Ann. Inst. Statist. Math.
Vol. 41, No. 2, 247-267 (1989)

APPROXIMATING EXPONENTIAL MODELS*

O. E. BARNDORFF-NIELSEN[1] AND P. E. JUPP[2]

[1]*Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark*
[2]*Department of Mathematical Sciences, University of St. Andrews, North Haugh, St. Andrews, KY16 9SS, U.K.*

(Received February 23, 1988)

Abstract. Approximation of parametric statistical models by exponential models is discussed, from the viewpoints of observed as well as of expected likelihood geometry. This extends a construction, in expected geometry, due to Amari. The approximations considered are parametrization invariant and local. Some of them relate to conditional models given exact or approximate ancillary statistics. Various examples are considered and the relation between the maximum likelihood estimators of the original model and the approximating models is studied.

Key words and phrases: Expected likelihood geometry, observed likelihood geometry, vector bundles, conditional inference, maximum likelihood estimation.

1. Introduction

The wide occurrence and usefulness of exponential models makes it natural to enquire to what extent and in what ways it is feasible and profitable to construct exponential models that approximate a given statistical model. This question seems, so far, to have been considered only from the viewpoint of approximations locally around a single member of the model, and this is also the viewpoint taken in this paper, though the global viewpoint appears to be at least as important.

Local approximation by normal models with essentially known variance matrices is, of course, a standard theme in statistics. Of some particular interest in connection with the setting of the present paper is the concept of local asymptotic normality (see for instance, Ibragimov and Has'minskii (1981) and Le Cam (1986)). Efron (1975) briefly pointed to the possibility of using higher order approximations by simply Taylor expanding the log model function around a parameter point. This approach

*Research partly supported by the Danish Science Research Council.

seems, however, too primitive to be of much use. A more promising and original approach was taken by Amari (1987) who relied on concepts and intuition of a differential geometric nature. The purpose of the present paper is partly to show that Amari's construction, which belongs to expected likelihood geometry (in the sense of Barndorff-Nielsen (1986)), has a counterpart in observed likelihood geometry, and partly to investigate properties of these constructions. (It is not immediate that an observed counterpart must exist, because the Hilbert bundle that Amari used for his construction has no observed analogue, due to the lack of an "observed inner product" of random variables.)

One of the potential uses of approximating exponential models is in the construction of approximate ancillaries, a theme that we do not pursue here, however.

Section 2 contains a few preliminaries. Section 3 describes Amari's definition of locally approximating exponential models and provides some elementary properties and several examples of this type of approximation, which we refer to as *expected exponential approximations*. The examples concern the von Mises distribution, nonlinear normal regression, and a simple model from quantum physics which is of some particular interest because it is neither an exponential model nor a transformation model. In Section 4 we then address the problem of carrying out a construction similar to Amari's but in the framework of observed, rather than expected, likelihood geometry. Elementary properties of the resulting *observed exponential approximations* are discussed. Section 5 is devoted to a more detailed discussion of the degree to which the expected or observed exponential approximations are close, locally, to the actual model, and to the closeness of the maximum likelihood estimators derived from these three types of model.

2. Preliminaries

In this section we introduce some notation and discuss in considerable generality the idea of projection onto a subspace.

We shall often consider functions $f: \Omega \to \mathbb{R}$ where Ω is an open set in $\mathbb{R}^d$ or, more generally, a d-dimensional manifold. Let $\omega_1,\dots,\omega_d$ denote a system of local coordinates on Ω. Then we put $f_i = \partial f/\partial\omega_i$. The following condensed notation will also be useful. For a *multi-index* $\alpha = (i_1, i_2,\dots, i_d)$ with each i_j a non-negative integer we put

$$|\alpha| = i_1 + i_2 + \cdots + i_d$$

and

$$f_\alpha = \frac{\partial^{|\alpha|} f}{\partial \omega_1^{i_1} \cdots \partial \omega_d^{i_d}} .$$

Similarly, given a function $l: \Omega \times \Omega \to \mathbb{R}$ expressed as $l(\omega_1, \ldots, \omega_d; \hat{\omega}_1, \ldots, \hat{\omega}_d)$, we put

$$l_{i;jk} = \frac{\partial^3 l}{\partial \omega_i \partial \hat{\omega}_j \partial \hat{\omega}_k} , \qquad \text{etc.}$$

and for multi-indices $\alpha = (i_1, \ldots, i_d)$ and $\beta = (j_1, \ldots, j_d)$ we put

$$l_{\alpha;\beta}(\omega; \hat{\omega}) = \frac{\partial^{|\alpha| + |\beta|} l(\omega, \hat{\omega})}{\partial \omega_1^{i_1} \cdots \partial \omega_d^{i_d} \partial \hat{\omega}_1^{j_1} \cdots \partial \hat{\omega}_d^{j_d}} .$$

Also we write

$$\not{l}_{i;jk}(\omega) = l_{i;jk}(\omega; \omega) , \qquad \text{etc.}$$

and

$$\not{l}_{\alpha;\beta}(\omega) = l_{\alpha;\beta}(\omega; \omega) .$$

Some further notational conventions are that for a multi-index $\alpha = (i_1, \ldots, i_d)$, we define $\alpha! = i_1! \, i_2! \cdots i_d!$ and $\omega^\alpha = \omega_1^{i_1} \cdots \omega_d^{i_d}$, and $\alpha \le \beta$ means that $i_r \le j_r$ for $r = 1, \ldots, d$ where $\beta = (j_1, \ldots, j_d)$.

As we shall make use of general projections of vector spaces onto subspaces we now recall this operation. Let V and W be vector spaces, $A: W \to V$ a linear mapping and $\varphi: V \times W \to \mathbb{R}$ a bilinear form. Then φ gives rise to a linear mapping $\vec{\varphi}: V \to W^*$, where W^* denotes the dual of W, given by

$$\vec{\varphi}(v)(w) = \varphi(v, w) .$$

If the composite $\vec{\varphi} \circ A: W \to V \to W^*$ is an isomorphism, then the *projection of V onto W* is defined by

$$\pi = (\vec{\varphi} \circ A)^{-1} \circ \vec{\varphi}: \qquad V \to W^* \to W .$$

The projection has the property $\pi \circ A = I: W \to W$. To obtain an expression for π in terms of coordinates, let $\{e_1, \ldots, e_p\}$ and $\{f_1, \ldots, f_q\}$ be bases of W and V, respectively. Let the matrices B and M have entries $b_{ij} = \varphi(f_i, e_j)$ and $m_{ij} = \varphi(Ae_i, e_j)$. Then the matrix of π with respect to these bases is $M^{-1}B$. If W is a subspace of V, φ is an inner product on V, and A is the inclusion of W in V, then π is just the orthogonal projection of V onto W.

3. Expected exponential approximations

In the present section we describe Amari's (1987) definition of approximating exponential models and we then go on to investigate some of its properties.

Let $\mathcal{M} = (\mathcal{X}, p, \Omega)$ be an arbitrary parametric statistical model, the parameter space Ω being a d-dimensional differentiable manifold (as in Section 2). We shall assume that the model function p has any requisite smoothness as a function on Ω. For a multi-index α the α-th derivative of the log-likelihood function at ω will be denoted by $l_\alpha(\omega)$. In view of the pleasant properties of exponential families, it is natural to consider approximating $\mathcal{M}$ (at least on some small portion of Ω) by a full or curved exponential family. For ω in Ω and for $r = 1, 2, \ldots$, Amari (1987) first constructs a full exponential model $\overline{E^r\mathcal{M}_\omega}$, the *r-th order full exponential approximation to $\mathcal{M}$ at ω*. This model $\overline{E^r\mathcal{M}_\omega}$ is the full exponential family generated by the element of $\mathcal{M}$ given by ω and by the log-likelihood derivatives $l_\alpha(\omega)$ for $1 \le |\alpha| \le r$. Thus one can regard $\overline{E^r\mathcal{M}_\omega}$ as having "r-th order contact with $\mathcal{M}$ at ω". It is easiest to visualise the particular case when $\mathcal{M}$ is a $(k, 1)$-exponential model. Then, if Θ denotes the canonical parameter space of the corresponding full exponential model and $\mathcal{M}$ has canonical parameter function $\theta : \Omega \to \Theta$, the family $\overline{E^r\mathcal{M}_\omega}$ has canonical parameters in the affine subspace of Θ through $\theta(\omega)$ parallel to that spanned by the derivatives $(d^i\theta/d\omega^i)(\omega)$, $1 \le i \le r$.

Amari's construction does more than provide a full exponential family $\overline{E^r\mathcal{M}_\omega}$ which is "close to $\mathcal{M}$ at ω". It also gives a "canonical" curved submodel $E^r\mathcal{M}_\omega$ of this full model: the parametric dimension of $E^r\mathcal{M}_\omega$ is the same as that of $\mathcal{M}$ (i.e., dimension d) and $E^r\mathcal{M}_\omega$ is, in a certain sense, closest to $\mathcal{M}$ among all submodels of this kind.

We shall refer to $E^r\mathcal{M}_\omega$ as the *r-th order expected exponential approximation to $\mathcal{M}$ at ω*. The term "expected" here refers to the expectations in (3.1) below used in the definition of $E^r\mathcal{M}_\omega$ and serves to distinguish $E^r\mathcal{M}_\omega$ from the "observed" exponential approximations $O^r\mathcal{M}_\omega$ introduced in the next section. There is a strong connection between this terminology and that of expected and observed statistical geometries.

To describe the detailed construction of $E^r\mathcal{M}_\omega$ it is useful to introduce the following notation. Let $\mathcal{H}_\omega = \{f : \mathcal{X} \to \mathbb{R} \mid E_\omega(f) = 0, \mathrm{Var}_{\omega'}(f) < \infty, \forall\, \omega' \in \Omega\}$ and equip $\mathcal{H}_\omega$ with the inner product

$$\langle f, g \rangle = E_\omega[fg], \quad f, g \in \mathcal{H}_\omega, \tag{3.1}$$

thus giving $\mathcal{H}_\omega$ the structure of a Hilbert space. For any multi-index α, define v_α in $\mathcal{H}_\omega$ by

$$v_\alpha = l_\alpha(\omega) - E_\omega[l_\alpha(\omega)]\,.$$

Also, for $r = 1,2,\ldots$, define M as the square matrix with entries $M_{\alpha\beta} = \langle v_\alpha, v_\beta\rangle$ for $|\alpha|, |\beta| \leq r$. Assume that M is invertible and denote the entries of M^{-1} by $M^{\alpha\beta}$. Then the model function $p_\omega^{[r]}$ of $E^r\mathcal{M}_\omega$ is given by

$$(3.2) \qquad p_\omega^{[r]}(x;\omega') = p(x;\omega) \cdot \exp\{\langle l(\omega') - l(\omega) + I(\omega,\omega'), v_\alpha\rangle M^{\alpha\beta} v_\beta - \kappa(\omega')\}\,,$$

where the convention of summing over repeated indices is used, $I(\omega,\omega')$ denotes the discrimination information $E_\omega[l(\omega) - l(\omega')]$, and κ is the log norming constant. (Of course, κ depends also on the point ω at which we approximate, but we shall consider ω as fixed and so shall suppress this dependence.) The intuitive idea behind this construction is the following. Let $\mathcal{T}_\omega^{(r)}$ denote the subspace of $\mathcal{H}_\omega$ spanned by $\{v_\alpha : |\alpha| \leq r\}$. The normed log-density-ratio function $\omega' \mapsto l(\omega') - l(\omega) + I(\omega,\omega')$ maps Ω into $\mathcal{H}_\omega$. Combining this with orthogonal projection Π^e of $\mathcal{H}_\omega$ onto $\mathcal{T}_\omega^{(r)}$ we obtain a function

$$\omega' \mapsto \langle\{l(\omega') - l(\omega) + I(\omega,\omega')\}, v_\alpha\rangle M^{\alpha\beta} v_\beta\,,$$

from Ω into $\mathcal{T}_\omega^{(r)}$ and so a function $\omega' \mapsto p_\omega^{[r]}(\cdot\,;\omega')$ given by (3.2) of Ω into the exponential family $\overline{E^r\mathcal{M}_\omega}$. A more detailed discussion of this construction is given in the following Remark. The union $\bigcup_{\omega\in\Omega}\mathcal{H}_\omega$ forms the total space of a vector bundle $\mathcal{H}$ over Ω. This is Amari's (1987) Hilbert bundle.

Remark. Let μ be a measure on $\mathcal{X}$ dominating all members of the family given by $\mathcal{M}$. Denote by M_1 the set of measures ν on $\mathcal{X}$ which are absolutely continuous with respect to μ and with $\nu(\mathcal{X}) = 1$. Then

$$TM_1 = \{(\nu, f) \mid \nu \in M_1, f\colon \mathcal{X} \to \mathbb{R}, E_\nu(f) = 0, \operatorname{Var}_{\nu'}(f) < \infty,\ \forall\, \nu' \in M_1\}\,,$$

together with the projection $\Pi\colon TM_1 \to M_1$ given by $\Pi(\nu, f) = \nu$ can be regarded as in some sense the "tangent bundle" to M_1 (cf. Dawid (1975)). Note that this vector bundle has a metric given by

$$\langle f, g\rangle = E_\nu[fg], \qquad (\nu, f), (\nu, g) \in TM_1\,.$$

Further, Amari's "Hilbert bundle" $\mathcal{H}$ over Ω is precisely the pull-back of TM_1 over the inclusion $\Omega \to M_1$ and the metric on $\mathcal{H}$ is that pulled back from TM_1. For details of the pull-back (or induced bundle) construction of vector bundles, see Husemoller (1966), p. 18 and 26.

The log-likelihood function for $\mathcal{M}$ determines a section $s\colon \Omega \to \mathcal{H}$ of the Hilbert bundle by

$$s(\omega) = l(\omega) - E_\omega[l(\omega)] .$$

Amari's 1-connection on $\mathscr{H}$ has parallel translation operator $T^{\omega}_{\omega'}: \mathscr{H}_{\omega'} \to \mathscr{H}_\omega$ from the fibre of $\mathscr{H}$ over ω' to that over ω given by

$$T^{\omega}_{\omega'}(f) = f - E_\omega[f] .$$

Then, given $\omega \in \Omega$, we can define the function $\Psi_\omega: \Omega \to \mathscr{H}_\omega$ by

$$\Psi_\omega(\omega') = T^{\omega}_{\omega'}(s(\omega')) - s(\omega) \tag{3.3}$$
$$= l(\omega') - l(\omega) + I(\omega, \omega') .$$

By analogy with the differential-geometric concept of development of paths in a manifold as described, for example, on pp. 45–46 and 77–79 of Lichnerowicz (1976), we propose the name *Amari development* or *expected development* for the above function Ψ_ω which lifts the parameter space Ω into the Hilbert space $\mathscr{H}_\omega$.

Let $\Pi^e: \mathscr{H}_\omega \to \mathscr{T}^{(r)}_\omega$ be the orthogonal projection onto $\mathscr{T}^{(r)}_\omega$. Then $\Pi^e \circ \Psi_\omega: \Omega \to \mathscr{T}^{(r)}_\omega$ gives rise to the above-mentioned inclusion of $E^r\mathscr{M}_\omega$ as a curved submodel of $\overline{E^r\mathscr{M}_\omega}$, the full exponential family generated by $\{l_\alpha(\omega): |\alpha| \le r\}$.

We proceed to list some simple but important properties of the approximating construction.

First we consider the effect of repeated random sampling. Given a model $\mathscr{M} = (\mathscr{X}, p, \Omega)$, let $\mathscr{M}^n = (\mathscr{X}^n, p^n, \Omega)$ denote the model obtained from $\mathscr{M}$ by taking random samples of size n, so that

$$p^n(x_1, \ldots, x_n; \omega) = \prod_{i=1}^{n} p(x_i; \omega) .$$

Standard calculations then show that $E^r(\mathscr{M}^n)_\omega = (E^r\mathscr{M}_\omega)^n$, so that we can denote this approximating exponential model without ambiguity by $E^r\mathscr{M}^n_\omega$.

Next we consider two important classes of models: exponential models and transformation models.

Suppose that $\mathscr{M}$ is a (k, d)-exponential model with canonical parameter space Θ.

Let t and $\theta(\omega)$ denote, respectively, the canonical statistic and the value of the canonical parameter corresponding to ω in Ω. Then $E^r\mathscr{M}_\omega$ is obtained from $\mathscr{M}$ by orthogonal projection Π^Σ (given by the inner product on span (Θ) obtained from $\Sigma = \mathrm{Var}_\omega(t)$) onto the subspace through $\theta(\omega)$ parallel to that spanned by $\theta_\alpha(\omega)$, $|\alpha| \le r$. To see this, note that if $\iota: \Theta \to \mathscr{H}_\omega$ is given by

$$\iota(\theta) = l(\theta) - l(\theta(\omega)) + I(\theta(\omega), \theta),$$

then the diagram in Fig. 1 commutes. Here, Π^e again denotes the orthogonal projection given by the inner product (3.1). Note that the inner product given by Σ is the expected information at $\theta(\omega)$ of the full exponential model generated by $\mathcal{M}$.

$$\begin{tikzcd} \Theta \arrow[r, "\iota"] \arrow[d, "\Pi^{\Sigma}"'] & \mathcal{H}_\omega \arrow[d, "\Pi^e"] \\ \Theta \arrow[r, "\iota"'] & \mathcal{H}_\omega \end{tikzcd}$$

Fig. 1. Relationship of the orthogonal projections Π^{Σ} and Π^e.

It may further be noted that $\mathcal{M} = E^r\mathcal{M}_\omega$ if and only if $\mathcal{M}$ is a (k, d)-exponential model and $\dim \operatorname{span}\{l_\alpha(\omega): |\alpha| \le r\} = k$. This dimensional condition is usually satisfied whenever

$$\sum_{i=1}^{r} \binom{d+i-1}{i} \ge k.$$

If $\mathcal{M}$ is a (composite) transformation model with an invariant dominating measure, that is if there is a group G acting on $\mathcal{X}$ and Ω such that

$$p(g\omega; gx) = p(\omega; x),$$

then the action of g on $\mathcal{X}$ induces the following G-action on the set $\bigcup_{\omega \in \Omega} E^r\mathcal{M}_\omega$ of r-th order exponential approximations. The element g of G sends $p_\omega^{[r]}(\cdot\,; \omega')$ to $p_\omega^{[r]}(g^{-1}\cdot\,; \omega') = p_{g\omega}^{[r]}(\cdot\,; g\omega')$.

Example 3.1. (*von Mises distributions with known concentration*) Let $\mathcal{M}$ be the von Mises model with known concentration parameter $\kappa > 0$. Then both $\mathcal{X}$ and Ω are the circle S^1 and

$$p(x;\omega)=\{I_0(\kappa)\}^{-1}e^{\kappa\cos(x-\omega)}\,, \tag{3.4}$$

where the dominating measure μ is the uniform probability measure on S^1. As $\mathscr{M}$ is a $(2,1)$-exponential model, it follows from the general construction that $E^1\mathscr{M}_\omega$ is a $(1,1)$-exponential model, corresponding in the plane of canonical parameters of the $(2,2)$-model generated by $\mathscr{M}$ to the tangent line at ω to the circle representing $\mathscr{M}$. Calculation of $E^1\mathscr{M}_\omega$ shows that

$$\begin{aligned} p_\omega^{[1]}(x;\omega') &= \{I_0(\kappa\sqrt{1+(\sin\delta)^2})\}^{-1} \\ &\quad\cdot\exp\{\kappa[\cos(x-\omega)+\sin\delta\sin(x-\omega)]\}\,, \end{aligned} \tag{3.5}$$

where $\delta=\omega'-\omega$. Thus in $E^1\mathscr{M}_\omega$ the parameter ω' corresponds to the von Mises distribution with concentration $\kappa\sqrt{1+(\sin\delta)^2}$ and mean direction $\omega+\tan^{-1}(\sin\delta)$. The geometrical description of this approximation is that $E^1\mathscr{M}_\omega$ is obtained from the circle representing $\mathscr{M}$ by projection parallel to the line between 0 (corresponding to $\kappa=0$) and ω onto the tangent to $\mathscr{M}$ at ω (see Fig. 2).

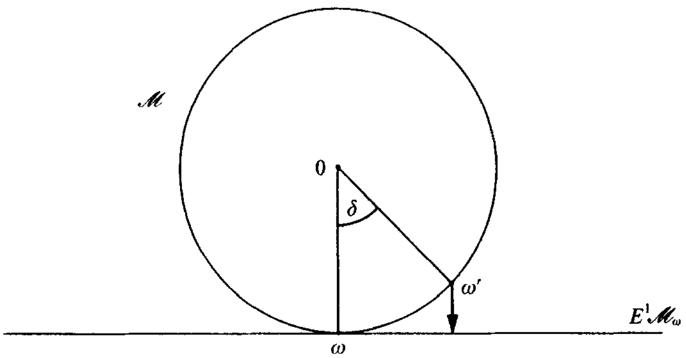

Fig. 2. Expected exponential approximation to von Mises model (3.4).

Example 3.2. (*Nonlinear regression*) Let $x_1,\dots,x_n$ be independent normal variates with a common variance σ^2 and with means $Ex_\nu=\xi_\nu(\omega)$, $\nu=1,\dots,n$. For simplicity we assume that σ^2 is known and that ω is one-dimensional. The r-th order expected exponential approximation with base point ω, i.e., $E^r\mathscr{M}_\omega$, may conveniently be described as follows. Letting $\lambda=\sigma^{-2}$ we find

$$\upsilon_\alpha=\lambda\sum_{\nu=1}^{n}(x_\nu-\xi_\nu)\xi_{\nu\alpha}\,,$$

$$l(\omega')-l(\omega)+I(\omega,\omega')=\lambda\sum_{\nu=1}^{n}(x_\nu-\xi_\nu)(\xi_\nu'-\xi_\nu)$$

and

$$M_{\alpha\beta} = \lambda \Xi_{\alpha\beta} ,$$

where $\xi_\nu = \xi_\nu(\omega)$, $\xi_{\nu\alpha} = \xi_{\nu\alpha}(\omega)$, $\xi'_\nu = \xi_\nu(\omega')$, and

$$\Xi_{\alpha\beta} = \sum_{\nu=1}^{n} \xi_{\nu\alpha}\xi_{\nu\beta} ,$$

$\xi_{\nu\alpha}$ denoting the α-th order derivative of ξ_ν. Now, for $\alpha = 1,\dots,r$, let

$$y_\alpha(\omega) = \sum_{\nu=1}^{n} (x_\nu - \xi_\nu(\omega))\xi_{\nu\alpha}(\omega) .$$

Then, under $E^r\mathscr{M}_\omega$, the vector $[y_\alpha(\omega)]$ is a sufficient statistic and

$$[y_\alpha(\omega)] \sim N_r(\eta_\alpha(\omega,\omega'), \sigma^2\Xi(\omega)) , \tag{3.6}$$

where

$$\eta_\alpha(\omega,\omega') = \sum_{\nu=1}^{n} \{\xi_\nu(\omega') - \xi_\nu(\omega)\}\xi_{\nu\alpha}(\omega) .$$

Thus, $E^r\mathscr{M}_\omega$ is again a nonlinear regression model, specified by (3.6).

Example 3.3. We shall consider a model from quantum physics discussed earlier, in the statistical literature, by Barnard (1971, 1982), Sprott (1977) and DiCiccio (1984) (see also Solmitz (1964)). The measurements concerned are of the cosine of the scattering angle for decaying Λ particles, and the model function for the cosine is

$$p(x;\omega) = \frac{1}{2}(1 - \omega x) , \tag{3.7}$$

the observed cosine x and the parameter ω (termed the parity non-conservation parameter) both varying in the interval $(-1, 1)$.

Note that (3.7), which we shall refer to as the Λ model, is neither an exponential model nor a transformation model.

With $l(\omega) = \log(1 - \omega x)$ we have

$$\dot{l}(\omega) = -x(1 - \omega x)^{-1} ,$$
$$\ddot{l}(\omega) = -x^2(1 - \omega x)^{-2} ,$$

from which we find

$$i(\omega) = \omega^{-3}\left[\frac{1}{2}\log\frac{1+\omega}{1-\omega} - \omega\right],$$

with similar, but more complicated, expressions for $E\{\dot{l}(\omega)\ddot{l}(\omega)\}$ and $E\{\ddot{l}^2\}$. Furthermore

$$\begin{aligned} &E_\omega[\dot{l}(\omega)\{l(\omega') - l(\omega)\}] \\ &\quad = \frac{1}{4}\left\{\log\frac{1+\omega'}{1-\omega'} - 2\omega' i(\omega') - \log\frac{1+\omega}{1-\omega} + 2\omega i(\omega)\right\}, \end{aligned}$$

whereas a fully explicit expression for

$$E_\omega[\ddot{l}(\omega)\{l(\omega') - l(\omega)\}] \tag{3.8}$$

does not exist in general.

However, for $\omega = 0$, (3.8) can be evaluated and this case is of some particular interest. In the early development interest centered on whether ω was, in fact, 0 and while this has now been firmly established not to be the case, the available evidence indicates that the actual value of ω is close to 0. We will therefore consider exponential approximations to (3.7) based at $\omega = 0$.

The second order expected exponential approximation turns out to be

$$p_0^{[2]}(x;\omega) = a(\theta(\omega))e^{-\theta_1(\omega)x - \theta_2(\omega)x^2}, \tag{3.9}$$

where

$$\theta_1(\omega) = \frac{3}{2}\omega^{-1} + \frac{3}{4}(1-\omega^{-2})\log\frac{1+\omega}{1-\omega},$$

$$\theta_2(\omega) = -\frac{5}{2} + \frac{15}{4}\omega^{-2} + \frac{15}{8}(\omega^{-1} - \omega^{-3})\log\frac{1+\omega}{1-\omega},$$

and

$$\begin{aligned} a(\theta)^{-1} &= \int_{-1}^{1} e^{-\theta_1 x - \theta_2 x^2}dx \\ &= \sqrt{\pi/\theta_2}\,e^{\theta_1^2/2(2\theta_2)}\left\{\Phi\left(\sqrt{2\theta_2}\left(\frac{\theta_1}{2\theta_2}+1\right)\right) - \Phi\left(\sqrt{2\theta_2}\left(\frac{\theta_1}{2\theta_2}-1\right)\right)\right\}, \end{aligned}$$

Φ denoting the distribution function of the standard normal distribution. Thus, (3.9) is the truncation of a normal model to the interval $(-1, 1)$.

Near $\omega = 0$, $\theta_1(\omega)$ and $\theta_2(\omega)$ behave as

$$\theta_1(\omega) = \omega + O(\omega^3)\,,$$

$$\theta_2(\omega) = \frac{1}{2}\,\omega^2 + O(\omega^4)\,.$$

4. Observed exponential approximations

Because observed geometries are in a sense closer to the data than expected geometries and because they often give rise to simpler formulae (see Barndorff-Nielsen (1987, 1988)) it is natural to seek a way of approximating models $\mathscr{M}$ by "observed" exponential families $O^r\mathscr{M}_\omega$ analogous to Amari's "expected" exponential families $E^r\mathscr{M}_\omega$ described in Section 3. The obvious choice for the full exponential family generated by $O^r\mathscr{M}_\omega$ is just the full exponential approximation $\overline{E^r\mathscr{M}_\omega}$, that is, the full exponential family generated by the element of $\mathscr{M}$ given by ω and by the log-likelihood derivatives l_α for $1 \le |\alpha| \le r$. However, offhand it is not at all clear what the curved subfamily $O^r\mathscr{M}_\omega$ should be. Amari's construction of $E^r\mathscr{M}_\omega$ makes essential use of $E_\omega[\upsilon_\alpha \upsilon_\beta]$ where $\upsilon_\alpha = l_\alpha - E_\omega[l_\alpha]$. The observed analogue of $E_\omega[\upsilon_\alpha \upsilon_\beta]$ is zero and so is of no use in constructing $O^r\mathscr{M}_\omega$.

A clue to the manner in which one might construct $O^r\mathscr{M}_\omega$ comes from the following two observations:

(i) the canonical statistic of $E^r\mathscr{M}_\omega$ is

$$\{l_\alpha(\omega;\,\cdot\,) \colon 1 \le |\alpha| \le r\}\,,$$

(ii) l_α is obtained from l by applying the differential operator

$$\partial_\alpha = \frac{\partial^{|\alpha|}}{\partial\omega_1^{i_1} \cdots \partial\omega_d^{i_d}} \qquad \text{where} \qquad \alpha = (i_1, \ldots, i_d)\,.$$

Thus one might proceed by projecting (in a sense to be clarified) a version of the log-likelihood function onto a space of differential operators. To make this precise, let $\mathscr{F}^r$ denote the set of C^r real-valued functions on Ω and let $\mathscr{D}^r$ denote the set of linear differential operators of order r and with zero constant part, that is, $\mathscr{D}^r = \operatorname{span}\{\partial_\alpha \colon 1 \le |\alpha| \le r\}$. Then, given ω in Ω, there is a bilinear mapping $\mathscr{D}^r \times \mathscr{F}^r \to \mathbb{R}$ given by

$$(\partial_\alpha, f) \mapsto \partial_\alpha f(\omega)\,.$$

Let $a \colon \mathscr{X} \to A$ be an auxiliary statistic for $\mathscr{M}$ such that $(\hat{\omega}, a)$ is a one-to-one transformation of the minimal sufficient statistic, where $\hat{\omega}$ is the maximum likelihood estimator of ω. Then there is a linear mapping $\iota \colon \mathscr{D}^r \to \mathscr{F}^r$ given by

$$\iota\left[\sum_{|\alpha|=1}^{r} b_\alpha \partial_\alpha\right] = \sum_{|\alpha|=1}^{r} b_\alpha l_\alpha(\omega; \cdot, a),$$

where ω is fixed, the coefficients b_α are constants, and the right-hand side is regarded as a function on Ω. Thus, as in Section 2, the above bilinear mapping yields a linear mapping $\mathscr{D}^r \to (\mathscr{D}^r)^*$. If (as we shall assume) this is an isomorphism, then it gives rise in a canonical way to a projection Π^o of $\mathscr{F}^r$ onto $\mathscr{D}^r$. Denote by $\mathscr{L}$ the space of all measurable functions on $\mathscr{X}$. Then the linear mapping δ: $\mathscr{D}^r \to \mathscr{L}$ defined by

$$\delta\left[\sum_{|\alpha|=1}^{r} b_\alpha \partial_\alpha\right] = \sum_{|\alpha|=1}^{r} b_\alpha l_\alpha(\omega; \cdot),$$

can be regarded as sending $\mathscr{D}^r$ into the canonical parameter space of $\overline{E^r\mathscr{M}_\omega}$. By analogy with the expected development (3.3), we define the *observed development* Ψ_ω^o: $\Omega \to \mathscr{F}^r$ by

$$\Psi_\omega^o(\omega') = l(\omega'; \cdot, a) - l(\omega; \cdot, a).$$

The corresponding parallel transport operator is the trivial one sending f at ω' to f at ω. We now define $O^r\mathscr{M}_\omega$, the *r-th order observed exponential approximation to $\mathscr{M}$ at* ω, to be the submodel of $\overline{E^r\mathscr{M}_\omega}$ with log-density functions with respect to $p(\omega; \cdot)$ obtained from the composite function $\delta \circ \Pi^o \circ \Psi_\omega^o$: $\Omega \to \mathscr{L}$. More explicitly, $O^r\mathscr{M}_\omega = (\mathscr{X}, q_\omega^{[r]}, \Omega)$ with

$$(4.1) \qquad q_\omega^{[r]}(x; \omega') = p(x; \omega) \cdot \exp\{[l_{;\alpha}(\omega'; \omega, a) - l_{;\alpha}(\omega; \omega, a)]\jmath^{\alpha;\beta} l_\beta(\omega; x) - \kappa(\omega')\},$$

where the summation convention is used and $\jmath^{\alpha;\beta}$ denotes the (α, β)-th element of the inverse of the matrix which has (α, β)-th element $\jmath_{\alpha;\beta}$ for $1 \leq |\alpha|, |\beta| \leq r$ assuming that this matrix is invertible. The log-norming constant $\kappa(\omega')$ depends also on ω. Note that, although (4.1) gives a coordinate-based description of $O^r\mathscr{M}_\omega$, the construction above is indeed independent of the coordinate system chosen.

From the viewpoint of conditional inference we shall be interested in the model $\mathscr{M}|_a = (\Omega, p(\cdot, \cdot|a), \Omega)$, where $p(\cdot; \omega'|a)$ denotes the probability density function of $\hat{\omega}$ conditional on the auxiliary a under the element of $\mathscr{M}$ given by ω'. In the present context it is natural to approximate $\mathscr{M}|_a$ by $O^r(\mathscr{M}|_a)$, the r-th order observed exponential approximation to $\mathscr{M}|_a$ at ω. That is, $O^r(\mathscr{M}|_a)_\omega = (\Omega, \tilde{q}_\omega^{[r]}, \Omega)$ with

$$(4.2) \qquad \tilde{q}_\omega^{[r]}(\hat{\omega}; \omega'|a) = p(\hat{\omega}; \omega|a) \cdot \exp\{[l_{;\alpha}(\omega'; \omega|a) - l_{;\alpha}(\omega; \omega|a)]\jmath^{\alpha;\beta} l_\beta(\omega; \hat{\omega}|a)\}$$

$$\cdot \exp\{-\kappa(\omega';\omega|a)\}.$$

Differentiation of the identity

$$l(\omega';\hat{\omega}|a) = l(\omega';\hat{\omega},a) + \chi(\omega',a),$$

where $\chi(\omega',a)$ denotes the logarithm of the appropriate norming constant, yields

$$l_{;\alpha}(\omega';\hat{\omega}|a) = l_{;\alpha}(\omega';\hat{\omega},a),$$
$$l_{\beta}(\omega';\hat{\omega}|a) = l_{\beta}(\omega';\hat{\omega},a) + \chi_{\beta}(\omega',a)$$

and

$$l_{\beta;\alpha}(\omega;\omega|a) = \mathfrak{t}_{\beta;\alpha},$$

for $|\alpha| \geq 1$. (This last equation is used in (4.2) to obtain the expression $\mathfrak{t}^{\alpha;\beta}$.) It follows that $O^r(\mathscr{M}|_a)_\omega$ can also be obtained from $O^r\mathscr{M}_\omega$ by first restricting the sample space to Ω (identified with a subspace of $\mathscr{X}$ by the mapping $\hat{\omega} \mapsto (\hat{\omega},a)$) and then renormalizing. We shall refer to $O^r(\mathscr{M}_{|a})_\omega$ as the *r-th order conditional observed exponential approximation to $\mathscr{M}$ at ω*. Formulae (4.1) and (4.2) should be compared with their "expected" analogue (3.2). Note that κ does not necessarily denote the same function in these three formulae.

Although $E^r\mathscr{M}_\omega$ and $O^r\mathscr{M}_\omega$ are both d-dimensional curved exponential submodels of $\overline{E^r\mathscr{M}_\omega}$, there appears to be no general relationship between $E^r\mathscr{M}_\omega$ and $O^r\mathscr{M}_\omega$.

We now investigate some simple aspects of $O^r\mathscr{M}_\omega$ analogous to those of $E^r\mathscr{M}_\omega$ considered in the second half of Section 3.

One might reasonably ask about the behaviour of observed exponential approximations under repeated random sampling. That is, how are $O^r(\mathscr{M}^n)_\omega$ and $(O^r\mathscr{M}_\omega)^n$ related? In general, it is not clear what auxiliary statistic is appropriate for $\mathscr{M}^n$. Even in the case of a location model (discussed in Example 4.1 below) where there is an "obvious" auxiliary, there seems to be no simple connection between $O^r(\mathscr{M}^n)_\omega$ and $(O^r\mathscr{M}_\omega)^n$— in contrast to the "expected" case.

For curved exponential models and for transformation models, the observed exponential approximations have properties similar to those of their "expected" counterparts.

For curved exponential models, observed exponential approximations of suitable order are exact. In fact, $\mathscr{M} = O^r\mathscr{M}_\omega$ (that is $p = q^{[r]}$) if and only if $\mathscr{M}$ is a curved (k,d)-exponential model with $k = \dim \operatorname{span}\{l_\alpha(\omega;\cdot): |\alpha| \leq r\}$.

Now suppose that $\mathcal{M}$ is a (composite) transformation model under the action of a group G. Suppose that the dominating measure μ of $\mathcal{M}$ is invariant under G and that the auxiliary statistic is G-invariant. Then the action of G on $\mathcal{X}$ induces the following actions of G on the sets $\bigcup_{\omega \in \Omega} O^r\mathcal{M}_\omega$ and $\bigcup_{\omega \in \Omega} O^r(\mathcal{M}_{|a})_\omega$. The element g of G sends $q_\omega^{[r]}(\cdot\,;\omega'|a)$ to $q_\omega^{[r]}(g^{-1}\cdot\,;\omega'|a)$ and $\tilde{q}_\omega^{[r]}(\cdot\,;\omega'|a)$ to $\tilde{q}_\omega^{[r]}(g^{-1}\cdot\,;\omega'|a)$. Further, we have

$$q_\omega^{[r]}(x;\omega'|a) = q_{g\omega}^{[r]}(gx;g\omega'|a)$$

and

$$\tilde{q}_\omega^{[r]}(\hat{\omega};\omega'|a) = \tilde{q}_{g\omega}^{[r]}(g\hat{\omega};g\omega'|a)\,.$$

Example 4.1. (*Location model*) Consider a location model with $\mathcal{X} = \Omega = \mathbb{R}$ and $p(x;\omega) = f(x-\omega)$ for some known probability density function f. A natural choice of auxiliary statistic for $\mathcal{M}^n$ is the configuration $(a_1,\ldots,a_n)$ defined by

$$a_i = x_i - \hat{\omega}\ ,$$

where $x_1,\ldots,x_n$ are the observations. Then

$$l(\omega;\hat{\omega},a) = -\sum_{i=1}^n g(a_i + \hat{\omega} - \omega)\ ,$$

where $g = -\log f$, and so

$$\ell_{s;t} = (-1)^{s+1}\sum_{i=1}^n g^{(s+t)}(a_i)\,.$$

Note, in particular, that for $r \geq 2$, the matrix $[\ell_{s;t}]$ is not symmetric. The observed exponential approximation $O^1\mathcal{M}_\omega^n$ has model function

$$q_\omega^{[1]}(x_1,\ldots,x_n;\omega') = \left\{\prod_{i=1}^n f(x_i-\omega)\right\}\varphi(\omega')$$
$$\cdot \exp\left\{w(a,\omega,\omega')\left[\sum_{i=1}^n g^{(2)}(a_i)\right]^{-1}\left[\sum_{i=1}^n g'(x_i-\omega)\right]\right\},$$

where $w(a,\omega,\omega') = \sum_{i=1}^n (g'(a_i) - g'(a_i+\omega-\omega'))$ and $\varphi(\omega')$ is a norming constant.

As the following example indicates, the condition that the matrix $[\ell_{\alpha;\beta}]$ is nonsingular, which underlies the definitions (4.1) and (4.2), can be presumed to be valid in great generality.

Example 4.2. Let $\lambda \in \mathbb{R}$, $\kappa > 0$ and $\varphi > 0$ be parameters, here considered known, and let $y_1, \ldots, y_n$ be a sample from the scale parameter family generated from the family with model function

$$cy^{\lambda-1} \exp\{-\kappa(y^{-1} + y^{\varphi})\},$$

where $y > 0$. This scale family constitutes a (2, 1) exponential model. (For $\varphi = 1$ the distributions considered belong to the family of generalized inverse Gaussian distributions.) Denoting the scale parameter by σ and letting $x_i = \log y_i$ and $\omega = \log \sigma$ we find that

$$\ell_{1;1} = \frac{\kappa}{2}\{\Sigma e^{-a_i} + \varphi^2 \Sigma e^{\varphi a_i}\},$$

$$\ell_{2;1} = \frac{\kappa}{2}\{\Sigma e^{-a_i} - \varphi^3 \Sigma e^{\varphi a_i}\} = -\ell_{1;2},$$

$$\ell_{2;2} = -\frac{\kappa}{2}\{\Sigma e^{-a_i} + \varphi^4 \Sigma e^{\varphi a_i}\},$$

where $a_i = x_i - \hat{\omega}$. Hence

$$\begin{vmatrix} \ell_{1;1} & \ell_{2;1} \\ \ell_{1;2} & \ell_{2;2} \end{vmatrix} = -\frac{\kappa^2}{4}(\Sigma e^{-a_i})(\Sigma e^{\varphi a_i})\varphi^2(\varphi+1)^2,$$

which is $\neq 0$ whatever the values of λ, κ, φ and $x_1, \ldots, x_n$.

5. Accuracy of approximations

One of the major purposes of introducing $O^r\mathcal{M}_\omega$ and $O^r(\mathcal{M}_{|a})_\omega$ is to obtain tractable approximations to $\mathcal{M}$ and $\mathcal{M}_{|a}$. In this section we investigate the closeness of these approximations.

Before going into detail, it is worth widening the discussion to include a general class of exponential families like $O^r\mathcal{M}_\omega$ which approximate $\mathcal{M}$ near ω. Recall that $\mathcal{L}$ denotes the space of all measurable real-valued functions on $\mathcal{X}$. In practice, we shall mostly be interested in certain convex subsets of $\mathcal{L}$ consisting of functions which satisfy appropriate regularity conditions. If Π is any projection of $\mathcal{L}$ onto the subspace spanned by $\{l_\alpha(\omega; \cdot): |\alpha| \leq r\}$, then Π gives rise to a curved exponential sub-model $\Pi\mathcal{M}_\omega = (\mathcal{X}, p^{[\Pi]}, \Omega)$ of $\overline{E^r\mathcal{M}_\omega}$ with

$$p^{[\Pi]}(x; \omega') = p(x; \omega) \exp\{\Pi(l(\omega'; x) - l(\omega; x)) - \kappa(\omega')\}, \tag{5.1}$$

where $\kappa(\omega')$ is an appropriate log-norming constant. To see that $O'\mathscr{M}_\omega$ can be constructed in this way, take $\mathscr{L}^r$ to be the space of C^r functions on $\mathscr{X}$ and define $\Phi^o_\omega: \Omega \to \mathscr{L}^r$ as the log-likelihood-ratio function given by

$$\Phi^o_\omega(\omega') = l(\omega'; \cdot) - l(\omega; \cdot) .$$

Let $\rho: \mathscr{L}^r \to \mathscr{F}^r$ be the restriction map which is given by

$$\rho(f)(\hat{\omega}) = f(\hat{\omega}, a) ,$$

and define $\Pi: \mathscr{L}^r \to \mathscr{L}^r$ by

$$\Pi = \delta \circ \Pi^o \circ \rho .$$

Then, Π is a projection and commutativity of the diagram in Fig. 3 shows that (4.1) is a special case of (5.1). Similarly, (4.2) and (3.2) are analogous to (5.1), with $(\mathscr{F}^r, \Pi^o)$ and $(\mathscr{H}_\omega, \Pi^e)$ used in place of $(\mathscr{L}^r, \Pi)$. Here Π^e denotes orthogonal projection of $\mathscr{H}_\omega$ onto span $\{l_\alpha(\omega): |\alpha| \leq r\}$.

The closeness of an approximating family $\Pi\mathscr{M}_\omega$ to $\mathscr{M}$ near ω can be investigated by using Taylor expansions and by exploiting the linear nature of projections. We have

$$\log p(x; \omega') = \log p(x; \omega) + l(\omega'; x) - l(\omega; x)$$

and

$$\log p^{[\Pi]}(x; \omega') = \log p(x; \omega) + \Pi(l(\omega'; x) - l(\omega; x)) - \kappa(\omega') .$$

Then, by Taylor expansion and using the fact that $\Pi l_\alpha(\omega) = l_\alpha(\omega)$ for $|\alpha| \leq r$, we have

$$\begin{tikzcd}
 & \mathscr{L}^r \arrow[r, "\Pi"] \arrow[dd, "\rho"] & \mathscr{L}^r \\
\Omega \arrow[ur, "\Phi^o_\omega"] \arrow[dr, "\Psi^o_\omega"'] & & \\
 & \mathscr{F}^r \arrow[r, "\Pi^o"] & \mathscr{D}^r \arrow[uu, "\delta"']
\end{tikzcd}$$

Fig. 3. Relationship of the orthogonal projections Π and Π^o.

(5.2) $$\begin{aligned}\log p^{[\Pi]}(x;\omega') &- \log p(x;\omega')\\ &= (\Pi - I)(l(\omega';x) - l(\omega;x)) - \kappa(\omega')\\ &= (\Pi - I)\left[\sum_{|\alpha|=1}^{2r+1} l_\alpha(\omega;x)\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2})\right] - \kappa(\omega')\\ &= (\Pi - I)\left[\sum_{|\alpha|=r+1}^{2r+1} l_\alpha(\omega;x)\frac{\delta^\alpha}{\alpha!}\right] + O(\|\delta\|^{2r+2}) - \kappa(\omega')\\ &= \sum_{|\alpha|=r+1}^{2r+1} C_\alpha(\omega;x)\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2}) - \kappa(\omega'),\end{aligned}$$

where

$$C_\alpha(\omega;x) = (\Pi - I)l_\alpha(\omega;x),$$

I denotes the identity map of $\mathscr{L}$, and $\delta = \omega' - \omega \to 0$.

Since $p^{[\Pi]}(\cdot\,;\omega')$ is a probability density function with respect to the underlying measure μ on $\mathscr{X}$, we have

$$\int \exp(\log p^{[\Pi]}(x;\omega'))d\mu(x) = 1.$$

It then follows from (5.2) that the norming constant $\kappa(\omega')$ is of the form

$$\begin{aligned}\exp\kappa(\omega') &= \int \exp\left\{\sum_{|\alpha|=r+1}^{2r+1} C_\alpha(\omega;x)\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2})\right\}p(x;\omega')d\mu(x)\\ &= \int\left\{1 + \sum_{|\alpha|=r+1}^{2r+1} C_\alpha(\omega;x)\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2})\right\}p(x;\omega')d\mu(x)\\ &= 1 + \sum_{|\alpha|=r+1}^{2r+1} E_{\omega'}[C_\alpha(\omega;x)]\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2}),\end{aligned}$$

under fairly general regularity conditions. Then

$$\kappa(\omega') = \sum_{|\alpha|=r+1}^{2r+1} E_{\omega'}[C_\alpha(\omega;x)]\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2})$$

and

(5.3) $$\begin{aligned}\log&\left[\frac{p^{[\Pi]}(x;\omega')}{p(x;\omega')}\right]\\ &= \sum_{|\alpha|=r+1}^{2r+1} B_\alpha(\omega,\omega';x)\frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2}),\end{aligned}$$

where

$$B_\alpha(\omega, \omega'; x) = C_\alpha(\omega; x) - E_{\omega'}[C_\alpha(\omega; x)] ,$$

so that $\Pi\mathscr{M}_\omega$ is r-th order close to $\mathscr{M}$ near ω. There is also corresponding closeness of derivatives of log-likelihood because differentiation of (5.3) yields

$$l_\beta^{[\Pi]}(\omega'; x) - l_\beta(\omega'; x) = \sum_{|\alpha|=r+1} B_\alpha(\omega, \omega; x) \frac{\delta^{\alpha-\beta}}{(\alpha-\beta)!} + O(\|\delta\|^{r+2-|\beta|}) , \tag{5.4}$$

where $l^{[\Pi]}(\omega'; x) = \log p^{[\Pi]}(x; \omega')$ and $\delta^{\alpha-\beta}/(\alpha-\beta)!$ is interpreted as 0 if any component of $\alpha - \beta$ is negative. In particular, putting $l^o(\omega'; x) = \log q_\omega^{[r]}(x; \omega')$, the versions of (5.3) and (5.4) for $O^r\mathscr{M}_\omega$ are

$$\log\left[\frac{q_\omega^{[r]}(x; \omega')}{p(x; \omega')}\right] = \sum_{|\alpha|=r+1}^{2r+1} B_\alpha(\omega, \omega'; x) \frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2}) \tag{5.5}$$

and

$$l_\beta^o(\omega'; x) - l_\beta(\omega'; x) = \sum_{|\alpha|=r+1} B_\alpha(\omega, \omega; x) \frac{\delta^{\alpha-\beta}}{(\alpha-\beta)!} + O(\|\delta\|^{r+2-|\beta|}) . \tag{5.6}$$

As $E^r\mathscr{M}_\omega$ and $O^r(\mathscr{M}_{|a})_\omega$ are also obtained by projection (but using $\mathscr{H}_\omega$ and $\mathscr{F}^r$ instead of $\mathscr{L}^r$), formulae analogous to (5.3) and (5.4) show that $E^r\mathscr{M}_\omega$ and $O^r(\mathscr{M}_{|a})_\omega$ are r-th order close to $\mathscr{M}$ and $\mathscr{M}_{|a}$, respectively, near ω. Thus for $O^r(\mathscr{M}_{|a})_\omega$, we have

$$\log\left[\frac{\tilde{q}^{[r]}(\hat{\omega}; \omega'|a)}{p(\hat{\omega}; \omega'|a)}\right] = \sum_{|\alpha|=r+1}^{2r+1} B_\alpha(\omega, \omega'; \hat{\omega}|a) \frac{\delta^\alpha}{\alpha!} + O(\|\delta\|^{2r+2}) , \tag{5.7}$$

where $B_\alpha(\omega, \omega'; \hat{\omega}|a) = C_\alpha(\omega, \omega'; \hat{\omega}|a) - E_{\omega'}[C_\alpha(\omega, \omega'; \hat{\omega})]$, the expectation being with respect to the conditional probability density function $p(\cdot\,; \omega'|a)$. Putting

$$\tilde{l}^{[r]}(\omega'; \hat{\omega}|a) = \log \tilde{q}^{[r]}(\hat{\omega}; \omega'|a)$$

and

$$l(\omega'; \hat{\omega}|a) = \log p(\hat{\omega}; \omega'|a) ,$$

differentiation of (5.5) yields

$$(5.8) \qquad \tilde{l}_{\beta}^{[r]}(\omega'; \hat{\omega}|a) - l_{\beta}(\omega'; \hat{\omega}|a)$$
$$= \sum_{|\alpha|=r+1} B_{\alpha}(\omega, \omega; \hat{\omega}) \frac{\delta^{\alpha-\beta}}{(\alpha-\beta)!} + O(\|\delta\|^{r+2-|\beta|}) .$$

Next we wish to compare the maximum likelihood estimators $\hat{\omega}$ and $\hat{\omega}^{[\Pi]}$ in $\mathscr{M}^n$ and $\Pi(\mathscr{M}^n)$, respectively, at least asymptotically for large n. Let Π_n denote the projection used in the construction of $\Pi(\mathscr{M}^n)_{\omega}$. Let $\mathscr{L}_n$ and $\mathscr{L}_{\infty}$ denote the spaces of all measurable real-valued functions on $\mathscr{X}^n$ and $\mathscr{X}^{\infty}$, respectively, where $\mathscr{X}^{\infty}$ is the set of infinite sequences of elements of $\mathscr{X}$. Then, Π_n, which is a projection of $\mathscr{L}_n$ onto some subspace, can also be regarded as a projection of $\mathscr{L}_{\infty}$ onto a subspace of $\mathscr{L}_{\infty}$. We shall assume that Π_n converges (in some sense) to a limiting projection Π, so that $\Pi_n = \Pi + O(n^{-1/2})$. For the case of observed exponential approximations $O^r(\mathscr{M}^n)_{\omega}$, this holds in particular if

$$(5.9) \qquad \lim_{n\to\infty} n^{-1} \ell_{\alpha;\beta}(\omega) \quad \text{exists for all } \alpha, \beta \text{ with } |\alpha|, |\beta| \le r .$$

Then, we have, under appropriate regularity conditions,

$$\begin{aligned}(5.10) \quad & n^{-1}B_{\alpha}(\omega, \omega; x) \\ &= n^{-1}(\Pi - I + O(n^{-1/2}))\, l_{\alpha}(\omega; x) \\ &\quad - E_{\omega}[n^{-1}(\Pi - I + O(n^{-1/2}))\, l_{\alpha}(\omega; x)] \\ &= \{n^{-1}(\Pi - I)\, l_{\alpha}(\omega; x) - E_{\omega}[n^{-1}(\Pi - I)\, l_{\alpha}(\omega; x)]\} + O_p(n^{-1/2}) \\ &= O_p(n^{-1/2}) .\end{aligned}$$

Then, taking $\omega' = \hat{\omega}^{[\Pi]}$ and $|\beta| = 1$ in (5.4) gives

$$0 - n^{-1} l_{\beta}(\hat{\omega}^{[\Pi]}; x)$$
$$= \sum_{|\alpha|=r+1} n^{-1} B_{\alpha}(\omega, \omega; x) \frac{\hat{\delta}^{\alpha-\beta}}{(\alpha-\beta)!} + O(\|\hat{\delta}\|^{r+1}) ,$$

and so

$$n^{-1}\hat{j}(\hat{\omega}^{[\Pi]} - \hat{\omega}) + O(\|\hat{\omega}^{[\Pi]} - \hat{\omega}\|^2) = O_p(n^{-1/2})O(\|\hat{\delta}\|^r) ,$$

where $\hat{j}$ denotes observed information at $\hat{\omega}$ and $\hat{\delta} = \hat{\omega}^{[\Pi]} - \omega$. As $\hat{\omega} - \omega = O_p(n^{-1/2})$, we obtain

$$\hat{\omega}^{[\Pi]} - \hat{\omega} = O_p(n^{-(r+1)/2}) . \tag{5.11}$$

Let $\hat{\omega}^e$, $\hat{\omega}^o$ and $\tilde{\omega}$ denote, respectively, the maximum likelihood estimators of ω' in the models $E^r\mathcal{M}^n_\omega$, $O^r(\mathcal{M}^n)_\omega$ and $O^r(\mathcal{M}_{|a})_\omega$. Then, as a special case of (5.11) we have

$$\hat{\omega}^e - \hat{\omega} = O_p(n^{-(r+1)/2}) . \tag{5.12}$$

A similar argument to the one above shows that, if condition (5.9) holds,

$$\hat{\omega}^o - \hat{\omega} = O_p(n^{-(r+1)/2}) . \tag{5.13}$$

Example 5.1. (*von Mises distribution with known concentration*) Consider again Example 3.1. For random samples $x_1,\dots,x_n$ from the distribution with model function (3.4) we can write

$$\sum_{i=1}^{n} (\cos x_i, \sin x_i) = r(\cos \hat{\omega}, \sin \hat{\omega}) ,$$

and $(r, \hat{\omega})$ is a one-to-one transformation of the minimal sufficient statistic. Calculation shows that $O^1\mathcal{M}_\omega$, the observed approximation based on the ancillary r, has model function

$$q^{[1]}(\hat{\omega}, r; \omega') = \{I_0(\kappa\sqrt{1 + (\sin \delta)^2})\}^{-n} \cdot \exp\{\kappa r[\cos(\hat{\omega} - \omega) + \sin\delta \sin(\hat{\omega} - \omega)]\} , \tag{5.14}$$

where $\delta = \omega' - \omega$. Differentiation of (5.14) shows that $\hat{\omega}^o$ satisfies

$$\begin{aligned} &n^{-1}r \sin(\hat{\omega} - \omega) \\ &= A(\kappa\sqrt{1 + (\sin \hat{\delta}^o)^2}) \sin \hat{\delta}^o\{1 + (\sin \hat{\delta}^o)^2\}^{-1/2} , \end{aligned}$$

where $\hat{\delta}^o = \hat{\omega}^o - \omega$ and $A(z) = I_1(z)/I_0(z)$. As $n^{-1}r - A(\kappa) = O_p(n^{-1/2})$, it follows from further calculation that $\hat{\omega}^o - \hat{\omega} = O_p(n^{-1})$ in agreement with (5.13). However, as $n^{-1}r - A(\kappa) \neq O_p(n^{-1})$, we have $\hat{\omega}^o - \hat{\omega} \neq O_p(n^{-3/2})$, so that we cannot replace $r + 1$ by a larger integer in (5.13).

Now consider the conditional observed approximation $O^1(\mathcal{M}|_r)_\omega$. By restriction of (5.14) and renormalization we obtain

$$(5.15) \quad \tilde{q}^{[1]}_{\omega}(\hat{\omega}; \omega' | r) = \{I_0(\kappa r\sqrt{1 + (\sin\delta)^2})\}^{-1} \cdot \exp\{\kappa r[\cos(\hat{\omega} - \omega) + \sin\delta \sin(\hat{\omega} - \omega)]\}.$$

Differentiation of (5.15) leads to

$$\sin(\hat{\omega} - \omega) = \sin(\tilde{\delta}) A(\kappa r\sqrt{1 + (\sin\tilde{\delta})^2})\{1 + (\sin\tilde{\delta})^2\}^{-1/2},$$

where $\tilde{\delta} = \tilde{\omega} - \omega$. Since $\kappa r = O_P(n)$ for $\kappa > 0$ and $A(z) = 1 - O(z^{-1})$ as $z \to \infty$ (see e.g., Mardia (1972), p. 288), it follows that $\tilde{\omega} - \omega = O_P(n^{-3/2})$. Thus $\tilde{\omega}$ can be considerably closer to $\hat{\omega}$ than $\hat{\omega}^o$ is.

REFERENCES

Amari, S.-I. (1987). Differential-geometrical theory of statistics—towards new developments, *Differential Geometry in Statistical Inference*, Lecture Notes—Monograph Series, 19–94, Institute of Mathematical Statistics, Hayward.

Barnard, G. A. (1971). Scientific inferences and day-to-day decisions, *Foundations of Statistical Inference*, (eds. V. P. Godambe and D. A. Sprott), 289–300, Holt, Rinehart and Winston of Canada, Toronto.

Barnard, G. A. (1982). Contribution to the discussion paper by D. V. Lindley, *Internat. Statist. Rev.*, **50**, 11–14.

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

Barndorff-Nielsen, O. E. (1987). Differential and integral geometry in statistical inference, *Differential Geometry in Statistical Inference*, Lecture Notes-Monograph Series, 95–161, Institute of Mathematical Statistics, Hayward.

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

Dawid, A. P. (1975). Contribution to discussion of "Defining the curvature of a statistical problem (with application to second order efficiency)" by B. Efron, *Ann. Statist.*, **3**, 1131–1134.

DiCiccio, T. J. (1984). On parameter transformations and interval estimation, *Biometrika*, **71**, 477–485.

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

Husemoller, D. (1966). *Fibre Bundles*, McGraw-Hill, New York.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*, Springer, Heidelberg.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer, New York.

Lichnerowicz, A. (1976). *Global Theory of Connections and Holonomy Groups*, Noordhoff International Publishing, Leyden.

Mardia, K. V. (1972). *Statistics of Directional Data*, Academic Press, New York.

Solmitz, E. T. (1964). Analysis of experiments in particle physics, *Annu. Rev. Nucl. Sci.*, **14**, 375–402.

Sprott, D. A. (1977). On the theory of statistical estimation, *Math. Sci.*, **2**, 127–137.