The Annals of Statistics
1983, Vol. 11, No. 3, 753–769

EXPONENTIAL MODELS WITH AFFINE DUAL FOLIATIONS

BY O. BARNDORFF-NIELSEN AND P. BLÆSILD

Aarhus University

Suppose an exponential model $\mathscr{M}$ is partitioned into submodels, all of the same parametric dimension. If each of these models corresponds to a linear hypothesis about the canonical parameter and also to a linear hypothesis about the mean parameter, we speak of the partitioning of $\mathscr{M}$ as an affine dual foliation. We study those cases where the parameter sets defining the hypotheses are parallel either in the canonical space or in the mean space, and obtain various characterisations and properties of these cases. It is shown, inter alia, that canonical parallelism and mean parallelism are related to likelihood independence of θ_1 and τ_2 (and hence to S-ancillarity and S-sufficiency), respectively to stochastic independence of $\hat{\theta}_1$ and $\hat{\tau}_2$. Here (θ_1, τ_2) denotes a mixed parametrisation of $\mathscr{M}$ and $\hat{\theta}_1$ and $\hat{\tau}_2$ are the maximum likelihood estimators of θ_1 and τ_2. Also, the two types of parallelism are characterised in terms of observed and expected information. Mean parallelism is closely related to the concept of reproductivity of exponential models that forms the subject of a separate paper. A number of requisite general results for exponential families are established, and these are also of some independent interest.

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

$$p(x; \theta) = a(\theta)b(x)e^{\theta \cdot t(x)}, \tag{1.1}$$

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 $\theta \in \mathrm{int}\Theta$ (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 $\tau(\mathrm{int}\Theta)$. The model $\mathscr{M}$ is assumed to be steep which is equivalent to $\mathscr{T} = \mathrm{int}C$, cf. Barndorff-Nielsen (1978) theorem 9.2. This is the case, in particular, if the canonical parameter domain Θ of (1.1) is open. We take vectors to be row vectors and denote the transpose of a vector v by v^*.

Suppose Θ is partitioned into subsets which, except perhaps for some singular cases, are differentiable submanifolds of Θ, of some fixed dimension d where $1 \le d < k$. Typically, an arbitrary parameter ψ of $\mathscr{M}$ induces such a partition. We then say that we have a *foliation* of Θ *of dimension d*, and the subsets of Θ that constitute the foliation are called *leaves*. To any d-dimensional foliation of Θ there is a dual foliation of the space $\mathscr{T}$, namely the partition of $\mathscr{T}$ induced from the partition of $\mathrm{int}\Theta$ by the mean value mapping τ. To distinguish between these two foliations we use the terms *θ-foliation* and *τ-foliation*. If each leaf of a foliation, of dimension d, of Θ or $\mathscr{T}$ is contained in a d-dimensional affine subspace, we call the foliation in question *affine*. Such foliations are generally of statistical interest, affine θ-foliations corresponding to linear hypotheses about the canonical parameter and affine τ-foliations corresponding to linear hypotheses about the mean value parameter. Cases where the dual foliations, of Θ and $\mathscr{T}$ respectively, are both affine should command special attention, and such cases are the objects of study in the present paper. We speak of models with this structure as exponential models with *affine dual foliations*, and our purpose is to derive properties of a general nature for such models.

To obtain useful general results it seems necessary to single out suitable subclasses of the class of all exponential models with affine dual foliations for separate study. We shall

Received September 1982; revised March 1983.

AMS 1980 *subject classifications.* Primary, 62E15; secondary, 62F99.

Key words and phrases. Cuts, duality, expected information, independence, Legendre transforms, mixed parametrisations, observed information, pivots, profile likelihood.

discuss two such subclasses here, namely those with *θ-parallel* and *τ-parallel* foliations, respectively. A third subclass, that of *θ-τ-cone* foliations will be briefly indicated at the end of the present section.

We say that the model $\mathscr{M}$ possesses a θ-parallel foliation of dimension d if there exists an affine dual foliation of $\mathscr{M}$ such that the leaves of the θ-foliation can be described as the intersections between Θ and the affine d-dimensional subspaces of R^k that are parallel to some given d-dimensional linear subspace of R^k. The concept of τ-parallel foliation is defined analogously. Let (t_1, t_2), (θ_1, θ_2) and (τ_1, τ_2) denote similar partitions of t, θ and τ. In discussing θ-parallel foliations, it causes no loss of generality to assume that the leaves of the foliation are the subsets of Θ obtained by fixing θ_1 at its various possible values, and analogously for τ-foliations. Various characterisations and properties of these two types of foliations are discussed in Sections 4 and 5, respectively. In particular, it is shown that the existence of a θ-parallel foliation is equivalent to the existence of a proper cut in $\mathscr{M}$. (For a discussion of the concept of a cut and of its intimate connections to S-ancillarity and S-sufficiency see Barndorff-Nielsen (1978), Sections 10.2–4.) This, in turn, is essentially equivalent to likelihood independence of θ_1 and τ_2. On the other hand, the existence of a τ-parallel foliation determined by fixing τ_2 at its potential values turns out to be closely related to stochastic independence of the maximum likelihood estimators $\hat{\theta}_1$ and $\hat{\tau}_2$, though a complete equivalence has not been established.

Legendre transforms turn up in a natural and useful way in the study of θ-parallel and τ-parallel foliations and for this reason we give in Section 2 the definition and some elementary properties of the Legendre transformation.

Section 3 contains various elementary general results for exponential models that are required in Sections 4 and 5. Several of these results are new and of independent statistical interest.

Throughout the paper we use the concept of an indefinite integral of a $r \times s$ matrix function of a s-dimensional vector. If $h(y) = \{h_{ij}(y)\}_{ij}$, where $y = (y_1, \cdots, y_s)$, we say that the r-dimensional function $H = (H_1, \cdots, H_r)$ of y is an indefinite integral of h if $\partial H^*/\partial y = \{\partial H_i/\partial y_j\}_{ij} = h$. A sufficient condition for h to possess an indefinite integral is that the derivatives $\partial h_{ij}/\partial y_k$ exist and are continuous and satisfy $\partial h_{ij}/\partial y_k = \partial h_{ik}/\partial y_j$ for all $i = 1, \cdots, r$, $j = 1, \cdots, s$ and $k = 1, \cdots, s$. Indefinite integrals will be denoted by the capital letters corresponding to the lower case letters used for the matrix functions.

The present report is strongly related to the report by Barndorff-Nielsen and Blæsild (1983). The paper by Bar-Lev and Reiser (1982) has been an important source for some of the developments in the two reports and together these reports provide inter alia an extension, from exponential models of order two to exponential models of arbitrary order, of the main results of Bar-Lev and Reiser (1982), as given in their Theorems 3.1 and 3.2.

As announced above, we conclude this section by indicating a third type of dual affine foliations that seems of some particular interest but for which we have not, so far, obtained any general results. We say that the model $\mathscr{M}$ has a *θ-τ-cone* foliation of dimension d if there exists an affine dual foliation of $\mathscr{M}$ such that each leaf of the θ-foliation can be described as the intersection between Θ and a d-dimensional cone, all the cones having the same top point, and similarly for the leaves of the τ-foliation.

The models corresponding to the gamma distributions, the von Mises-Fisher distributions in arbitrary dimensions, the hyperboloid distributions in arbitrary dimensions, and Fisher's gamma hyperbola all provide examples of exponential models with θ-τ-cone foliations.

In the investigation of models with such foliations the mixed parametrisation (θ_1, τ_2) does not seem to be of relevance, in contrast to the study of models with θ-parallel foliations and τ-parallel foliations that form the subject of Sections 4 and 5 below. We desist here from further remarks concerning exponential models with θ-τ-cone foliations, noting, however, that the models in the four examples mentioned above are all exponential transformation models.

2. The Legendre transform. Let f be a real differentiable function defined on an open subset U of R^k and define a new function $\check{f}$ on U by

$$\check{f}(x) = x \cdot y - f(x) \tag{2.1}$$

where

$$y = (Df)(x) = \frac{\partial f}{\partial x}(x)$$

is the gradient of f at x. We speak of $\check{f}$ as the *Legendre transform* of f. If the gradient mapping Df is a one-to-one function, one may think of $\check{f}$ as a function of y, and we shall then, for simplicity, write $\check{f}(y)$ instead of $\check{f}((Df)^{-1}(y))$. This is the case, in particular, provided f has continuous partial derivatives of the second order and provided the Hessian of f

$$(D^2 f)(x) = \frac{\partial^2 f}{\partial x \partial x^*}(x)$$

is nonsingular throughout U. In these circumstances the Legendre transform of $\check{f}(y)$ is well defined and we have $(D\check{f})(y) = x$ and

$$\check{\check{f}}(x) = y \cdot x - \check{f}(y) = f(x),$$

i.e. applying the Legendre transformation twice recovers the original function. Furthermore,

$$(D^2\check{f})(y) = \{(D^2 f)(x)\}^{-1}.$$

The classical areas of application of the Legendre transformation are certain types of variational problems and (partial) differential equations, cf. Courant and Hilbert (1952, 1953) and Kamke (1930, 1974). More recently, the Legendre transformation has acquired some prominence in convex analysis, Rockafellar (1970). In the study of exponential families of probabilities distributions, the Legendre transform of the cumulant generating function is a useful tool, with a simple statistical interpretation; this is discussed in Barndorff-Nielsen (1978). The latter special instance of the transform also occurs naturally in the present study, cf. Sections 3, 4, and 5.

We shall consider the extension of (2.1) to vector-valued functions. Suppose f takes values in R^d and is defined on an open subset of R^k. We assume that f is differentiable and define the Legendre transform $\check{f}$ of f as the function on U and with values in R^d which is given by

$$\check{f}(x) = x(Df)^*(x) - f(x)$$

where $(Df)^*$ is the transpose of the $d \times k$ matrix $Df = \partial f^*/\partial x$.

Later on we shall refer to the following elementary property of the Legendre transform.

LEMMA 2.1. *Suppose f has continuous partial derivatives of the first order in the open subset U of R^k and that $0 \in U$. If the Legendre transform $\check{f}$ of f is constant on U then f is an affine function on U, i.e. $f(x) = xA + B$ for some constant $k \times d$ matrix A and some constant d-dimensional vector B.* □

PROOF. Let C be a constant vector of dimension d and suppose that $\check{f}(x) = C$ for $x \in U$. There is no loss of generality in assuming that $d = 1$ and $C = 0$, and then we have

$$x \cdot (Df)(x) = f(x).$$

At first, assume that the unit sphere is contained in U. In this case it is well-known that the partial differential equation for f has a solution of the form $f(x) = |x| f(e(x))$ where $e(x)$ denotes the unit vector in the direction of x. Since f is differentiable also at the origin, we therefore have $f(te) = tf(e)$ for any k-dimensional unit vector e and any scalar t such that $te \in U$. Differentiating this equation with respect to t and setting $t = 0$, we obtain $e \cdot (Df)(0) = f(e)$ whence $te \cdot (Df)(0) = f(te)$, i.e. $f(x) = x \cdot (Df)(0)$ for $x \in U$. The proof is now easily completed. □

3. Some elementary general results for exponential models. A variety of elementary general results for exponential families are presented in this section. These results are prerequisites for the developments in Sections 4 and 5, but they are also of independent interest.

Let $\theta = (\theta_1, \theta_2)$ be a partition of θ into components θ_1 and θ_2 of dimensions k_1 and k_2, respectively, and let $t = (t_1, t_2)$ and $\tau = (\tau_1, \tau_2)$ denote similar partitions of t and of the mean value parameter $\tau = E_\theta t$, which is defined on $\text{int}\Theta$, the interior of Θ. The mixed parameter (θ_1, τ_2) plays an important role in the following discussion and the first three of the results of the present section concern the mixed parametrisation. (Strictly speaking, since we take $\text{int}\Theta$ as the domain of definition of τ, (θ_1, τ_2) provides a parametrisation only for those probability measures (1.1) for which $\theta \in \text{int}\Theta$.)

The first lemma is an extension of Theorem 8.4 in Barndorff-Nielsen (1978) from regular to steep exponential models.

LEMMA 3.1. *Suppose that the full exponential model* (1.1) *is steep. Let Θ_1 denote the possible values of the component θ_1, i.e. $\Theta_1 = \{\theta_1 \mid \text{there exist a } \theta_2 \text{ such that } (\theta_1, \theta_2) \in \Theta\}$, and let C_2 denote the closed convex hull of the support of the marginal distribution of t_2. Then*

$$(\theta_1, \tau_2)(int\Theta) = int\Theta_1 \times intC_2,$$

i.e. the components θ_1 and τ_2 of the mixed parameter (θ_1, τ_2) are variation independent. □

PROOF. Suppose $\theta_0 = (\theta_{01}, \theta_{02}) \in \text{int}\,\Theta$ and let $\mathscr{M}(\theta_{01})$ denote the submodel of $\mathscr{M}$ whose parameter domain is determined by $\Theta(\theta_{01}) = \{\theta \in \Theta \mid \theta = (\theta_{01}, \theta_2)\}$. Clearly $\mathscr{M}(\theta_{01})$ is a full and steep exponential model of order k_2 with minimal representation

$$q(x; (\theta_{01}, \theta_2)) = \frac{p(x; (\theta_{01}, \theta_2))}{p(x; (\theta_{01}, \theta_{02}))} = \frac{a(\theta_{01}, \theta_2)}{a(\theta_{01}, \theta_{02})} e^{(\theta_2 - \theta_{02})\cdot t_2(x)}.$$

If $\Theta_2(\theta_{01})$ denotes the set $\{\theta_2 : (\theta_{01}, \theta_2) \in \Theta\}$, it follows from Theorem 9.2 in Barndorff-Nielsen (1978), that $\tau_2(\text{int}\Theta_2(\theta_{01})) = \text{int}\, C_2$ independently of the value of θ_{01}. □

The next lemma presents the Jacobian matrix for the mapping taking the mixed parameter (θ_1, τ_2) into the mixed parameter (τ_1, θ_2). In order to formulate this result recall that if κ denotes the cumulant transform of (1.1), i.e.

$$\kappa(\theta) = -\ln a(\theta),$$

then $\tau = (D\kappa)(\theta)$ and $\Sigma = (D^2\kappa)(\theta)$ is the variance matrix of t. Let $\Delta = \Sigma^{-1}$ and let

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \tag{3.1}$$

be the partitions of Σ and Δ such that Σ_{11} and Δ_{11} are $k_1 \times k_1$ matrices.

LEMMA 3.2.. *The Jacobian matrix for the mapping taking (θ_1, τ_2) into (τ_1, θ_2) is*

$$\begin{aligned} J_{(\tau_1,\theta_2)}(\theta_1, \tau_2) &= \frac{\partial(\tau_1, \theta_2)^*}{\partial(\theta_1, \tau_2)} \\ &= \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1} \end{bmatrix} = \begin{bmatrix} \Delta_{11}^{-1} & -\Delta_{11}^{-1}\Delta_{12} \\ \Delta_{21}\Delta_{11}^{-1} & \Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12} \end{bmatrix}. \quad \square \end{aligned} \tag{3.2}$$

PROOF. Using the well-known result that

$$J_{(\tau_1,\tau_2)}(\theta_1, \theta_2) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

it follows that

$$(3.3)\qquad J_{(\theta_1,\theta_2)}(\theta_1, \tau_2) = J^{-1}_{(\theta_1,\tau_2)}(\theta_1, \theta_2) = \begin{bmatrix} I_{k_1} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I_{k_1} & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1} \end{bmatrix},$$

where I_{k_1} denotes the $k_1 \times k_1$ identity matrix. An application of the chain rule implies that

$$\begin{aligned} J_{(\tau_1,\theta_2)}(\theta_1, \tau_2) &= J_{(\tau_1,\theta_2)}(\theta_1, \theta_2) J_{(\theta_1,\theta_2)}(\theta_1, \tau_2) \\ &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & I_{k_2} \end{bmatrix} \begin{bmatrix} I_{k_1} & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{12} & \Sigma_{22}^{-1} \end{bmatrix}. \end{aligned}$$

The last equality in (3.2) expresses standard formulae for inverses of partitioned matrices. □

We may apply (3.2) to obtain

$$\frac{\partial \kappa(\theta_1, \tau_2)}{\partial \theta_1} = \tau_1 + \tau_2 \frac{\partial \theta_2^*(\theta_1, \tau_2)}{\partial \theta_1} = \tau_1 - \tau_2 \frac{\partial \tau_1(\theta_1, \tau_2)}{\partial \tau_2^*}$$

i.e., in an obvious notation,

$$(3.4)\qquad \frac{\partial \kappa(\theta_1, \tau_2)}{\partial \theta_1} = -\check{\tau}_1(\tau_2 | \theta_1).$$

The following two lemmas are concerned, respectively, with the expected (or Fisherian) information for the mixed parameter (θ_1, τ_2) and with the observed profile information for θ_1.

LEMMA 3.3. *The expected information for* (θ_1, τ_2) *is*

$$i(\theta_1, \tau_2) = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} = \begin{bmatrix} \Delta_{11}^{-1} & 0 \\ 0 & \Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12} \end{bmatrix}. \qquad \square$$

PROOF. Since

$$i(\theta_1, \tau_2) = \frac{\partial(\theta_1, \theta_2)}{\partial(\theta_1, \tau_2)^*} i(\theta_1, \theta_2) \frac{\partial(\theta_1, \theta_2)^*}{\partial(\theta_1, \tau_2)}$$

the result follows from (3.3) and the fact that $i(\theta_1, \theta_2) = \Sigma$. □

The log-likelihood function ℓ for θ corresponding to the observed value t of the canonical statistic of (1.1) is

$$(3.5)\qquad \ell(\theta; t) = \theta \cdot t - \kappa(\theta)$$

and hence the log-likelihood function ℓ_0 for (θ_1, τ_2) is

$$(3.6)\qquad \ell_0((\theta_1, \tau_2); t) = \ell(\theta(\theta_1, \tau_2); t)$$

and the observed information is

$$j((\theta_1, \tau_2); t) = -\frac{\partial^2 \ell_0((\theta_1, \tau_2); t)}{\partial(\theta_1, \tau_2)\partial(\theta_1, \tau_2)^*}.$$

Let $\tilde{\ell}_0$ be the profile (or partially maximised) log-likelihood function for θ_1, i.e.

$$\tilde{\ell}_0(\theta_1; (t_1, t_2)) = \sup_{\tau_2} \ell_0((\theta_1, \tau_2); (t_1, t_2)) = \ell_0((\theta_1, t_2); (t_1, t_2)),$$

where the last equality presupposes $t \in \operatorname{int} C$. The information function corresponding to $\tilde{\ell}_0$ will be denoted by

$$\tilde{j}(\theta_1; (t_1, t_2)) = -\frac{\partial^2 \tilde{\ell}_0(\theta_1; (t_1, t_2))}{\partial \theta_1 \partial \theta_1^*}.$$

Finally, let $j_{11}(\theta_1, \tau_2)$ and $i_{11}(\theta_1, \tau_2)$ stand for the (1, 1)-bloc element of $j(\theta_1, \tau_2)$ and $i(\theta_1, \tau_2)$, respectively. Thus, by Lemma 3.3, $i_{11}(\theta_1, \tau_2) = \Delta_{11}^{-1}$. With these notations one has:

LEMMA 3.4. *The observed information function $\tilde{j}$ based on the profile log-likelihood function for θ_1 depends, considered as a function of (θ_1, t_1, t_2) on $(\mathrm{int}\Theta_1) \times \mathrm{int}C$, on (θ_1, t_2) only and equals $j_{11}(\theta_1, t_2)$ and also $i_{11}(\theta_1, t_2)$, i.e.*

$$\tilde{j}(\theta_1; (t_1, t_2)) = \tilde{j}(\theta_1; t_2) = j_{11}(\theta_1, t_2) = i_{11}(\theta_1, t_2). \qquad \square$$

PROOF. Let b and x be vectors of dimension r and s, respectively, and let $M = M(x)$ be a $r \times s$ matrix differentiable with respect to $x = (x_1, \cdots, x_s)$. We then define the product $b \times \partial M/\partial x^*$ to be the $s \times s$ matrix given by

$$\begin{bmatrix} b \dfrac{\partial M}{\partial x_1} \\ \vdots \\ b \dfrac{\partial M}{\partial x_s} \end{bmatrix}.$$

Differentiating (3.6) twice with respect to (θ_1, τ_2) we find

$$\begin{aligned} j((\theta_1, \tau_2); (t_1, t_2)) &= i(\theta_1, \tau_2) - (t - \tau) \times \frac{\partial}{\partial(\theta_1, \tau_2)^*}\left(\frac{\partial(\theta_1, \theta_2)^*}{\partial(\theta_1, \tau_2)}\right) \\ &= i(\theta_1, \tau_2) - (t_2 - \tau_2) \times \frac{\partial}{\partial(\theta_1, \tau_2)^*}[-\Sigma_{22}^{-1}\Sigma_{21}, \Sigma_{22}^{-1}] \end{aligned}$$

from which we obtain, using (3.3) and $t \in \mathrm{int}C$, that

$$j((\theta_1, t_2); (t_1, t_2)) = i(\theta_1, t_2). \tag{3.7}$$

It is well known that, whether the model is exponential or not, the observed formation function (i.e. the inverse of the observed information function) calculated from a profile likelihood equals the relevant part of the observed formation function from the full likelihood (Richards, 1961, Patefield, 1977), i.e. in mathematical terms

$$\tilde{j}^{-1}(\theta_1) = j^{11}(\theta_1, t_2),$$

where $j^{11}(\theta_1, t_2)$ denotes the (1, 1)-bloc element of the inverse of the matrix $j(\theta_1; t_2)$. Lemma 3.4 follows from this result in conjunction with (3.7) and Lemma 3.3. $\square$

The first conclusion of Lemma 3.4 may be formulated as saying that the estimated observed formation function for θ_1 does not depend on t_1. This may be compared to the well known result that the full observed information $j(\theta)$ for θ does not depend on t, as it equals the expected information function $i(\theta)$.

We shall repeatedly use the fact if P_θ denotes the probability measure given by (1.1), if u is a statistic and if $p(u; \theta)$ is the density of the lifted measure uP_θ with respect to some σ-finite measure dominating the class $\{uP_\theta : \theta \in \Theta\}$ of marginal distributions of u then, for any elements θ_0 and θ of Θ, we have

$$p(u; \theta) = E_{\theta_0}\left\{\frac{dP_\theta}{dP_{\theta_0}} \,\middle|\, u\right\} p(u; \theta_0), \tag{3.8}$$

cf. Barndorff-Nielsen (1978), Section 8.2 (iii).

Quite generally, we shall use the notation

$$x \sim EM(t(x); \theta)$$

to indicate that a random variate x follows an exponential model with $t(x)$ as the canonical statistic and θ as the associated canonical parameter.

Under the exponential model (1.1) it may happen that the marginal distributions of a component t_2, say, of the minimal canonical statistic t constitute an exponential model. The final lemma in this section is concerned with the particularly simple case in which $t_2 \sim EM((H(t_2), t_2); \theta)$ where H is some k_1-dimensional vector function. We use the symbol $\perp$ to indicate stochastic independence between random variates.

LEMMA 3.5. *Under the exponential model* (1.1), *let* $t = (t_1, t_2)$ *be a partition of the canonical statistic and consider a* k_1*-dimensional statistic of the form* $H(t_2)$ *for some function* H. *Then*

$$t_2 \sim EM((H(t_2), t_2); \theta) \Leftrightarrow t_1 - H(t_2) \perp t_2 \Leftrightarrow t_1 - H(t_2) \sim EM(t_1 - H(t_2); \theta_1).$$

In this case the distribution of $t_1 - H(t_2)$ *depends on* θ_1 *only, and, with*

$$p(t_2; \theta) = a_0(\theta)e^{\theta \cdot (H(t_2), t_2)}$$

as an exponential representation of the model for t_2, *the Laplace transform of* $t_1 - H(t_2)$ *may be expressed as*

$$E_{\theta_1} e^{\lambda \cdot \{t_1 - H(t_2)\}} = \frac{a_0(\theta + (\lambda, 0))}{a_0(\theta)} \frac{a(\theta)}{a(\theta + (\lambda, 0))},$$

where the right hand side, in fact, depends on θ_1 *only.* □

This lemma is identical to Theorem 2.1 of Barndorff-Nielsen and Blæsild (1983) and the proof is given in that paper. (Note also the reference there to Bar-Lev, 1983.)

We conclude this section by pointing out an important duality between the maximised log-likelihood function and the cumulant transform of the exponential model (1.1). This duality, which will be used in Section 5, is established via the Legendre transformation, discussed in Section 2.

Letting $\hat{\ell}$ denote the maximised log-likelihood function

$$\hat{\ell}(t) = \sup_\theta \ell(\theta; t)$$

we have, using (3.5), that

$$\check{\kappa}(\theta) = \hat{\ell}(\tau) \tag{3.9}$$

for $\theta \in \text{int}\Theta$ and $\tau = \tau(\theta) \in \mathscr{T} = \tau(\text{int}\Theta)$. Conversely, the Legendre transform of $\hat{\ell}(\tau)$ equals $\kappa(\theta)$. Furthermore, again for $\theta \in \text{int}\Theta$ and $\tau = \tau(\theta) \in \mathscr{T}$, we have the duality relations

$$\frac{\partial \kappa}{\partial \theta} = \tau \qquad \frac{\partial \hat{\ell}}{\partial \tau} = \theta \tag{3.10}$$

$$\frac{\partial^2 \kappa}{\partial \theta \partial \theta^*} = \Sigma \qquad \frac{\partial^2 \hat{\ell}}{\partial \tau \partial \tau^*} = \Sigma^{-1}. \tag{3.11}$$

In particular, the gradient of $\hat{\ell}$ provides the inverse of the mean value mapping $\tau(\theta)$. (For further discussion, see Barndorff-Nielsen, 1978, Section 9.1).

4. Models with θ-parallel foliations. In the discussion of these models we may without loss of generality assume that the foliation of intΘ is of the form $\{\text{ri}\Theta(\theta_1):\theta_1 \in \text{int}\Theta_1\}$, where ri$\Theta(\theta_1)$ denotes the relative interior of the set of parameter values $\theta = (\theta_1, \theta_2)$ with fixed θ_1. Let $\mathscr{T}(\theta_1) = \tau(\text{ri}\Theta(\theta_1))$. Since the model is assumed to be θ-parallel $\mathscr{T}(\theta_1)$ is the intersection of a k_2-dimensional affine subspace and $\mathscr{T} = \tau(\text{int}\Theta)$. Consequently, there exists a $k \times k_1$ matrix $d(\theta_1)$ and a $1 \times k_1$ vector $e(\theta_1)$ such that

$$\mathscr{T}(\theta_1) = \{\tau \in \mathscr{T}: \tau d(\theta_1) + e(\theta_1) = 0\}, \quad \theta_1 \in \text{int}\Theta_1. \tag{4.1}$$

By assumption, the rank of $d(\theta_1)$ is k_1 for every $\theta_1 \in \operatorname{int}\Theta_1$. If

$$d(\theta_1) = \begin{bmatrix} d_1(\theta_1) \\ d_2(\theta_1) \end{bmatrix}$$

denotes a partition of $d(\theta_1)$ such that $d_1(\theta_1)$ is a $k_1 \times k_1$ matrix one has

$$\tau d(\theta_1) + e(\theta_1) = \tau_1 d_1(\theta_1) + \tau_2 d_2(\theta_1) + e(\theta_1).$$

The matrix $d_1(\theta_1)$ must be regular. In fact, if this was not so then for $\theta_1 \in \operatorname{int}\Theta_1$ and some non-null $1 \times k_1$ vector $\phi = \phi(\theta_1)$ one would have $d_1(\theta_1)\phi^* = 0$ whence, setting $\phi_2^* = d_2(\theta_1)\phi^*$ and $\phi_0 = e(\theta_1)\phi^*$, it follows that

$$\tau_2 \cdot \phi_2 + \phi_0 = 0 \quad \text{for any} \quad \tau_2 \in \mathscr{T}_2, \tag{4.2}$$

where $\mathscr{T}_2$ denotes the set of the possible values of τ_2; according to Lemma 3.1, $\mathscr{T}_2$ is the same as the set $\mathscr{T}(\theta_1)_2 = \{\tau_2 : \tau_2 = \tau_2(\theta), \theta \in \Theta(\theta_1)\}$. Furthermore, since $d(\theta_1)$ has rank k_1 and $d_1(\theta_1)\phi^* = 0$ the vector ϕ_2 is different from 0. Relation (4.2) now yields a contradiction, for $\mathscr{T}_2$ is an open subset of R^{k_2}. Thus, letting $h(\theta_1) = d_2(\theta_1)d_1(\theta_1)^{-1}$ and $k(\theta_1) = -e(\theta_1)d_1(\theta_1)^{-1}$ we may rewrite (4.1) as

$$\mathscr{T}(\theta_1) = \{\tau \in \mathscr{T} : \tau_1 = -\tau_2 h(\theta_1) + k(\theta_1)\}, \quad \theta_1 \in \operatorname{int}\Theta_1.$$

This shows that τ_1, considered as a function of the mixed parameter (θ_1, τ_2), is of the form

$$\tau_1(\theta_1, \tau_2) = -\tau_2 h(\theta_1) + k(\theta_1). \tag{4.3}$$

We are now set to show:

THEOREM 4.1. *Consider the exponential model $\mathscr{M}$ with exponential representation* (1.1). *The following five statements are equivalent.*

(i) *$\mathscr{M}$ has a θ-parallel foliation, given* (*without loss of generality*) *by* $\{\operatorname{ri}\Theta(\theta_1) : \theta_1 \in \operatorname{int}\Theta_1\}$.

(ii) $\tau_1(\theta_1, \tau_2)$ *is of the form*

$$\tau_1(\theta_1, \tau_2) = -\tau_2 h(\theta_1) + k(\theta_1). \tag{4.4}$$

(iii) $\theta_2(\theta_1, \tau_2)$ *is of the form*

$$\theta_2(\theta_1, \tau_2) = H(\theta_1) + m(\tau_2). \tag{4.5}$$

(iv) $\kappa(\theta_1, \tau_2)$ *is of the form*

$$\kappa(\theta_1, \tau_2) = K(\theta_1) + \check{M}(\tau_2). \tag{4.6}$$

(v) *t_2 is a proper cut of size k_2 in $\mathring{\mathscr{P}} = \{P_\theta : \theta \in \operatorname{int}\Theta\}$, i.e. t_2 is S-sufficient for τ_2 and S-ancillary for θ_1.* □

REMARK. A number of examples of exponential models with proper cuts may be found in Barndorff-Nielsen (1978).

PROOF. The implication (i) ⇒ (ii) has already been established and the converse is trivial.

From (3.2) one obtains

$$\frac{\partial \theta_2^*}{\partial \theta_1}(\theta_1, \tau_2) = \Sigma_{22}^{-1}\Sigma_{21} = -\frac{\partial \tau_1}{\partial \tau_2^*}(\theta_1, \tau_2) \tag{4.7}$$

from which the equivalence of (ii) and (iii) follows. Specifically, (4.4) and (4.7) imply that

$$\frac{\partial \theta_2^*}{\partial \theta_1}(\theta_1, \tau_2) = h(\theta_1)$$

and letting H be a function of θ_1 with values in R^{k_2} such that $\partial H^*/\partial\theta_1 = h$ we obtain formula (4.5) for some vector function m. To prove the converse implication (iii) $\Rightarrow$ (ii) note that (4.5) and (4.7) imply

$$\frac{\partial \tau_1}{\partial \tau_2^*}(\theta_1, \tau_2) = -\frac{\partial H^*}{\partial \theta_1}(\theta_1) = -h(\theta_1)$$

from which we obtain formula (4.4) for some vector function k.

To prove (iii) $\Rightarrow$ (iv) we observe, using (4.4) and (4.5), that

$$\begin{aligned}\frac{\partial \kappa}{\partial \theta_1}(\theta_1, \tau_2) &= \frac{\partial \kappa}{\partial \theta_1}(\theta_1, \theta_2) + \frac{\partial \kappa}{\partial \theta_2}(\theta_1, \theta_2)\frac{\partial \theta_2^*}{\partial \theta_1}(\theta_1, \tau_2)\\ &= \tau_1 + \tau_2 h(\theta_1) = k(\theta_1)\end{aligned}$$

and hence

$$\kappa(\theta_1, \tau_2) = K(\theta_1) + L(\tau_2),$$

where K and L are real valued functions of θ_1 and τ_2, respectively, and where $\partial K/\partial\theta_1 = k$. Now,

$$\frac{\partial L}{\partial \tau_2}(\tau_2) = \frac{\partial \kappa}{\partial \tau_2}(\theta_1, \tau_2) = \frac{\partial \kappa}{\partial \theta_2}(\theta_1, \theta_2)\frac{\partial \theta_2^*}{\partial \tau_2}(\theta_1, \tau_2) = \tau_2 \frac{\partial m^*}{\partial \tau_2}(\tau_2)$$

which implies that

$$L(\tau_2) = \tau_2 \cdot m(\tau_2) - M(\tau_2) = \check{M}(\tau_2),$$

where M is an indefinite integral of m. This shows that (iii) $\Rightarrow$ (iv).

The equivalence of (iii) and (v) follows from Theorem 10.4 of Barndorff-Nielsen (1978).

The proof of Theorem 4.1 is now completed by showing that (iv) $\Rightarrow$ (ii). Differentiating (4.6) with respect to θ_1 and using (3.4) we find

$$-\check{\tau}_1(\tau_2 \mid \theta_1) = k(\theta_1).$$

There is no loss of generality in assuming $0 \in \mathscr{T}_2$ and Lemma 2.1 then implies that τ_1 is of the form (4.4). □

It may be noted that when (iii) and (iv) hold then the representation (1.1) may be rewritten as

$$p(x; \theta_1, \tau_2) = b(x)\exp(-\check{M}(\tau_2) + m(\tau_2)\cdot t_2)\exp(-K(\theta_1) + \theta_1 \cdot t_1 + H(\theta_1)\cdot t_2).$$

Using Theorem (4.1) and formula (3.2), the following two alternative characterisations of θ-parallelism are easily established.

THEOREM 4.2. *Let $\mathscr{M}$ be the exponential model with representation* (1.1). *Then the following three conditions are equivalent.*

(i) *$\mathscr{M}$ has a θ-parallel foliation of the form considered in Theorem* 4.1.

(ii) *$\Sigma_{12}\Sigma_{22}^{-1}$ considered as a function of (θ_1, τ_2) depends on θ_1 only.*

(iii) *The* (2, 2)*-bloc element in the partition of $i(\theta_1, \tau_2)$, i.e. $i_{22}(\theta_1, \tau_2) = \Sigma_{22}^{-1}$, depends on τ_2 only.* □

5. Models with τ-parallel foliations. Because of the duality between intΘ and $\mathscr{T}$, as embodied in the two convex functions $\kappa(\theta)$ and $\hat{\ell}(\tau)$ and the relations (3.10) and (3.11), we may without further proof assert the equivalence of the dual versions of the statements (i)–(iv) in Theorem 4.1. This is done in Theorem 5.1.

More interesting, however, is the problem of finding the appropriate dual statement to that of (v) in Theorem 4.1 since (v), essentially, states that θ_1 and τ_2 are likelihood independent, in the sense of Barndorff-Nielsen (1978), and this represents the statistical content of the concept of θ-parallelism. From the duality viewpoint, it seems natural to investigate whether the exponential models with τ-parallel foliations are those in which $\hat{\theta}_1$ and $\hat{\tau}_2$ are stochastically independent. (It may be noted that $\hat{\theta}_1$ and $\hat{\tau}_2$ are always asymptotically independent, cf. Lemma 3.3.) We are able to prove that the independence, under mild conditions, implies τ-parallelism, whereas at present we can prove the converse only in the special cases discussed at the end of this section.

However, before taking up this problem, we shall present an alternative statistical characterisation of τ-parallelism in terms of the observed profile information for θ_1. This description is based on still another characterisation, given in Theorem 5.2.

THEOREM 5.1. *Consider the exponential model $\mathscr{M}$ with exponential representation* (1.1). *Then the following four statements are equivalent.*

(i) *$\mathscr{M}$ has a τ-parallel foliation, given* (*without loss of generality*) *by* $\{\mathrm{ri}\, \mathscr{T}(\tau_2) : \tau_2 \in \mathscr{T}_2\}$.

(ii) $\theta_2(\theta_1, \tau_2)$ *is of the form*

$$\theta_2(\theta_1, \tau_2) = -\theta_1 h(\tau_2) + k(\tau_2). \tag{5.1}$$

(iii) $\tau_1(\theta_1, \tau_2)$ *is of the form*

$$\tau_1(\theta_1, \tau_2) = m(\theta_1) + H(\tau_2). \tag{5.2}$$

(iv) $\hat{\ell}(\theta_1, \tau_2)$ *is of the form*

$$\hat{\ell}(\theta_1, \tau_2) = \check{M}(\theta_1) + K(\tau_2). \tag{5.3}$$

Furthermore, if one of the conditions (i)–(iv) *is fulfilled* (1.1) *may be rewritten as*

$$p(x; \theta_1, \tau_2) = b(x)\exp(\theta_1 \cdot \check{H}(\tau_2) - \check{K}(\tau_2) - M(\theta_1))\exp(\theta_1 \cdot \{t_1 - t_2 h^*(\tau_2)\} + k(\tau_2) \cdot t_2). \square \tag{5.4}$$

PROOF. Because of the remark made above, only formula (5.4) needs to be proved. As mentioned in Section 3, $\kappa(\theta)$ equals the Legendre transform of $\hat{\ell}(\tau)$, i.e.

$$\kappa(\theta) = \tau \cdot \theta - \hat{\ell}(\theta)$$

and from (5.1)–(5.3) one finds

$$\begin{aligned} \kappa(\theta_1, \tau_2) &= \theta_1 \cdot \{m(\theta_1) + H(\tau_2)\} + \{-\theta_1 h(\tau_2) + k(\tau_2)\} \cdot \tau_2 - \check{M}(\theta_1) - K(\tau_2) \\ &= -\theta_1 \cdot \check{H}(\tau_2) + \check{K}(\tau_2) + M(\theta_1). \end{aligned} \tag{5.5}$$

Inserting this and (5.1) into (1.1) the proof is completed. $\square$

The following corollary is an immediate consequence of (5.4).

COROLLARY 5.1. *Suppose $\mathscr{M}$ has a τ-parallel foliation of the form considered in Theorem* 5.1. *Then the Laplace transform of the quantity*

$$p = t_1 - t_2 h^*(\tau_2) + \check{H}(\tau_2)$$

is given by

$$E_\theta\{\exp(\lambda \cdot p)\} = \exp(M(\theta_1 + \lambda) - M(\theta_1)).$$

Consequently, the distribution of p depends on θ_1 only, i.e. p is a pivot provided θ_1 is known. $\square$

Quite similar to Theorem 4.2 we have:

THEOREM 5.2. *Let $\mathscr{M}$ be the exponential model with representation* (1.1). *Then the following three conditions are equivalent.*

(i) $\mathscr{M}$ *has a* τ*-parallel foliation of the form considered in Theorem* 5.1.

(ii) $\Sigma_{12}\Sigma_{22}^{-1}$ *considered as a function of* (θ_1, τ_2) *depends on* τ_2 *only.*

(iii) *The* (1, 1)*-bloc element in the partition of* $i(\theta_1, \tau_2)$, *i.e.* $i_{11}(\theta_1, \tau_2) = \Delta_{11}^{-1}$, *depends on* θ_1 *only.* □

As noted in Section 3, the observed profile information $\tilde{j}(\theta_1)$ for θ_1 in general depends on the observation (t_1, t_2) only through t_2. Thus we have the following characterisation in statistical terms of the concept of τ-parallelism.

COROLLARY 5.2. *Consider the exponential model* $\mathscr{M}$ *with representation* (1.1) *and suppose* $t(X) = \mathrm{int}C$, *where* $t(X)$ *denotes the range of* t. *Then* $\mathscr{M}$ *has a* τ*-parallel foliation of the form considered in Theorem* 5.1 *if and only if the observed profile information* $\tilde{j}(\theta_1)$ *does not depend on* t_2 *(and hence not on* t*).* □

PROOF. Since for every $\tau_2 \in \mathscr{T}_2$ there exists a $x \in X$ such that $\tau_2 = t_2 = t_2(x)$, the statement that $\tilde{j}(\theta_1; t)$ does not depend on t is equivalent, according to Lemmas 3.3 and 3.4, to the statement that Δ_{11}^{-1} considered as a function of (θ_1, τ_2) does not depend on τ_2. An application of Theorem 5.2 completes the proof. □

The function $\ell^{\dagger}$, defined by

$$\ell^{\dagger}(\tau; \hat{\theta}) = \tau\cdot\hat{\theta} - \hat{\ell}(\tau)$$

can, because of (3.5) and (3.9), be considered as the dual of the log-likelihood function ℓ, and a characterisation of θ-parallelism similar to that of τ-parallelism in Corollary 5.2 can be given in terms of $\ell^{\dagger}$. If $\tilde{j}^{\dagger}(\tau_2; \hat{\theta})$ denotes the observed profile "information" on τ_2 calculated from $\ell^{\dagger}$, it can be shown that $\tilde{j}^{\dagger}(\tau_2; \hat{\theta}) = \Sigma_{22}^{-1}(\hat{\theta}_1, \tau_2)$ and the dual version of Corollary 5.2 now follows from Theorem 4.2.

Before we discuss the relation between independence of $\hat{\theta}_1$ and $\hat{\tau}_2$ and τ-parallelism, we give in Theorem 5.3 three conditions which separately are equivalent to the former condition provided one has τ-parallelism.

For a sample $x_1, \cdots, x_n$, $n \geq 1$, of independent observations from the distribution (1.1) we denote the corresponding model by $\mathscr{M}_n$. A minimal sufficient statistic of $\mathscr{M}_n$ is $\bar{t} = (\bar{t}_1, \bar{t}_2) = ((1/n)\sum_i t_1(x_i), (1/n)\sum_i t_2(x_i))$ and the corresponding canonical parameter is $\theta_n = n\theta$. It follows from (5.2) that if $\mathscr{M}$ has a τ-parallel foliation then so does $\mathscr{M}_n$ and, with obvious notation, one has $H_n = H$, $m_n = m(n^{-1}\cdot)$ and $k_n = nk$.

THEOREM 5.3. *Assume that the exponential model* (1.1) *has a* τ*-parallel foliation of the form considered in Theorem* 5.1. *If, in addition, the maximum likelihood estimate* $\hat{\theta}$ *of* θ *exists with probability* 1 *(i.e.* $\bar{t} \in \mathrm{int}C$ *with probability* 1*) then the following four conditions are equivalent*

(i) $\hat{\theta}_1 \perp \hat{\tau}_2$

(ii) $\bar{t}_1 - H(\bar{t}_2) \perp \bar{t}_2$

(iii) $\hat{\tau}_2 = \bar{t}_2 \sim \mathrm{EM}((H(\bar{t}_2), \bar{t}_2); n\theta)$

(iv) $m(\hat{\theta}_1) = \bar{t}_1 - H(\bar{t}_2) \sim \mathrm{EM}(\bar{t}_1 - H(\bar{t}_2); n\theta_1)$.

Furthermore, if the conditions (i)–(iv) *are fulfilled then*

(v) *the distribution of the quantity*

$$q = H(\bar{t}_2) - H(\tau_2) + (\tau_2 - \bar{t}_2)h^*(\tau_2) \tag{5.6}$$

depends only on θ_1. □

PROOF. It follows immediately from (5.2) that m is a one-to-one function of θ_1 and hence, since the maximum likelihood estimate of θ satisfies $\tau(\hat{\theta}) = \bar{t}$ for $\bar{t} \in \mathrm{int}C$, the equivalence of (i) and (ii) is proved. The equivalence of these and (iii) and (iv) follow from Lemma 3.5.

To prove (v) let

$$\bar{p} = \bar{t}_1 - \bar{t}_2 h^*(\tau_2) + \check{H}(\tau_2).$$

Since $\hat{\theta}_1$ and q are independent (due to (i)) and since

$$\bar{p} - m(\hat{\theta}_1) = (\bar{t}_1 - \bar{t}_2 h^*(\tau_2) + \check{H}(\tau_2)) - (\bar{t}_1 - H(\bar{t}_2)) = q,$$

it follows from Corollary 5.1 that

$$\begin{aligned} E_\theta\{\exp(\lambda\cdot q)\} &= (E_\theta\{\exp(n^{-1}\lambda\cdot p)\})^n / E_\theta\{\exp(\lambda\cdot m(\hat{\theta}_1))\} \\ &= \exp\{n(M(\theta_1 + n^{-1}\lambda) - M(\theta_1))\}/E_\theta\{\exp(\lambda\cdot m(\hat{\theta}_1))\} \end{aligned}$$

and using (iv) the proof of (v) is completed. □

To prove that the independence of $\hat{\theta}_1$ and $\hat{\tau}_2$ implies that the model considered has a τ-parallel foliation, we need the following lemma.

Lemma 5.1. *Consider the exponential model $\mathscr{M}$ with representation* (1.1). *Suppose there exists two stochastically independent statistics s_1 and s_2 such that the correspondence between t and (s_1, s_2) is one-to-one, and such that neither s_1 nor s_2 is a sufficient statistic. Then $\mathscr{M}$ has an exponential representation of the form*

$$a_1(\tilde{\theta}_1)a_2(\tilde{\theta}_2)\exp(\tilde{\theta}_1\cdot\{\tilde{f}(s_1(x)) + \tilde{g}_1(s_2(x))\} + \tilde{\theta}_2\cdot\tilde{g}_2(s_2(x))), \quad x \in X,$$

where

(i) *$\tilde{f}$ is one-to-one considered as a function of s_1,*

(ii) *$\tilde{g} = (\tilde{g}_1, \tilde{g}_2)$ is one-to-one considered as a function of s_2*

and

(iii) *$\tilde{\theta}_1$ parametrises the class of marginal distributions of s_1, i.e. the marginal distribution of s_1 depends on $\tilde{\theta}$ through $\tilde{\theta}_1$ alone, and different values of $\tilde{\theta}_1$ give different distributions of s_1.* □

Proof. Without loss of generality we assume that $0 \in \Theta$. Using (3.8) we obtain that

$$p(s_1, s_2; \theta) = \frac{d(s_1, s_2)P_\theta}{d(s_1, s_2)P_0} = z(s_1; \theta)w(s_2; \theta)$$

for some functions z and w. It follows that

$$\frac{p(s_1, s_2; \theta)}{p(s_{01}, s_2; \theta)} = \frac{p(s_1, s_{02}; \theta)}{p(s_{01}, s_{02}; \theta)}, \quad \theta \in \Theta, \tag{5.7}$$

for all s_1, s_{01}, s_2 and s_{02}. Simultaneous application of (5.7) to a set of linearly independent values $\theta_1, \cdots, \theta_k$ of θ implies that

$$\begin{aligned} t(s_1, s_2) &= t(s_1, s_{02}) + t(s_{01}, s_2) + t(s_{01}, s_{02}) \\ &= f(s_1) + g(s_2). \end{aligned} \tag{5.8}$$

Let d denote the dimension of the smallest affine subspace containing $f(s_1(X))$. Since neither s_1 nor s_2 is minimal sufficient, it follows from (5.8) that $0 < d < k$.

Now, let $v_0, v_1, \cdots, v_k$ be linearly independent vectors such that $\{v_1, \cdots, v_k\}$ is a basis of R^k and such that

$$f(s_1(X)) \subseteq v_0 + \operatorname{span}\{v_1, \cdots, v_d\}.$$

If $v_0 = \sum_{i=1}^k c_i v_i$ one has, with obvious notation,

$$
\begin{aligned}
\theta \cdot t &= \theta \cdot (f(s_1) + g(s_2)) \\
&= \theta \cdot \left(\textstyle\sum_{i=1}^{d} (f_i(s_1) + g_i(s_2) + c_i)v_i + \sum_{i=d+1}^{k} (g_i(s_2) + c_i)v_i\right) \\
&= \tilde{\theta}_1 \cdot \{\tilde{f}(s_1) + \tilde{g}_1(s_2)\} + \tilde{\theta}_2 \cdot \tilde{g}_2(s_2),
\end{aligned} \tag{5.9}
$$

where

$$\tilde{\theta}_1 = \theta[v_1^*, \cdots, v_d^*], \quad \tilde{\theta}_2 = \theta[v_{d+1}^*, \cdots, v_k^*],$$

$$\tilde{f}(s_1) = (f_1(s_1), \cdots, f_d(s_1)),$$

$$\tilde{g}_1(s_2) = (g_1(s_2) + c_1, \cdots, g_d(s_2) + c_d)$$

and

$$\tilde{g}_2(s_2) = (g_{d+1}(s_2) + c_{d+1}, \cdots, g_k(s_2) + c_k).$$

Using (5.9), the proof of Lemma 5.1 is easily completed. □

THEOREM 5.4. *Consider the exponential model $\mathscr{M}$ with representation* (1.1). *If $t(X)$ = int$C$ and $\hat{\theta}_1$ and $\hat{\tau}_2$ are stochastically independent then $\mathscr{M}$ has a τ-parallel foliation, of the form discussed in Theorem* 5.1. □

PROOF. From Lemma 5.1 with $(s_1, s_2) = (\hat{\theta}_1, \hat{\tau}_2) = (\hat{\theta}_1, t_2)$, it follows that $\mathscr{M}$ has representation

$$a_1(\tilde{\theta}_1)a_2(\tilde{\theta}_2)\exp(\tilde{\theta}_1 \cdot \tilde{t}_1 + \tilde{\theta}_2 \cdot \tilde{t}_2),$$

where

$$\tilde{t}_1 = \tilde{f}(\hat{\theta}_1) + g_1(t_2), \quad \tilde{t}_2 = \tilde{g}_2(t_2).$$

According to Lemma 8.1 in Barndorff-Nielsen (1978), there exists a $k \times k$ matrix A and a $1 \times k$ vector B such that

$$t = \tilde{t}A + B.$$

Letting

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

be a partition of A such that A_{11} is a $d \times k_1$ matrix, and writing $B = (B_1, B_2)$ where B_1 is a $1 \times k_1$ vector, one finds

$$t_1 = (\tilde{f}(\hat{\theta}_1) + g_1(t_2))A_{11} + g_2(t_2)A_{21} + B_1 = m(\hat{\theta}_1) + H(t_2).$$

The assumption $t(X)$ = int$C$ now implies that

$$\tau_1 = m(\theta_1) + H(\tau_2)$$

which, according to Theorem 5.1, completes the proof. □

Using Theorems 5.3 and 5.4, we obtain the following corollary.

COROLLARY 5.3. *Consider the exponential model $\mathscr{M}$ with representation* (1.1). *Under the assumption $t(X)$ = int$C$, stochastic independence of $\hat{\theta}_1$ and $\hat{\tau}_2$ implies that $\hat{\tau}_2 = t_2 \sim EM((H(t_2), t_2); \theta)$ for some k_1-dimensional function H.* □

In the discussion of whether τ-parallelism implies independence of $\hat{\theta}_1$ and $\hat{\tau}_2$, we take formula (5.1) as a basis and consider in succession the three cases corresponding, respectively, to $h(\tau_2)$ constant, $k(\tau_2)$ constant, and both $h(\tau_2)$ and $k(\tau_2)$ non-constant.

Constancy of $h(\tau_2)$ occurs if and only if the τ-parallel foliation given by (5.1) is also θ-parallel, as follows from Theorem 4.1 (iii). In this case we have:

THEOREM 5.5. *If the τ-parallel foliation given by* (5.1) *is also a θ-parallel foliation then h is constant,*

$$t_1 - t_2 h^* \perp t_2,$$

and the distribution of $t_1 - t_2 h^*$ depends on θ_1 only while the distribution of t_2 depends on $\theta_2 + \theta_1 h$ only. □

PROOF. In order to prove the independence it suffices, cf. Section 9.2 in Barndorff-Nielsen (1978), to prove that $t_1 - t_2 h^*$ and t_2 are uncorrelated. From (3.2) and (5.1) it follows that

$$\operatorname{cov}(t_1 - t_2 h^*, t_2) = \operatorname{cov}(t_1 - t_2 \Sigma_{22}^{-1} \Sigma_{21}, t_2) = \Sigma_{12} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22} = 0.$$

The proof is completed by noting that

$$\theta \cdot t = \theta_1 \cdot (t_1 - t_2 h^*) + (\theta_2 + \theta_1 h) \cdot t_2. \qquad \square$$

If $k(\tau_2)$ is constant, we may without loss of generality assume that $k(\tau_2) = 0$. For this case, which is characterised by the following theorem, we have

$$\Sigma_{22}^{-1} = -\theta_1 \times \frac{\partial h}{\partial \tau_2^*}$$

as is seen from (3.2) and (5.1).

THEOREM 5.6. *Suppose $\mathscr{M}$ has a τ-parallel foliation of the form discussed in Theorem* 5.1, *i.e.* $\theta_2(\theta_1, \tau_2) = -\theta_1 h(\tau_2) + k(\tau_2)$. *Furthermore, suppose $c \operatorname{int}\Theta \subseteq \operatorname{int}\Theta$ for every scalar $c > 1$ and let $D(\theta)$ denote the distribution of t_2. Then the following two statements are equivalent.*

(i) $k(\tau_2) = 0$ *for every* $\tau_2 \in \mathscr{T}_2$

(ii) *The distribution of* $\bar{t}_2$ *is* $D(n\theta)$.

Furthermore, if (i) *is fulfilled the function h is one-to-one.* □

PROOF. Suppose (i) holds, i.e.

$$\theta_2(\theta_1, \tau_2) = -\theta_1 h(\tau_2). \tag{5.10}$$

Since θ_1 and τ_2 are variation independent (cf. Lemma 3.1), formula (5.10) implies that

$$\begin{aligned} h(\tau_2) = h(\tilde{\tau}_2) &\Rightarrow \theta_2(\theta_1, \tau_2) = \theta_2(\theta_1, \tilde{\tau}_2) \\ &\Rightarrow (\theta_1, \theta_2(\theta_1, \tau_2)) = (\theta_1, \theta_2(\theta_1, \tilde{\tau}_2)) \\ &\Rightarrow (\theta_1, \tau_2) = (\theta_1, \tilde{\tau}_2) \Rightarrow \tau_2 = \tilde{\tau}_2 \end{aligned}$$

and consequently h is one-to-one.

The formula

$$\tau_2(c\theta) = \tau_2(\theta), \quad c > 1, \tag{5.11}$$

is the basis for the proof of the implication (i) $\Rightarrow$ (ii). To prove this formula, let $\tilde{\tau}_2 = \tau_2(c\theta)$ and $\tau_2 = \tau_2(\theta)$. Using (5.10) we have

$$c\theta_2 = -c\theta_1 h(\tilde{\tau}_2) \Rightarrow \theta_2 = -\theta_1 h(\tilde{\tau}_2) \Rightarrow (\theta_1, \theta_2(\theta_1, \tilde{\tau}_2)) = (\theta_1, \theta_2(\theta_1, \tau_2)) \Rightarrow \tau_2 = \tilde{\tau}_2.$$

From (5.5) it follows that the cumulant transform of t_2 is

$$\begin{aligned} \kappa_{t_2}(\lambda; \theta) &= \kappa(\theta_1, \theta_2 + \lambda) - \kappa(\theta_1, \theta_2) \\ &= -\theta_1 \cdot \{\check{H}(\tau_2(\theta_1, \theta_2 + \lambda)) - \check{H}(\tau_2(\theta_1, \theta_2))\}. \end{aligned}$$

The cumulant transform of $\bar{t}_2$ is, using (5.11),

$$\begin{aligned}\kappa_{\bar{t}_2}(\lambda;\theta) &= n\kappa_{t_2}(n^{-1}\lambda;\theta)\\ &= -n\theta_1\cdot\{\check{H}(\tau_2(\theta_1,\theta_2+n^{-1}\lambda)) - \check{H}(\tau_2(\theta_1,\theta_2))\}\\ &= -n\theta_1\cdot\{\check{H}(\tau_2(n\theta_1,n\theta_2+\lambda)) - \check{H}(\tau_2(n\theta_1,n\theta_2))\}\\ &= \kappa_{t_2}(\lambda;n\theta)\end{aligned}$$

and so $\bar{t}_2$ has distribution $D(n\theta)$.

To prove the converse implication, note that (ii) implies that

$$\tau_2(n\theta) = \tau_2(\theta) = \tau_2.$$

Inserting this into (5.1) we find

$$\theta_2 = -\theta_1 h(\tau_2) + k(\tau_2)$$

and

$$n\theta_2 = -n\theta_1 h(\tau_2) + k(\tau_2).$$

Consequently, $k(\tau_2) = 0$ and the proof of Theorem 5.6 is completed. □

We conclude the discussion of exponential models with τ-parallel foliation for which the function k is constant by considering the situation where the component $t_2(x)$ of the minimal canonical statistic $t(x)$ is in one-to-one correspondence with x. It then causes no loss of generality to assume that $t_2(x) = x$. Collecting results from above we have:

COROLLARY 5.4. *Suppose $t_2(x) = x$, i.e. the exponential model $\mathcal{M}$ is of the form*

$$\frac{dP_\theta}{d\mu}(x) = a(\theta)b(x)\exp(\theta_1\cdot u(x) + \theta_2\cdot x),$$

and assume moreover that

$$\theta_2(\theta_1,\tau_2) = -\theta_1 h(\tau_2).$$

If, in addition, $\bar{t}\in \operatorname{int}C$ with probability 1, $c\operatorname{int}\Theta\subseteq\operatorname{int}\Theta$ for every scalar $c>1$, and u is continuous then one has

(i) $\bar{x}\sim P_{n\theta}$

(ii) $h(\tau_2) = \partial u^*(\tau_2)/\partial\tau_2$

(iii) $\hat{\theta}_1\perp\hat{\tau}_2$, *or equivalently*, $\bar{u}-u(\bar{x})\perp\bar{x}$

(iv) $m(\hat{\theta}_1) = \bar{u}-u(\bar{x})\sim EM(\bar{u}-u(\bar{x});n\theta_1)$

and the Laplace transform of $\bar{u}-u(\bar{x}) = m(\hat{\theta}_1)$ is

$$\begin{aligned}&E_{\theta_1}\{\exp(\lambda\cdot m(\hat{\theta}_1))\}\\ &\quad = \exp(-\{M(n\theta_1+\lambda)-M(n\theta_1)\}+n\{M(\theta_1+n^{-1}\lambda)-M(\theta_1)\})\end{aligned}\tag{5.12}$$

for $\theta_1\in\operatorname{int}\Theta_1$.

(v) *The distribution of the quantity*

$$q = u(\bar{x}) - u(\tau_2) + (\tau_2-\bar{x})h^*(\tau_2)$$

depends only on θ_1. □

PROOF. Theorem 5.6 implies that (i) is true. From (i) it follows, using Lemma 3.5, that $\bar{u}-u(\bar{x})\perp\bar{x}$ and that $\bar{u}-u(\bar{x})\sim EM(\bar{u}-u(\bar{x});n\theta_1)$. Using the continuity of u one has

$$\bar{u}-u(\bar{x})\to\tau_1-u(\tau_2) = m(\theta_1)+H(\tau_2)-u(\tau_2)\quad\text{a.s.}$$

as $n \to \infty$. Since the distribution of $\bar{u} - u(\bar{x})$ depends on θ_1 only $H(\tau_2) - u(\tau_2)$ must be constant. Assertion (ii) is now immediate and (iii), (iv) and (v) follow from Theorem 5.3.

It remains to prove formula (5.12). From (i) and (iii) in conjunction with Lemma 3.5 we find

$$E_\theta\{\exp(\lambda \cdot m(\hat{\theta}_1))\} = \frac{a(n\theta + (\lambda, 0))}{a(n\theta)} \frac{a(\theta)^n}{a(\theta + (n^{-1}\lambda, 0))^n}. \tag{5.13}$$

According to (5.5) we have

$$a(\theta) = \exp(\theta_1 \cdot \check{H}(\tau_2) - M(\theta_1))$$

and inserting this in (5.13) and using (5.11) we obtain (5.12). □

Finally, we present an example where neither $h(\tau_2)$ nor $k(\tau_2)$ of (5.1) is constant. For this case we have not been able to establish the independence of $\hat{\theta}_1$ and $\hat{\tau}_2$ in general. However, we do not know of any counterexamples. In special cases it is often easy to establish the independence as in:

EXAMPLE 5.1. Let u and v be positive variates such that the distribution of u is the inverse Gaussian distribution $N^-(\chi, \psi)$ with probability density function

$$\frac{\sqrt{\chi}}{\sqrt{2\pi}} \exp(\sqrt{\chi\psi}) u^{-3/2} \exp(-(\chi u^{-1} + \psi u)/2)$$

and such that the distribution of v given u is the Γ-distribution with probability density function

$$\frac{1}{\Gamma(\beta u)} \nu^{\beta u} v^{\beta u - 1} \exp(-\nu v).$$

It follows that the distribution of (u, v) has probability density function

$$\frac{\sqrt{\chi}}{\sqrt{2\pi}} \exp(\sqrt{\chi\psi}) u^{-3/2} v^{\beta u - 1} \frac{1}{\Gamma(\beta u)} \beta^{\beta u} \exp(-\tfrac{1}{2}\chi u^{-1} - \alpha u - \nu v) \tag{5.14}$$

where

$$\alpha = \tfrac{1}{2}\psi - \beta \ln(\nu/\beta).$$

For β fixed (5.14) is of the form (1.1) with

$$t(u, v) = (u^{-1}, u, v)$$

and

$$\theta = (-\tfrac{1}{2}\chi, -\alpha, -\nu).$$

Since

$$\tau = (1/\chi + \sqrt{\psi/\chi}, \sqrt{\chi/\psi}, \sqrt{\chi/\psi}\, \beta/\nu)$$

one has

$$\tau_1 = m(\theta_1) + H(\tau_{21}, \tau_{22}),$$

where

$$m(\theta_1) = -\tfrac{1}{2}\theta_1^{-1}$$

and

$$H(\tau_{21}, \tau_{22}) = \tau_{21}^{-1}.$$

This shows that the family of distributions of (u, v) for fixed β has a τ-parallel foliation.

In the present case the function k of (5.1) is not constant. In fact,

$$k(\tau_{21}, \tau_{22}) = \theta_1 h(\tau_{21}, \tau_{22}) + (\theta_{21}, \theta_{22}) = \theta_1(-\tau_{21}^{-2}, 0) + (\theta_{21}, \theta_{22})$$
$$= (\beta \ln(\tau_{21}/\tau_{22}), -\beta\tau_{21}/\tau_{22}).$$

However, the following argument shows that $\hat{\theta}_1$ and $\hat{\tau}_2$ are independent in this case. From (5.14) one finds the cumulant transform for (u, v) to be

$$\kappa_{(u,v)}(\zeta, \eta) = \sqrt{2\chi(\alpha + \beta \ln(\nu/\beta))} - \sqrt{2\chi(\alpha - \zeta + \beta \ln((\nu - \eta)/\beta))}.$$

If $[N^-, \Gamma](\chi, \alpha, \nu; \beta)$ denotes the distribution (5.14) it follows that the distribution of $(\bar{u}, \bar{v})$, corresponding to n independent observations, is $[N^-, \Gamma](n\chi, n\alpha, n\nu; n\beta)$ which implies that $(\bar{u}, \bar{v}) \sim EM((\bar{u}^{-1}, \bar{u}, \bar{v}); n\theta)$. The independence now follows from Theorem 5.3. □

For further examples illustrating the theory in this section, including the normal distribution, the inverse Gaussian distribution, the gamma distribution and several distributions obtained by combining these three distributions, we refer to Bar-Lev and Reiser (1982) and Barndorff-Nielsen and Blæsild (1983). Still other examples are provided by the Wishart distribution, the p-dimensional normal distribution and a large class of submodels of this distribution including, for instance, the model for multivariate two-way analysis of variance.

Acknowledgment. We are indebted to A. Holst Andersen for a critical reading of the manuscript.

REFERENCES

Bar-Lev, S. K. (1983). A characterization of certain statistics in exponential models whose distributions depend on a sub-vector of parameters only. *Ann. Statist.* **11** 746–752.

Bar-Lev, S. K. and Reiser, B. (1982). An exponential subfamily which admits UMPU tests based on a single test statistic. *Ann. Statist.* **10** 979–989.

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

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

Courant, R. and Hilbert, D. (1952). *Methods of Mathematical Physics.* Vol. I. Interscience, New York.

Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics.* Vol. II. Interscience, New York.

Kamke, E. (1930). *Differentialgleichungen Reeller Funktionen.* Akademische Verlagsgesellschaft, Leipzig.

Kamke, E. (1974). *Differentialgleichungen. Lösungsmetoden und Lösungen.* Chelsea, Bronx.

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

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

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press, New Jersey

Afdeling for Teoretisk Statistik
Matematisk Institut
Aarhus Universitet
DK-8000 Aarhus C
Denmark