E. (1988). *Parametric Statistical Models and Likelihood*, Springer, Heidelberg.

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

Barndorff-Nielsen, O. E. and Blæsild, P. (1983*b*). Reproductive exponential families, *Ann. Statist.*, **11**, 770–782.

Barndorff-Nielsen, O. E. and Blæsild, P. (1993). Orthogeodesic models, *Ann. Statist.*, **21**, 1018–1039.