

On large deviations and choice of ancillary for p^* and r^*

OLE E. BARNDORFF-NIELSEN¹ and ANDREW T.A. WOOD^{2*}

¹*Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark*

²*School of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK*

The large deviation properties of p^* , the approximation to the conditional density of the maximum likelihood estimator, and r^* , the modified directed likelihood, are studied. Attention is restricted to curved exponential models. Various specifications of an approximate ancillary, which are required in the construction of p^* and r^* , are considered, including: a modified directed likelihood ancillary, a^* , and an unmodified directed likelihood ancillary, a^o . It is shown that if a^* is used then p^* and r^* achieve saddlepoint accuracy on both normal and large deviation regions; if, on the other hand, a^o is used in the construction of p^* , then saddlepoint accuracy is not achieved, though the relative error still stays bounded on large deviation regions. It is also shown that if a^o rather than a^* is held fixed in the sample space differentiations needed to calculate r^* , then saddlepoint accuracy is still attained in both normal and large deviation regions. On a first impression, the last result is a little surprising because, in a repeated sampling framework, a^o is only ancillary to order $O(n^{-1/2})$. However, this finding is also of direct practical interest because, from the point of view of calculation, a^o is often substantially easier to work with than a^* . An important aspect of our approach is the development of guidelines, referred to as Laplace-spa calculus, for the construction of invariant saddlepoint-style approximations to marginal and conditional densities. Finally, connections between recent work by Jensen and the results of this paper are discussed and clarified.

Keywords: conditional inference; Laplace's approximation; likelihood ancillary; relative error; saddlepoint approximation; tail-area approximation

1350–7265 © 1998 Chapman & Hall

1. Introduction

This paper is concerned with the large deviation properties of p^* and r^* in (m, d) exponential models (i.e. curved exponential models). In notation which is described in more detail in Section 2, p^* is the approximation $p^* = c|\hat{j}|^{1/2} e^{\ell - \hat{\ell}}$ to the (conditional) density of the maximum likelihood estimator and $r^* = r + r^{-1} \log(u/r)$ is the modified directed likelihood, where r is the unmodified directed, or signed, likelihood and u is a certain test statistic. Further details of p^* and r^* are given by Barndorff-Nielsen (1983; 1986; 1991), McCullagh (1987), Fraser (1988), Reid (1988), Skovgaard (1990) and Barndorff-Nielsen and Cox (1994).

*To whom correspondence should be addressed. e-mail: atw@maths.bath.ac

1350–7265 © 1998 Chapman & Hall

In (m, d) exponential models, m is the dimension of the canonical statistic t , and d is the dimension of $\hat{\omega}$, the maximum likelihood estimator of the parameter ω . In (m, m) exponential models, $\hat{\omega}$ is a minimal sufficient statistic and in this case p^* is known to be equivalent to a normalized saddlepoint approximation; see, for example, Barndorff-Nielsen and Cox (1994). However, to determine p^* when $d < m$, it is necessary to specify a second statistic, a , of dimension $k = m - d$ with the following properties: (i) $(\hat{\omega}, a)$ should be jointly sufficient; and (ii) a should be exactly or approximately ancillary. Typically a will be constructed in such a way that it has an asymptotic k -dimensional standard normal distribution.

A normal deviation region is one in which $\hat{\omega} - \omega$ and $\bar{a} = n^{-1/2}a$ are both of order $O(n^{-1/2})$, where n is the sample size in a repeated sampling framework. A number of authors have studied the accuracy of p^* in normal deviation regions: the basic finding is that, provided a is ancillary to order $O(n^{-1})$ (and locally ancillary to order $O(n^{-3/2})$), p^* has error $O(n^{-3/2})$ on normal deviation regions; see Barndorff-Nielsen and Cox (1994) for further details and references. However, under general (m, d) exponential models with $d < m$, little attention has been given to the properties of p^* in large deviation regions (that is, regions in which one or both of $\hat{\omega} - \omega$ and \bar{a} are of the order $O(1)$), but see Skovgaard (1990). We comment on Skovgaard's discussion in more detail in Section 6.

Much of our attention will be focused on two particular types of approximate ancillary: the modified and unmodified directed likelihood ancillaries, a^* and a^o , respectively. It is shown that if a^* is used in the construction of p^* , then both have saddlepoint accuracy. That is, for either normal or large deviations of a^* , the p^* approximation to the conditional density of $\hat{\omega}$ given a^* has relative error of order $O(n^{-3/2})$ and $O(n^{-1})$ for normal and large deviation regions of $\hat{\omega} - \omega$, respectively. If, on the other hand, a^o is used in the construction of p^* , then saddlepoint accuracy is not achieved, though the relative error still stays bounded on large deviation regions.

In the case of r^* , our conclusions are more subtle: the standard normal approximation to the distribution of r^* may be interpreted as an approximation to the conditional distribution of r^* given a^* regardless of whether we keep a^* or a^o fixed in the required sample space differentiations. On a first impression, this finding may seem surprising because a^o is only ancillary to order $O(n^{-1/2})$. This finding is also of direct practical interest because, from the point of view of calculation, a^o is usually far easier to work with than a^* .

We also give some consideration to a general class of first-order ancillaries which includes the Efron–Hinkley and score ancillaries. It is shown that, if an ancillary in this class is used in the construction of p^* , then typically the relative error in p^* does not stay bounded if the ancillary varies over a large deviation region.

Our strategy is to compare p^* , and the standard normal approximation to the distribution of r^* , with certain saddlepoint-type approximations. This approach is closely connected to recent developments in saddlepoint and related approximations to marginal and conditional distributions involving nonlinear statistics. A number of authors have been involved in these developments, including Tierney *et al.* (1989), Fraser (1990), DiCiccio *et al.* (1990), Barndorff-Nielsen (1990; 1991), Daniels and Young (1991), DiCiccio and Martin (1991), Jensen (1992) and Jing and Robinson (1994). The basic ideas may be summarized as follows. Given an approximation to the joint density of a set of variables, for example a saddlepoint approximation, we estimate marginal densities by applying Laplace's approximation to the

appropriate integral of the joint density approximation, and we estimate conditional densities as the joint density estimate divided by the appropriate marginal density estimate. A more detailed specification of this approach is given in Section 4. In the case of one-dimensional variables, tail probabilities may then be estimated by applying Temme's (1982) method to the relevant marginal or conditional density approximation.

Empirical evidence suggests that approximations of the type indicated above are typically very accurate; see, for example, Barndorff-Nielsen and Cox (1994), Barndorff-Nielsen and Schmidli (1995) and Jensen (1995). Moreover, the mathematical form of these approximations suggests that they will generally have desirable relative error properties on large deviation regions. However, except in particular cases, rigorous mathematical proofs are lacking, particularly when conditioning is involved. Rigorous justification of these approximations under general conditions is an open problem, though the technical difficulties can be avoided by restricting attention to 'local' large deviations. Further discussion is given in Section 3.

In Section 2, more detailed definitions of p^* , r^* and the approximate ancillaries a^* and a^o are given. In Section 3 our main results are presented, and proofs are given in Section 5. Saddlepoint approximations for nonlinear statistics are reviewed in Section 4, and the issue of normalization is also discussed. The main focus is on the construction of invariant forms of these approximations, and the resulting guidelines are referred to as Laplace-spa calculus (where spa is an abbreviation for saddlepoint approximation). We conclude in Section 6 with some discussion of our results, and we also clarify the connection between the results in this paper and those of Jensen (1992). An illustrative example is considered in Sections 2 and 6.

Finally, some remarks on notation. For $q \geq 1$ and vectors $x_1 \in \mathbb{R}^{p_1}, \dots, x_q \in \mathbb{R}^{p_q}$, we identify $\mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_q}$ with $\mathbb{R}^{p_1 + \dots + p_q}$ and define the following norm on $\mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_q}$:

$$\|x_1, \dots, x_q\| = \left\{ \sum_{j=1}^q \sum_{i=1}^{p_j} (x_j^i)^2 \right\}^{1/2}.$$

For each $\gamma > 0$, we define the open ball

$$B_\gamma = \{(x_1, \dots, x_q): \|x_1, \dots, x_q\| < \gamma\} \subset \mathbb{R}^{p_1 + \dots + p_q}.$$

The dimensions of the vectors concerned will usually not be made explicit. A bar over a quantity such as r , u or a will indicate multiplication by $n^{-1/2}$: i.e. $\bar{r} = n^{-1/2}r$, $\bar{u} = n^{-1/2}u$ and so on. The motivation for doing this will be explained in Section 2.4. The notation for likelihood derivatives will follow Barndorff-Nielsen and Cox (1994).

2. Preliminaries

2.1. Curved exponential models

We begin with the definition of a prime exponential model. Consider a model \mathcal{M} with model function of the form

$$p(y; \theta) = a_0(y) \exp \{ \theta^T s(y) - \kappa(\theta) \} \quad \theta \in \Theta \quad (2.1)$$

where $\Theta = \{ \theta: \int a_0(y) \exp \{ \theta^T s(y) \} \mu(dy) < \infty \}$, μ is a dominating measure, and the canonical parameter θ and canonical statistic s are assumed to have dimension m . Write Θ^0 for the interior of Θ , and let \mathcal{M}^0 denote the model obtained from \mathcal{M} by deleting all $p(\cdot; \theta)$ for which $\theta \in \Theta \setminus \Theta^0$. Consider the following condition: for any $\theta_0 \in \Theta^0$ and for any point $\theta_1 \in \Theta \setminus \Theta^0$, we have

$$\frac{d}{d\lambda} \kappa(\lambda\theta_1 + (1-\lambda)\theta_0) \rightarrow \infty$$

as λ tends to 1 from below. If this condition is satisfied, then \mathcal{M} , \mathcal{M}^0 and $\kappa(\theta)$ are said to be *steep*. By definition, \mathcal{M}^0 is a *prime* exponential model if it is steep. Important properties of prime exponential models are summarized in Barndorff-Nielsen and Cox (1994, pp. 63–64).

Now suppose that the canonical parameter θ is restricted to be a smooth function of a parameter ω of dimension $d < m$, and domain of variation Ω . The resulting model, referred to below as \mathcal{M}_Ω , has a model function of the form

$$p(y; \omega) = a_0(y) \exp [\theta^T(\omega) s(y) - \kappa\{\theta(\omega)\}], \quad \omega \in \Omega. \quad (2.2)$$

Assuming that $\theta(\Omega)$ does not lie in an affine subspace of Θ , (2.2) defines a curved exponential model or, more specifically, an (m, d) exponential model. The original model (2.1) is referred to as the ambient (m, m) exponential model for the curved exponential model (2.2). Throughout this paper it is assumed that the ambient model is a prime exponential model. This is in fact the case for most exponential models of applied interest.

It will be assumed that $s(y) = \sum_{i=1}^n y_i$, where the y_i are realizations of independent and identically distributed random vectors, and we shall take as our starting point the density, with respect to Lebesgue measure on \mathbb{R}^m , of the minimal sufficient statistic $t = n^{-1}s(y)$. This density is given by

$$p(t; \omega) = a_1(t) \exp \{ n[\theta^T(\omega)t - K\{\theta(\omega)\}] \}, \quad (2.3)$$

where $K\{\theta(\omega)\} = n^{-1}\kappa\{\theta(\omega)\}$. In general, the density (2.3) need not exist. However, throughout this paper we shall be working under an assumption, given in (3.1) below, which ensures the existence of this density for n sufficiently large.

The assumption that the y_i are independent and identically distributed can certainly be weakened significantly. The key requirement is essentially that the cumulants of the canonical statistic nt should be $O(n)$.

2.2. The p^* formula

The p^* formula for curved exponential models is now introduced. Taking (2.3) as our starting point, let a be a quantity of dimension $k = m - d$ with the following two properties, for n sufficiently large: (i) there is a smooth one–one relationship between the canonical statistic t and $(\hat{\omega}, a)$, where $\hat{\omega}$ is the maximum likelihood estimator of ω under model \mathcal{M}_Ω ; (ii) a is approximately distribution constant under model \mathcal{M}_Ω . For example, a^* defined in

(2.11) below satisfies these requirements; more detailed discussion is given in Sections 2.4 and 2.5. For reasons to be explained there, it will generally be more convenient and appropriate to work with $\bar{a} = n^{-1/2}a$ rather than a .

Let $p(\hat{\omega}; \omega|\bar{a})$ denote the true conditional density of the maximum likelihood estimator, $\hat{\omega}$, given \bar{a} . Then $p(\hat{\omega}; \omega|\bar{a})$ may be approximated by $p^*(\hat{\omega}; \omega|\bar{a})$, with the latter given by

$$p^*(\hat{\omega}; \omega|\bar{a}) = c|j(\hat{\omega})|^{1/2} \exp \{ \ell(\omega) - \ell(\hat{\omega}) \}, \quad (2.4)$$

where $\ell(\omega)$ and $\ell(\hat{\omega})$ are the log-likelihood for the model evaluated at the true parameter and the maximum likelihood estimator, respectively; $j(\hat{\omega})$ is the observed information evaluated at the maximum likelihood estimator; and $c = c(\omega, \bar{a})$ is an exact or approximate normalizing constant; see Section 4 for further discussion. Further details of the p^* formula are given in Barndorff-Nielsen (1980; 1983), Barndorff-Nielsen and Cox (1994), McCullagh (1987) and Skovgaard (1990).

The log-likelihood $\ell(\omega)$ which appears in (2.4) is given by

$$\begin{aligned} \ell(\omega) &\equiv \ell(\omega; \hat{\omega}, \bar{a}) = n[\theta^T(\omega)t - K\{\theta(\omega)\}] \\ &= n[\theta^T(\omega)h(\hat{\omega}, \bar{a}) - K\{\theta(\omega)\}] \end{aligned} \quad (2.5)$$

where $t = h(\hat{\omega}, \bar{a})$ for some one-one function h . The $d \times d$ matrix $j(\hat{\omega})$ in (2.4) denotes the observed information function

$$j(\hat{\omega}) = j(\omega; \hat{\omega}, \bar{a}) = \frac{\partial^2 \ell(\omega; \hat{\omega}, \bar{a})}{\partial \omega \partial \omega^T}.$$

In general, the p^* formula provides a very good approximation to p . For transformation models, p^* is generally exact. For (m, d) exponential models that are not transformation models, it is known that p^* typically has relative error of order $O(n^{-3/2})$ on normal deviation regions, provided that a is ancillary to order $O(n^{-1})$; see Barndorff-Nielsen (1983) and Barndorff-Nielsen and Cox (1994). A detailed discussion of the accuracy of p^* in exponential models with $d = m$ is given by Barndorff-Nielsen and Cox (1979).

However, for $d < m$, there are no clear results in the literature on the accuracy of p^* in large deviation regions. This question is addressed in Theorem 3.1 below.

2.3. The r^* formula

We shall view $\omega = (\psi, \chi)$ as consisting of a scalar parameter of interest, ψ , and a vector, χ , of $d - 1$ nuisance parameters. We write $\hat{\chi}_\psi$ for the partial maximum likelihood estimator of χ when ψ is prescribed and held fixed. The directed, or signed, likelihood is defined by

$$r_\psi = \text{sign}(\hat{\psi} - \psi)[2\{\ell(\hat{\omega}) - \ell(\hat{\omega}_\psi)\}]^{1/2}, \quad (2.6)$$

where $\hat{\omega}_\psi = (\psi, \hat{\chi}_\psi)$, and the modified directed likelihood r_ψ^* is defined by

$$r_\psi^* = r_\psi + r_\psi^{-1} \log(u_\psi / r_\psi), \quad (2.7)$$

where

$$u_\psi = [[\ell_{\chi;\hat{\omega}}(\hat{\omega}_\psi)\ell_{;\hat{\omega}}(\hat{\omega}) - \ell_{;\hat{\omega}}(\hat{\omega}_\psi)] / \{ |j_{\chi\chi}(\hat{\omega}_\psi)| |j(\hat{\omega})| \}]^{1/2} \quad (2.8)$$

and $[.]$ is a partitioned matrix whose last column is given by

$$\ell_{;\hat{\omega}}(\omega) = \ell_{;\hat{\omega}}(\omega; \hat{\omega}, \bar{a}) = \left(\frac{\partial \ell(\omega; \hat{\omega}, \bar{a})}{\partial \hat{\omega}^q}; q = 1, \dots, d \right),$$

and the $d \times (d - 1)$ submatrix consisting of the first $d - 1$ columns is given by

$$\ell_{\chi;\hat{\omega}}(\omega) = \ell_{\chi;\hat{\omega}}(\omega; \hat{\omega}, \bar{a}) = \left(\frac{\partial^2 \ell(\omega; \hat{\omega}, \bar{a})}{\partial \chi^s \partial \hat{\omega}^q}; q = 1, \dots, d; s = 1, \dots, d - 1 \right).$$

An important point here is that \bar{a} is held fixed while the sample space differentiation $\partial/\partial\hat{\omega}$ is performed. Note that, apart from in (m, m) exponential models, it is essential to specify \bar{a} if we are to calculate u_ψ in (2.8). For further details of the notation used in (2.8), see Barndorff-Nielsen (1991) and Barndorff-Nielsen and Cox (1994).

2.4. Directed likelihood ancillaries

A procedure for constructing asymptotically normal ancillaries in (m, d) exponential models will now be described. This type of construction was suggested in Barndorff-Nielsen (1986); see also Barndorff-Nielsen and Cox (1994, p. 235).

As before, let θ be the canonical parameter of the ambient (m, m) exponential model. Consider an alternative parametrization (ω, η) of the ambient model, where ω has dimension d , η has dimension k and $m = d + k$. For convenience, we write $\theta = \theta(\omega, \eta)$. The idea is to choose the parameter η so that, for some fixed $\eta_{(0)}$, the constraint $\eta = \eta_{(0)}$ determines \mathcal{M}_Ω , the (m, d) exponential model under study. As before, we assume that $\omega = (\psi, \chi)$, where ψ is the scalar parameter of interest, and χ is a nuisance parameter vector of dimension $d - 1$. Our aim is to conduct inference for ψ under \mathcal{M}_Ω , the model determined by the constraint $\eta = \eta_{(0)}$, treating χ as a nuisance parameter vector. To construct r_ψ^* in (2.7), we need to specify an ancillary of dimension k which is kept fixed while the sample space differentiations in u_ψ are performed. We now explain how to construct two, mutually related, k -dimensional approximate ancillaries, a^* and a^o , based on directed likelihoods.

Consider the following sequence of hypotheses involving $\eta = (\eta^1, \dots, \eta^k)$ and $\eta_{(0)} = (\eta_{(0)}^1, \dots, \eta_{(0)}^k)$.

H_0 : ω unspecified, $\eta = \eta_{(0)}$.

H_j : ω unspecified, η^1, \dots, η^j unspecified,

$$\eta^{j+1} = \eta_{(0)}^{j+1}, \dots, \eta^k = \eta_{(0)}^k, \quad 1 \leq j \leq k - 1.$$

H_k : ω unspecified, η unspecified.

Observe that $H_0 \subset H_1 \subset \dots \subset H_k$, the dimension of H_j is $d + j$, and H_k corresponds to the ambient (m, m) exponential model.

The log-likelihood $\ell(\omega, \eta)$ for the ambient prime exponential model may be written

$$\ell(\omega, \eta) \equiv \ell(\omega, \eta; \hat{\eta}_{(k)}, \hat{\omega}_{(k)}) = n[\theta^T(\omega, \eta)\nabla_{\theta}K\{\theta(\hat{\omega}_{(k)}, \hat{\eta}_{(k)})\} - K\{\theta(\omega, \eta)\}], \quad (2.9)$$

where $\nabla_{\theta}K$ is the θ -gradient of K and $(\hat{\omega}_{(j)}, \hat{\eta}_{(j)})$ is the maximum likelihood estimator of (ω, η) under hypothesis H_j . The corresponding directed likelihoods are given by

$$r_j = \text{sign}(\hat{\eta}_{(j)}^j - \eta_{(0)}^j)[2\{\ell(\hat{\omega}_{(j)}, \hat{\eta}_{(j)}) - \ell(\hat{\omega}_{(j-1)}, \hat{\eta}_{(j-1)})\}]^{1/2}, \quad 1 \leq j \leq k. \quad (2.10)$$

We now define a sequence u_1, \dots, u_k in reverse order, using (2.8) repeatedly, as follows.

Step 1. Calculate u_k in (2.8) in the usual way, treating η_k as the interest parameter and $(\omega, \eta_1, \dots, \eta_{k-1})$ as the nuisance parameter vector of length $d + k - 1$. In this case there is no ancillary \bar{a} . Then put $r_k^* = r_k + r_k^{-1} \log(u_k/r_k)$, with r_k given in (2.10).

Step 2. For each $j = k - 1, k - 2, \dots, 1$, calculate u_j treating $(\omega, \eta_1, \dots, \eta_{j-1})$ as the nuisance parameter vector, η_j as the interest parameter, and keeping r_{j+1}^*, \dots, r_k^* fixed in the sample space differentiations; i.e. $\bar{a} = (\bar{r}_{j+1}^*, \dots, \bar{r}_k^*)$ in (2.8). Then put $r_j^* = r_j + r_j^{-1} \log(u_j/r_j)$.

Then the modified directed likelihood ancillary is defined by

$$\bar{a}^* = (\bar{r}_1^*, \dots, \bar{r}_k^*), \quad (2.11)$$

and the corresponding ancillary based directly on the sequence of unmodified directed likelihoods is given by

$$\bar{a}^o = (\bar{r}_1, \dots, \bar{r}_k), \quad (2.12)$$

where the presence of a bar indicates multiplication by $n^{-1/2}$.

Under the hypothesis H_0 : ω unspecified, $\eta = \eta_{(0)}$, where the constraint $\eta = \eta_{(0)}$ determines the model \mathcal{M}_{Ω} , both \bar{a}^* and \bar{a}^o are asymptotically multivariate normal with an $N_k(0, I)$ distribution, where I is the $k \times k$ identity matrix. However, in normal deviation regions, \bar{a}^* is typically standard normal to order $O(n^{-3/2})$, whereas \bar{a}^o is only standard normal to order $O(n^{-1/2})$.

Observe that the ancillaries \bar{a}^* and \bar{a}^o are not invariant with respect to the choice of the sequence of hypotheses $H_1 \subset \dots \subset H_k$.

We now digress briefly to explain the reason for introducing the bar in quantities such as \bar{a}^* and \bar{r}_{ψ}^* . Recall that the bar indicated multiplication by $n^{-1/2}$ so that, for example, $\bar{a}^* = n^{-1/2}a^*$ and $\bar{r}_{\psi}^* = n^{-1/2}r_{\psi}^*$. Consider the curved exponential log-likelihood $\ell(\omega, \eta_0; \hat{\eta}_{(k)}, \hat{\omega}_{(k)})$ defined in (2.9). Note that $(\hat{\omega}_{(k)}, \hat{\eta}_{(k)})$ is a minimal sufficient statistic, since H_k corresponds to the ambient exponential model.

In the construction of r_{ψ}^* , based on an ancillary a with an asymptotic $N_k(0, I)$ distribution, we first transform from $(\hat{\omega}_{(k)}, \hat{\eta}_{(k)})$ to $(\hat{\omega}, \bar{a})$, where $\hat{\omega}$ is the maximum likelihood estimator of ω under hypothesis H_0 . More explicitly, we may write

$$(\hat{\omega}_{(k)}, \hat{\eta}_{(k)}) = g(\hat{\omega}, \bar{a})$$

for some function g , where g is one–one for $(\hat{\omega}, \bar{a})$ at least in a small neighbourhood of $(\omega, 0)$. Then, because of the presence of the factor $n^{-1/2}$ in $\bar{a} = n^{-1/2}a$, both g and the inverse mapping g^{-1} have Jacobians whose norms are bounded for $(\hat{\omega}, \bar{a})$ in a neighbourhood of $(\omega, 0)$. A related point is that $n^{1/2}(\hat{\omega} - \omega)$ and a are asymptotically normal, whereas $\hat{\omega} - \omega$ and \bar{a} are both $O_p(n^{-1/2})$. Thus, when we consider how the log-likelihood and its derivatives depend on the ancillary, it is appropriate to view them as being functions of \bar{a} rather than a , at least in a repeated sampling framework. For related discussion, see Barndorff-Nielsen (1984, Section 1).

2.5. A class of first-order ancillaries

We now define a class of first-order ancillaries via Properties 2.1 and 2.2 below. A typical member of this class will be written $\bar{a} = n^{-1/2}a$.

Property 2.1. For each $\omega \in \mathcal{M}_\Omega$ there exists a function g , independent of n , such that $(\hat{\omega}, \bar{a}) = g(t)$, where t is the canonical statistic and g is smooth and one–one for all t which lie in a fixed neighbourhood of τ_ω , the expectation of t under ω . Note that this assumption implies that $n^{1/2}\{\bar{a}(t) - \alpha\} \rightarrow N_k(0, \Sigma)$ in distribution, where the limiting covariance matrix Σ has full rank k and $\alpha = \bar{a}(\tau_\omega)$.

Property 2.2. The quantities α and Σ defined above do not depend on ω .

Properties 2.1 and 2.2 have several implications which it will be useful to note. First, the implicit function theorem allows us to view the unmodified directed ancillary \bar{a}^o as a smooth function of $\hat{\omega}$ and \bar{a} , at least in a sufficiently small neighbourhood of the form $\|\hat{\omega} - \omega, \bar{a} - \alpha\| < \gamma$; we shall emphasize this by writing $\bar{a}^o = G(\hat{\omega}, \bar{a})$. Second, we may assume without loss of generality that

$$\begin{aligned} \bar{a}_q^o &= \bar{a}_q + \frac{1}{2!} G_q^{ij} \bar{a}_i \bar{a}_j + \frac{1}{3!} G_q^{ijk} \bar{a}_i \bar{a}_j \bar{a}_k \dots \\ &= \bar{a}_q + H_q^{ij} \bar{a}_i \bar{a}_j, \end{aligned} \tag{2.13}$$

where the summation convention has been used, the quantities $G_q^{ij}, G_q^{ijk}, \dots$ are smooth functions of $\hat{\omega}$ only, and the H_q^{ij} are smooth functions of $\hat{\omega}$ and \bar{a} .

An important type of ancillary which satisfies the above assumptions is the score ancillary, which when $k=1$ coincides with the Efron–Hinkley ancillary; see the continuation of Example 2.1 in Section 6 for an example involving the Efron–Hinkley ancillary, and Barndorff-Nielsen and Cox (1994) for further discussion. However, there are sensible approximate ancillaries which do not possess the properties indicated above; for example, the modified directed ancillary does not fall in this class because the corresponding function g depends on n .

We now discuss an example which we shall return to in Section 6.

Example 2.1. Define $Exp(\lambda)$ to be the exponential distribution with density $\lambda e^{-\lambda x}$ on $x > 0$, and let X_1, \dots, X_n and Y_1, \dots, Y_n be independent samples from the $Exp(\chi)$ and $Exp(\psi)$ distributions, respectively. We follow Barndorff-Nielsen (1980), Pedersen (1981) and Barndorff-Nielsen and Cox (1994, p. 229) and consider the (2,1) exponential model defined by imposing the constraint $\chi\psi e^\psi = 1$, and viewing χ as a function of ψ . The log-likelihood in the curved model is given by

$$\ell(\psi; x, y) = -n\{\chi(\psi)x + \psi y\},$$

where

$$x = n^{-1} \sum_{i=1}^n x_i, \quad y = n^{-1} \sum_{i=1}^n (y_i + 1),$$

and the maximum likelihood estimator $\hat{\psi}$ in this submodel is the unique solution of

$$e^\psi(1 + \psi^{-1}) = y/x. \tag{2.14}$$

The Efron–Hinkley ancillary is defined by $\hat{\gamma}^{-1}(\hat{j}/\hat{i} - 1)$, where i and j are the expected and observed information and γ is the Efron curvature in the (2,1) submodel, and the hat indicates evaluation of these quantities at the maximum likelihood estimator $\hat{\psi}$, the solution of (2.14).

As in the above references, we define the functions $\Psi_\pm(\psi) = \{1 + (1 + \psi)^{\pm 2}\}^{-1/2}$ and write $\hat{\Psi}_\pm$ for $\Psi_\pm(\hat{\psi})$. Direct calculation shows that the Efron–Hinkley ancillary is given by

$$\bar{a} = (y - 1 - \hat{\psi}^{-1})\hat{\psi}\hat{\Psi}_- \tag{2.15}$$

and the (unmodified) directed likelihood ancillary, \bar{a}^o , is given by

$$\frac{1}{2}(\bar{a}^o)^2 = (\hat{\Psi}_+ + \hat{\Psi}_-)\bar{a} - \log(1 + \hat{\Psi}_+\bar{a}) - \log(1 + \hat{\Psi}_-\bar{a}), \tag{2.16}$$

where $\bar{a}^o = \bar{a}^o(\hat{\psi})$, \bar{a} is chosen to have the same sign as \bar{a} .

3. Main results

Before going on to present our main results, we state two key assumptions.

Assumption 3.1 Preliminaries. The ambient model \mathcal{M} is a prime (m, m) exponential model. The parametrization (ω, η) is smooth in the sense that (ω, η) may be viewed as a one–one function of the canonical parameter θ with continuous partial derivatives up to and including those of the sixth order. The submodel obtained by fixing η at some prescribed value is written \mathcal{M}_Ω ; the dependence on η will be suppressed in our notation.

Assumption 3.2 Integrability of the characteristic function. For some $\theta \in \Theta^0$, there exists a $\nu = \nu(\theta) > 0$ such that

$$\int_{\tau \in \mathbb{R}^m} |\xi_\theta(\tau)|^{\nu(\theta)} d\tau < \infty, \tag{3.1}$$

where $\xi_\theta(\tau) = \exp \{K(\theta + i\tau) - K(\theta)\}$ is the characteristic function of the distribution with density (2.3) when the sample size, n , equals 1.

One consequence of Assumption 3.2 is the following: for each $\theta \in \Theta^0$ there exists an integer $n_0 = n_0(\theta)$ such that, when $n \geq n_0$, the distribution of the canonical statistic t is absolutely continuous with respect to Lebesgue measure on \mathbb{R}^m ; see Lemma 5.3 below.

Assumptions 3.1 and 3.2 are very mild smoothness requirements, appropriate for continuous models. It would be desirable to formulate, as far as possible, appropriate conditions for the purely discrete and ‘mixed’ cases, in which the densities are absolutely continuous with respect to counting measure, or a product of counting and Lebesgue measure, respectively.

The main results of the paper are now presented. Recall, from Section 2, that \bar{r}_ψ^* is the modified directed likelihood and \bar{r}_ψ the unmodified directed likelihood; and that \bar{a}^* is the modified directed likelihood ancillary and \bar{a}^o the unmodified directed likelihood ancillary, both based on a fixed sequence of hypotheses $H_1 \subset \dots \subset H_k$. We shall also denote by \bar{a} a typical member of the class of first-order ancillaries defined in Section 2.5.

To determine $r_\psi^* = r_\psi + r_\psi^{-1} \log(u_\psi/r_\psi)$, it is necessary to calculate sample space derivatives to obtain u_ψ ; see (2.8). In Theorem 3.3 below, we consider two alternative ways of calculating these sample space derivatives: (i) keeping \bar{a}^* fixed (i.e. taking $\bar{a} = \bar{a}^*$ in (2.8)); (ii) keeping \bar{a}^o fixed. We shall write the corresponding modified directed likelihoods as r_ψ^* and r_ψ^o respectively.

Theorem 3.1. *Suppose that Assumptions 3.1 and 3.2 hold. Then, in some large deviation region,*

$$p(\hat{\omega}; \omega|\bar{a}^*) = p^*(\hat{\omega}; \omega|\bar{a}^*)\{1 + O(n^{-1}\|\hat{\omega} - \omega\|) + O(n^{-2})\} \quad (3.2)$$

and

$$p(\hat{\omega}; \omega|\bar{a}^o) = p^*(\hat{\omega}; \omega|\bar{a}^o)\{1 + O(\|\bar{a}^o\|)O(\|\hat{\omega} - \omega\|) + O(n^{-1}\|\hat{\omega} - \omega\|) + O(n^{-2})\}. \quad (3.3)$$

Now consider a first-order ancillary \bar{a} which satisfies Properties 2.1 and 2.2 given in Section 2.5. Then, for $\hat{\omega}$ in some large deviation region and \bar{a} in some normal deviation region,

$$0 < A^{-1} < \liminf_{n \rightarrow \infty} \frac{p^*(\hat{\omega}; \omega|\bar{a})}{p(\hat{\omega}; \omega|\bar{a})} \leq \limsup_{n \rightarrow \infty} \frac{p^*(\hat{\omega}; \omega|\bar{a})}{p(\hat{\omega}; \omega|\bar{a})} < A, \quad (3.4)$$

where $A \in (0, \infty)$ is a constant.

Theorem 3.2. *If Assumptions 3.1 and 3.2 hold, then*

$$P_\omega(r_\psi^* > y|\bar{a}^*) = \{1 - \Phi(y)\}\{1 + O(n^{-3/2}y) + O(n^{-3/2})\}, \quad (3.5)$$

where the error terms are uniform for $y \in (y_0, n^{1/2}\delta)$. Here, $y_0 \in (-\infty, \infty)$ is fixed but arbitrary and $\delta > 0$ is sufficiently small.

Remark 3.1. In order to simplify the statements of Theorems 3.1, 3.2 and Corollary 3.4 below, we have left a certain technical detail implicit: that we only allow the canonical statistic t to range over a bounded open set which contains the true expected value of t . This detail may seem insignificant from a practical point of view (and, theoretically, the probability that t lies outside any such set is exponentially small). However, to obtain fully rigorous results in useful generality, it does appear that one needs to invoke a condition of this kind. The key problems are rigorous justification, in general settings, of (i) Laplace's approximation and (ii) conditional Temme (1982) tail-area approximations. Both (i) and (ii) can be justified in broad generality when t ranges over a bounded open set; but without this restriction on t , the technical difficulties seem to be intractable. In the proofs of Theorems 3.1 and 3.2 below, we do take explicit account of the need to restrict t to a bounded set; more detailed discussion is given in Barndorff-Nielsen and Wood (1995).

Remark 3.2. In Theorem 3.1, it is assumed that the constant c in (2.4) is calculated in a specific way:

$$c = \left\{ Lap_1 \int_{\mathbb{R}^d} |\hat{j}|^{1/2} \exp \{ \ell(\omega) - \ell(\hat{\omega}) \} d\hat{\omega} \right\}^{-1},$$

where Lap_1 denotes a two-term Laplace approximation to the integral and is explained in more detail in Section 4.

Remark 3.3. In (m, m) exponential models there is no need for an ancillary because $\hat{\omega}$ is minimal sufficient for ω . In this case, all reference to ancillaries should be omitted in the statements of Theorems 3.1 and 3.2, and, in particular, statement (3.4) is redundant.

Remark 3.4. The role of y_0 in Theorem 3.2 is purely cosmetic: its purpose is to rule out large deviations in the lower tail so as to simplify the specification of the error terms in (3.5). Since by hypothesis y is bounded below, it follows directly from (3.5) that the conditional probability of the complementary event $r_\psi^* \leq y$ is given by

$$P_\omega(r_\psi^* \leq y/\bar{a}^*) = \Phi(y) \{ 1 + O\{1 - \Phi(y)\} O(n^{-3/2}y + n^{-3/2}) \}.$$

Remark 3.5. Comparison of (3.3) with (3.2) shows that an additional relative error term

$$O(\|\bar{a}^o\|)O(\|\hat{\omega} - \omega\|)$$

has appeared in (3.3), although it should be noted that this term is always zero when $m = d$ and also in exponential transformation models, as defined in Barndorff-Nielsen and Cox (1994). The additional relative error term in (3.3) is of order $O(1)$ if both $\hat{\omega}$ and \bar{a}^o lie in large deviation regions, of order $O(n^{-1/2})$ if one lies in a large deviation region and the other lies in a normal deviation region, and of order $O(n^{-1})$ if both $\hat{\omega}$ and \bar{a}^o lie in normal deviation regions. Thus, the additional term will typically dominate the relative error in (3.3).

Remark 3.6. Typically, (3.4), which was anticipated by Skovgaard (1990), fails if \bar{a} varies over a large deviation region. See Section 6 for further discussion and an example.

Theorem 3.3. *Suppose that Assumption 3.1 holds. Then, in some large deviation region,*

$$\bar{r}_{\psi}^* - \bar{r}_{\psi}^o = O(n^{-2}).$$

An important implication of Theorems 3.2 and 3.3 may be summarized in the following corollary. Let us write $r_{\psi\text{obs}}^*$ and $r_{\psi\text{obs}}^o$ for the ‘observed’ values of r_{ψ}^* and r_{ψ}^o based on a notional sample.

Corollary 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then, for any $y_0 \in (-\infty, \infty)$ and some $\delta > 0$, we have*

$$P_{\omega}[r_{\psi}^o > r_{\psi\text{obs}}^o | \bar{a}^*] = \{1 - \Phi(r_{\psi\text{obs}}^o)\} \{1 + O(n^{-1} \bar{r}_{\psi\text{obs}}^o) + O(n^{-3/2})\}, \quad (3.6)$$

where the relative errors are uniformly of the stated order for $r_{\psi\text{obs}}^o \in (y_0, n^{1/2}\delta)$.

Proof. Using Theorems 3.2 and 3.3, we obtain

$$\begin{aligned} P_{\omega}[r_{\psi}^o > r_{\psi\text{obs}}^o | \bar{a}^*] &= P_{\omega}[r_{\psi}^* > r_{\psi\text{obs}}^o + r_{\psi}^* - r_{\psi}^o | \bar{a}^*] \\ &= P_{\omega}[r_{\psi}^* > r_{\psi\text{obs}}^o + O(n^{-3/2}) | \bar{a}^*] \\ &= 1 - \Phi\{r_{\psi\text{obs}}^o + O(n^{-3/2})\} \{1 + O(n^{-3/2} r_{\psi\text{obs}}^o) + O(n^{-3/2})\} \\ &= \{1 - \Phi(r_{\psi\text{obs}}^o)\} \{1 + O(n^{-1} \bar{r}_{\psi\text{obs}}^o) + O(n^{-3/2})\}, \end{aligned}$$

as required. □

It is clear that in general $\bar{r}_{\psi\text{obs}}^o$, obtained by keeping \bar{a}^o fixed in the sample space differentiations, will be substantially easier to calculate than $\bar{r}_{\psi\text{obs}}^*$, which is obtained by keeping \bar{a}^* fixed in the sample space differentiations. The importance of result (3.6) given in Corollary 3.4 is that, if we work with $\bar{r}_{\psi\text{obs}}^o$ instead of $\bar{r}_{\psi\text{obs}}^*$, the resulting inference can still be interpreted as being conditional on \bar{a}^* rather than \bar{a}^o , and in addition we do not increase the order of magnitude of the relative error in the standard normal approximation to the distribution of \bar{r}_{ψ}^* .

4. Laplace-spa calculus

The main element in our strategy for proving Theorems 3.1 and 3.2 is a careful analysis of certain related saddlepoint-style approximations. As mentioned in Section 1, a number of authors have contributed in the development of saddlepoint-style approximations in recent years. The main ingredients of these approximations are (i) ‘classical’ saddlepoint approximations; (ii) smooth nonlinear transformations; and (iii) the use of Laplace’s approximation to obtain approximations to marginal and conditional densities.

On a first impression, there is non-uniqueness in the specification of these approximations,

due the fact that Laplace's approximation can be implemented in different ways. However, it turns out that if one wishes the resulting approximations to possess certain desirable invariance properties, then there is only one way to proceed. In this section, we present explicit guidelines for the construction of invariant marginal and conditional density approximations, and the issue of normalization is also addressed. We refer to these developments as Laplace-spa calculus (where 'spa' is an abbreviation for 'saddlepoint approximation'). More detailed discussion is given in Barndorff-Nielsen and Wood (1995).

In all applications of Laplace's approximation considered in Section 5, it is always possible (perhaps after some manipulation) to write the integral to be approximated in a particular way: specifically, in the form

$$\int_{\mathbb{R}^m} D_{(n)}(x) e^{nI(x)} dx \tag{4.1}$$

where $D_{(n)} \sim D_0 + n^{-1}D_1 + \dots$, and the functions $I(x)$, $D_0(x)$, $D_1(x)$, ... do not depend on n .

Let \check{x} be a stationary maximum of $I(x)$. We define the *Lap* functionals

$$Lap_0 \int_{\langle \check{x} \rangle} D_n(x) e^{nI(x)} dx = n^{-m/2} e^{n\check{I}} (2\pi)^{m/2} |-\nabla\nabla^T\check{I}|^{-1/2} \check{D}_0 \tag{4.2}$$

and

$$Lap_1 \int_{\langle \check{x} \rangle} D_n(x) e^{nI(x)} dx = n^{-m/2} e^{n\check{I}} (2\pi)^{m/2} |-\nabla\nabla^T\check{I}|^{-1/2} \{\check{D}_0 + n^{-1}\check{d}_1\}, \tag{4.3}$$

where $\check{I} = I(\check{x})$, $\check{D}_0 = D_0(\check{x})$ and so on. The quantity \check{d}_1 , which is given explicitly by, for example, Shun and McCullagh (1995), is a function of \check{D}_1 , the first and second derivatives of D_0 , and the second, third and fourth derivatives of I , all evaluated at $x = \check{x}$. The region of integration is written as $\langle \check{x} \rangle$ to highlight the fact that the calculation of Lap_0 and Lap_1 depends only on the behaviour of the functions $I(x)$ and $D_{(n)}(x)$ in a neighbourhood of \check{x} .

Lap_0 and Lap_1 give, respectively, the one-term and two-term Laplace approximation to the integral (4.1). We have used explicit terminology for these widely used approximations partly because we need to be precise about what we mean by the approximation, and partly to emphasize that the right-hand sides of (4.2) and (4.3) are well defined even when the integral in (4.1) does not exist. An important consequence of insisting that the integral in (4.1) has the stated form is that the Lap_0 and Lap_1 approximations are invariant with respect to smooth one-one transformations of the form $y = g(x)$ defined in a neighbourhood of $x = \check{x}$. Further discussion of invariance is given in Barndorff-Nielsen and Wood (1995).

We now move on to saddlepoint approximations to densities. Let t denote the sample mean of a sum of independent and identically distributed random vectors of dimension m . Assume that the cumulant generating function, $K(\theta)$, of a single observation exists in a neighbourhood of $\theta = 0$, and so is analytic there. Denote the mean of t by μ .

The leading term in the saddlepoint expansion for the density of t may be written

$$p_0^{SP}(t) = (2\pi)^{-m/2} n^{-m/2} |\nabla_\theta \nabla_\theta^T K(\hat{\theta})|^{-1/2} e^{n\{K(\hat{\theta}) - \hat{\theta}^T t\}}$$

where $\nabla_{\theta}\nabla_{\theta}^{\top}K(\hat{\theta})$ denotes the Hessian of $K(\theta)$ evaluated at $\theta = \hat{\theta}$, and $\hat{\theta} = \hat{\theta}(t)$ is the unique solution to the saddlepoint equation $\nabla_{\theta}K(\theta) = t$. The saddlepoint approximation based on the first two leading terms in the saddlepoint expansion may be written

$$p_1^{\text{SP}}(t) = p_0^{\text{SP}}(t)\{1 + n^{-1}Q(t)\} \quad (4.4)$$

where Q is a certain polynomial function of t ; see, for example, McCullagh (1987, Chapter 6) where an expression for Q is given.

Now consider transformed variables $w = w(t)$ where w is a smooth, one-to-one function of t in a neighbourhood of $t = \mu$. The resulting saddlepoint approximations to the density of w may be written

$$p_i^{\text{SP}}(w) = p_i^{\text{SP}}(t)J(t; \omega), \quad i = 0, 1,$$

where $J(t; w)$ is the Jacobian of the transformation from t to w , and t and J are viewed as functions of w .

Marginal and conditional densities are now considered. More detailed discussion of some aspects may be found in Jensen (1995, Section 4.1). Write $w = (x, y)$, where x and y are subvectors of w . Natural definitions for saddlepoint-type approximations to the marginal density of y , and the conditional density of x given y , are, for $i = 0, 1$,

$$\begin{aligned} p_i^{\text{SP}}(y) &= \text{Lap}_i \int_{\langle \tilde{x}_y \rangle} p_i^{\text{SP}}(x, y) dx \\ p_i^{\text{SP}}(x|y) &= p_i^{\text{SP}}(x, y) / p_i^{\text{SP}}(y), \end{aligned} \quad (4.5)$$

where \tilde{x}_y is the locally unique maximizer of the exponent $K(\hat{\theta}) - \hat{\theta}^{\top}t$, viewed as a function of x and y with y fixed. Strictly speaking, we should use a notation which indicates that $p_i^{\text{SP}}(y)$ and $p_i^{\text{SP}}(x|y)$ have been obtained using Laplace's approximation; however, the expressions given are notationally convenient and should not cause confusion. It follows easily from the above definitions, and from basic properties of Laplace's approximation, that

$$p_1^{\text{SP}}(y) = p_0^{\text{SP}}(y)\{1 + n^{-1}Q_1(y) + O(n^{-2})\} \quad (4.6)$$

and

$$p_1^{\text{SP}}(x|y) = p_0^{\text{SP}}(x|y)\{1 + n^{-1}Q_2(x|y) + O(n^{-2})\}, \quad (4.7)$$

where Q_1 and Q_2 are smooth functions of their respective arguments.

Finally, we define the normalized saddlepoint approximations

$$p_0^{\text{NSP}}(y) = p_0^{\text{SP}}(y) / \text{Lap}_1 \int_{\langle \tilde{y} \rangle} p_0^{\text{SP}}(y) dy$$

and

$$p_0^{\text{NSP}}(x|y) = p_0^{\text{SP}}(x|y) / \text{Lap}_1 \int_{\langle \tilde{x}_y \rangle} p_0^{\text{SP}}(x|y) dx.$$

The motivation for normalization is the following:

$$p_1^{\text{SP}}(y) = p_0^{\text{NSP}}(y)\{1 + O(n^{-1}\|y - \check{y}\|) + O(n^{-2})\} \quad (4.8)$$

and

$$p_1^{\text{SP}}(x|y) = p_0^{\text{NSP}}(x|y)\{1 + O(n^{-1}\|x - \check{x}_y\|) + O(n^{-2})\}, \quad (4.9)$$

i.e. p_0^{NSP} is a better approximation to p_1^{SP} than is p_0^{SP} ; compare (4.8) with (4.6) and (4.9) with (4.7). The possibility of normalizing via Laplace's approximation was suggested by Skovgaard (1985, Remark 4.11). Detailed derivations of (4.8) and (4.9) are given in Barndorff-Nielsen and Wood (1995, Section 3.3).

The development in this section only requires that the cumulant generating function, K , exists in a neighbourhood of the origin, and that the transformation $w = w(t)$ be smooth and one-one for t in a neighbourhood of $\mu = \nabla_{\theta}K(0)$. Thus, the accuracy of the normalized saddlepoint approximation p_0^{NSP} is governed only by the accuracy of p_1^{SP} .

5. Proofs

5.1. Preliminary lemmas

Our first result concerns the difference $\bar{r}_{\psi}^* - \bar{r}_{\psi}$. Observe that on sets of the form $(\bar{r}_{\psi}, \bar{a}^*, \hat{\chi}_{\psi} - \chi) \in B_{\gamma}$ with $\gamma > 0$ sufficiently small, we may view \bar{r}_{ψ}^* as a function of \bar{r}_{ψ} , $\hat{\chi}_{\psi}$ and \bar{a}^* . However, \bar{r}_{ψ}^* and \bar{a}^* and \bar{u}_{ψ} , but not \bar{r}_{ψ} and $\hat{\chi}_{\psi}$, will depend on n , as noted in Section 2.5.

Lemma 5.1. *If Assumption 3.1 holds, then there exists a $\gamma = \gamma(\omega) > 0$ such that*

$$\bar{r}_{\psi}^* - \bar{r}_{\psi} = n^{-1}S_0(\bar{r}_{\psi}^*, \bar{a}^*, \hat{\chi}_{\psi}) + O(n^{-2}), \quad (5.1)$$

where the $O(n^{-2})$ term is uniformly of that order for $(\bar{r}_{\psi}^*, \bar{a}^*, \hat{\chi}_{\psi} - \chi) \in B_{\gamma}$. Moreover S_0 , which may be chosen to be independent of n , is a smooth function on B_{γ} .

Proof. It follows from the definition of u_{ψ} in (2.8), and Assumption 3.1, that we may view \bar{u}_{ψ} as a function of \bar{r}_{ψ} , \bar{a}^* and $\hat{\chi}_{\psi}$. Note that \bar{u}_{ψ} also depends on n , since \bar{u}_{ψ} depends on \bar{a}^* and \bar{a}^* depends on n . However, Taylor expansion of \bar{u}_{ψ} about $\bar{r}_{\psi} = 0$, keeping $\hat{\chi}_{\psi}$ and \bar{a}^* fixed, shows that we may write

$$\bar{u}_{\psi} = \bar{r}_{\psi} + \bar{r}_{\psi}^2 Q_1(\hat{\chi}_{\psi}, \bar{a}^*) + \bar{r}_{\psi}^3 Q_2(\hat{\chi}_{\psi}, \bar{a}^*) + O(\bar{r}_{\psi}^4) + O(n^{-1}\bar{r}_{\psi}^2), \quad (5.2)$$

where the functions Q_1 and Q_2 do not depend on n . Note that the dependence of \bar{u}_{ψ} on n has been absorbed into the $O(n^{-1}\bar{r}_{\psi}^2)$ remainder term in (5.2). Consequently,

$$\bar{r}_{\psi}^{-1}\bar{u}_{\psi} - 1 = \bar{r}_{\psi} Q_1(\hat{\chi}_{\psi}, \bar{a}^*) + \bar{r}_{\psi}^2 Q_2(\hat{\chi}_{\psi}, \bar{a}^*) + O(\bar{r}_{\psi}^3) + O(n^{-1}\bar{r}_{\psi}) \quad (5.3)$$

and

$$\bar{r}_{\psi}^{-1}(\bar{r}_{\psi}^{-1}\bar{u}_{\psi} - 1) = Q_1(\hat{\chi}_{\psi}, \bar{a}^*) + \bar{r}_{\psi} Q_2(\hat{\chi}_{\psi}, \bar{a}^*) + O(\bar{r}_{\psi}^2) + O(n^{-1}), \quad (5.4)$$

where the two remainder terms are smooth functions of \bar{r}_{ψ} , $\hat{\chi}_{\psi}$ and \bar{a}^* , and uniformly of the stated order, for $\|\bar{r}_{\psi}, \bar{a}^*, \hat{\chi}_{\psi} - \chi\|$ sufficiently small.

It follows that there exists a function S_0 , independent of n , such that

$$\begin{aligned} S_0(\bar{r}_\psi, \bar{a}^*, \hat{\chi}_\psi) &= \bar{r}_\psi^{-1} \log(\bar{u}_\psi / \bar{r}_\psi) + O(n^{-1}) \\ &= \bar{r}_\psi^{-1} \log\{1 + (\bar{r}_\psi^{-1} \bar{u}_\psi - 1)\} + O(n^{-1}) \\ &= \bar{r}_\psi^{-1} (\bar{r}_\psi^{-1} \bar{u}_\psi - 1) g(\bar{r}_\psi^{-1} \bar{u}_\psi - 1) + O(n^{-1}) \end{aligned} \quad (5.5)$$

where, for any $\lambda \in (0, 1)$, $g(x) = x^{-1} \log(1 + x)$ is analytic for $|x| \leq 1 - \lambda$. In view of (5.2)–(5.5), S_0 is smooth for $(\bar{r}_\psi, \bar{a}^*, \hat{\chi}_\psi - \chi) \in B_{2_\gamma}$ when $\gamma > 0$ is sufficiently small. Since

$$\bar{r}_\psi^* - \bar{r}_\psi = n^{-1} S_0(\bar{r}_\psi, \bar{a}^*, \hat{\chi}_\psi) + O(n^{-2}),$$

it follows that $|\bar{r}_\psi^* - \bar{r}_\psi| = O(n^{-1})$ uniformly on B_{2_γ} ; therefore, for $(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi - \chi) \in B_\gamma$ we may replace \bar{r}_ψ by \bar{r}_ψ^* in S_0 with uniform $O(n^{-2})$ error; hence (5.1) is valid.

Lemma 5.2. *Suppose that Assumption 3.1 holds. Then*

$$p^\dagger(r_\psi^*, \hat{\chi}_\psi; \omega | \bar{a}^*) = \phi(r_\psi^*) p^\dagger(\hat{\chi}_\psi; \chi | \bar{r}_\psi^*, \bar{a}^*) J(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi) \quad (5.6)$$

where

$$\begin{aligned} J &= J(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi) = e^{(r_\psi^* - r_\psi)^2/2} |\partial \bar{r}_\psi^* (\bar{r}_\psi, \bar{a}^*, \hat{\chi}_\psi) / \partial \bar{r}_\psi|^{-1} \\ &= 1 + n^{-1} S_1(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi) + O(n^{-2}) \end{aligned} \quad (5.7)$$

uniformly for $(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi - \chi) \in B_\gamma$ when $\gamma > 0$ is sufficiently small. Moreover, S_1 , which may be chosen to be independent of n , is a smooth function on B_γ . Note: p^\dagger is obtained by choosing $c = (2\pi)^{-d/2}$ in (2.4).

Proof. We first establish identity (5.6). From formulae (2.7) and (3.3) of Barndorff-Nielsen (1991) it follows that

$$|(\hat{\psi}, \hat{\chi})_{/(r_\psi, \hat{\chi}_\psi)}| = r_\psi u_\psi^{-1} |j_{\chi\chi}(\hat{\omega}_\psi)|^{1/2} |j(\hat{\omega})|^{-1/2},$$

where $|(\hat{\psi}, \hat{\chi})_{/(r_\psi, \hat{\chi}_\psi)}|$ is the determinant of the Jacobian of the transformation from $(\hat{\psi}, \hat{\chi})$ to $(r_\psi, \hat{\chi}_\psi)$. Therefore the determinant of the Jacobian of the transformation from $(\hat{\psi}, \hat{\chi})$ to $(\bar{r}_\psi^*, \hat{\chi}_\psi)$ is given by

$$r_\psi u_\psi^{-1} |j_{\chi\chi}(\hat{\omega}_\psi)|^{1/2} |j(\hat{\omega})|^{-1/2} |\partial \bar{r}_\psi (r_\psi, \bar{a}^*, \hat{\chi}_\psi) / \partial \bar{r}_\psi|^{-1}.$$

Writing $|\hat{j}_{\chi\chi}|$ and $|\hat{j}|$ for $|j_{\chi\chi}(\hat{\omega}_\psi)|$ and $|j(\hat{\omega})|$, respectively, we obtain

$$\begin{aligned} p^\dagger(r_\psi^*, \hat{\chi}_\psi; \omega | \bar{a}^*) &= r_\psi u_\psi^{-1} |\hat{j}_{\chi\chi}|^{1/2} |\hat{j}|^{-1/2} |\partial \bar{r}_\psi^* / \partial \bar{r}_\psi|^{-1} p^\dagger(\hat{\omega}; \omega | \bar{a}^*) \\ &= r_\psi u_\psi^{-1} |\hat{j}_{\chi\chi}|^{1/2} |\hat{j}|^{-1/2} |\partial \bar{r}_\psi^* / \partial \bar{r}_\psi|^{-1} (2\pi)^{-d/2} |\hat{j}|^{1/2} e^{\ell(\omega) - \ell(\hat{\omega}_\psi) - r_\psi^2/2} \\ &= r_\psi u_\psi^{-1} |\partial \bar{r}_\psi^* / \partial \bar{r}_\psi|^{-1} \phi(r_\psi) p^\dagger(\hat{\chi}_\psi; \omega | \bar{r}_\psi^*, \bar{a}^*) \\ &= \phi(r_\psi^*) p^\dagger(\hat{\chi}_\psi; \omega | \bar{r}_\psi^*, \bar{a}^*) e^{(r_\psi^* - r_\psi)^2/2} |\partial \bar{r}_\psi^* / \partial \bar{r}_\psi|^{-1} \\ &= \phi(r_\psi^*) p^\dagger(\hat{\chi}_\psi; \omega | \bar{r}_\psi^*, \bar{a}^*) J(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi), \end{aligned}$$

say. Using Lemma 5.1 we have

$$\begin{aligned} \partial \bar{r}_\psi^*(\bar{r}_\psi, \bar{a}^*, \hat{\lambda}_\psi) / \partial \bar{r}_\psi &= 1 + n^{-1} \partial S_0(\bar{r}_\psi, \bar{a}^*, \hat{\lambda}_\psi) / \partial \bar{r}_\psi \\ &= 1 + n^{-1} \partial S_0(\bar{r}_\psi^*, \bar{a}^*, \hat{\lambda}_\psi) / \partial \bar{r} + O(n^{-2}) \end{aligned} \quad (5.8)$$

and

$$e^{(r_\psi^* - r_\psi)^2 / 2} = 1 + n^{-1} S_0^2(\bar{r}_\psi^*, \bar{a}^*, \hat{\lambda}_\psi) / 2 + O(n^{-2}). \quad (5.9)$$

Thus (5.7) follows directly from Lemma 5.1, (5.8) and (5.9). \square

Lemma 5.3. *Suppose that Assumptions 3.1 and 3.2 hold and let $\bar{a}^?$ denote one of \bar{a}^* , \bar{a}^o or \bar{a} . Then there exists a $\gamma = \gamma(\omega) > 0$ such that*

$$p(\hat{\omega}, \bar{a}^?; \omega) = p_1^{\text{SP}}(\hat{\omega}, \bar{a}^?; \omega) \{1 + O(n^{-2})\}$$

where the $O(n^{-2})$ term is uniform for $(\hat{\omega} - \omega, \bar{a}^?) \in B_\gamma$.

Proof. A convenient reference is Jensen (1995); see, in particular, Lemmas 2.3.4 and 2.3.5 given there. \square

Lemma 5.4. *Let \bar{a}^* , \bar{a}^o be as defined in (2.11) and (2.12), and let us view \bar{a}^o as a function of $\hat{\omega}$ and \bar{a}^* , and write $\bar{a}^o = h(\hat{\omega}, \bar{a}^*)$. If Assumption 3.1 holds then there exists a $\gamma = \gamma(\omega) > 0$ such that $\|\bar{a}^* - \bar{a}^o\| = O(n^{-1})$, where the $O(n^{-1})$ error is uniform for $(\hat{\omega} - \omega, \bar{a}^*) \in B_\gamma$. Moreover, if g (not depending on n) is a differentiable function of $\hat{\omega}$ and \bar{a}^o , and if we define*

$$f(\hat{\omega}, \bar{a}^*) \equiv g\{\hat{\omega}, h(\hat{\omega}, \bar{a}^*)\} \quad \text{for } (\hat{\omega} - \omega, \bar{a}^*) \in B_\gamma,$$

then

$$\frac{\partial f}{\partial \hat{\omega}} = \frac{\partial g}{\partial \hat{\omega}} + n^{-1} G,$$

where the quantity $G = G(\hat{\omega}, \bar{a}^*)$ has dimension d and satisfies $\|G\| = O(1)$ uniformly for $(\hat{\omega} - \omega, \bar{a}^*) \in B_\gamma$. The functions h , f and G (but not g) will depend on n . More specifically, $h = h_0 + n^{-1} h_1 + O(n^{-2})$, where h_0 and h_1 do not depend on n , with similar formulae for f and G .

Proof. The first part of the lemma follows from $k = m - d$ applications of Lemma 5.1.

Write $\bar{a}^o = \bar{a}^* + n^{-1} A$, where $\|A\| = O(1)$ uniformly for $(\hat{\omega} - \omega, \bar{a}^*) \in B_\gamma$. Interpreting $\hat{\omega}$, \bar{a}^o , \bar{a}^* and A as row vectors and differentiating, we obtain

$$\begin{aligned} \frac{\partial f}{\partial \hat{\omega}} &= \frac{\partial g}{\partial \hat{\omega}} + \frac{\partial g}{\partial \bar{a}^o} \frac{\partial (\bar{a}^o)^T}{\partial \hat{\omega}} \\ &= \frac{\partial g}{\partial \hat{\omega}} + \frac{\partial g}{\partial \bar{a}^o} \left\{ n^{-1} \frac{\partial A^T}{\partial \hat{\omega}} + \frac{\partial (\bar{a}^*)^T}{\partial \hat{\omega}} \right\} \\ &= \frac{\partial g}{\partial \hat{\omega}} + \frac{\partial g}{\partial \bar{a}^o} \left\{ n^{-1} \frac{\partial A^T}{\partial \hat{\omega}} + 0 \right\} \\ &= \frac{\partial g}{\partial \hat{\omega}} + n^{-1} G, \quad \text{where } G = \frac{\partial g}{\partial \bar{a}^o} \frac{\partial A^T}{\partial \hat{\omega}}. \end{aligned}$$

The statement that $\|G\| = O(1)$ uniformly for $(\hat{\omega} - \omega, \bar{a}^*) \in B_\gamma$ follows since $G = G(\hat{\omega}, \bar{a}^*)$ is a continuous function of its arguments. \square

5.2. Proof of Theorem 3.1

As noted in Remark 3.1, we shall assume that t varies over a fixed bounded open set which contains the expected value, μ of t . To be specific, and without loss of generality, we shall assume that this bounded set is of the form

$$\Xi_\gamma = \{t: \|t - \mu\| < \gamma\},$$

where $\gamma > 0$ is suitably small. There is no reason, other than convenience, for choosing Ξ_γ to be a ball; other types of set could equally well be used. A convenient way to incorporate the restriction on t is to condition on the event $t \in \Xi_\gamma$. We will follow this approach, and it should be understood that, strictly speaking, we are approximating the density $p(\hat{\omega}; \omega|\bar{a}^2, t \in \Xi_\gamma)$ rather than the density $p(\hat{\omega}; \omega|\bar{a}^2)$ in Theorem 3.1.

The preliminary step in the proof is to establish that under Assumptions 3.1 and 3.2,

$$p(\hat{\omega}; \omega|\bar{a}^2, t \in \Xi_\gamma) = p_0^{\text{NSP}}(\hat{\omega}; \omega|\bar{a}^2)\{1 + O(n^{-1}\|\hat{\omega} - \omega\|) + O(n^{-2})\}$$

uniformly for $\|\hat{\omega} - \omega, \bar{a}^2\| < \gamma_1$, with γ and γ_1 sufficiently small. Viewing $\hat{\omega}$ and \bar{a}^2 as functions of the canonical statistic t , we define the sets

$$D_\gamma = \{(\hat{\omega}(t), \bar{a}^2(t)): t \in \Xi_\gamma\}, \quad E_\gamma = \{\bar{a}^2(t): t \in \Xi_\gamma\}$$

and, for each $\bar{a}^2 \in E_\gamma$,

$$F_\gamma(\bar{a}^2) = \{\hat{\omega}: (\hat{\omega}, \bar{a}^2) \in D_\gamma\}.$$

Then

$$\begin{aligned} p(\hat{\omega}; \omega|\bar{a}^2, t \in \Xi_\gamma) &= \frac{p(\hat{\omega}, \bar{a}^2)}{\int_{F_\gamma(\bar{a}^2)} p(\hat{\omega}, \bar{a}^2; \omega) d\hat{\omega}} \\ &= \frac{p_1^{\text{SP}}(\hat{\omega}, \bar{a}^2; \omega)\{1 + O(n^{-2})\}}{\{1 + O(n^{-2})\} \text{Lap}_1 \int_{F_\gamma(\bar{a}^2)} p(\hat{\omega}, \bar{a}^2; \omega) d\hat{\omega}} \\ &= \frac{p_1^{\text{SP}}(\hat{\omega}; \bar{a}^2; \omega)\{1 + O(n^{-2})\}}{\text{Lap}_1 \int_{F_\gamma(\bar{a}^2)} p_1^{\text{SP}}(\hat{\omega}, \bar{a}^2; \omega)\{1 + O(n^{-2})\} d\hat{\omega}} \\ &= \frac{p_1^{\text{SP}}(\hat{\omega}, \bar{a}^2; \omega)\{1 + O(n^{-2})\}}{p_1^{\text{SP}}(\bar{a}^2; \omega)\{1 + O(n^{-2})\}} \\ &= p_1^{\text{SP}}(\hat{\omega}; \omega|\bar{a}^2)\{1 + O(n^{-2})\} \\ &= p_0^{\text{NSP}}(\hat{\omega}; \omega|\bar{a}^2)\{1 + O(n^{-1}\|\hat{\omega} - \omega\|) + O(n^{-2})\}, \end{aligned} \quad (5.10)$$

as required.

The argument leading to (5.10) requires some further explanation. When γ and γ_1 are sufficiently small, Assumption 3.1 implies the following: for each $\bar{a}^? \in D_\gamma$, $F_\gamma(\bar{a}^?)$ is a bounded, open, connected subset of \mathbb{R}^d with a smooth boundary. Consequently, the first equality leading to (5.10) is legitimate. The second equality follows from an application of Lemma 5.3 to the numerator; and the Lap_1 approximation to the integral in the denominator can be justified rigorously since the region of integration is a bounded set; see, for example, Jensen (1995a, Theorem 3.1.3). The third equality follows from another application of Lemma 5.3, while the fourth and fifth steps follow from definitions given in Section 4. The final line follows from (4.9).

The remainder of the proof is divided into three parts, corresponding to statements (3.2), (3.3) and (3.4).

Proof of (3.2). Using Lemma 5.2 repeatedly ($k = m - d$ times, to be precise) we have

$$p^\dagger(\hat{\omega}, a^*; \omega) = p^\dagger(\hat{\omega}; \omega | \bar{a}^*) \left\{ \prod_{q=1}^k J_q \right\} \left\{ \prod_{q=1}^k \phi(a_q^*) \right\} \quad (5.11)$$

where $J_1 \dots, J_k$ are the versions of the factor J defined in (5.7) corresponding to the hypotheses $H_1 \subset \dots \subset H_k$ specified in Section 2.4. Write $D_1 = \prod_{q=1}^k J_q$. Repeated application of Lemma 5.2 also tells us that

$$D_1 = D_1(\hat{\omega}, \bar{a}^*) = 1 + n^{-1} S_2(\hat{\omega}, \bar{a}^*) + O(n^{-2}) \quad (5.12)$$

uniformly for $(\bar{a}^*, \hat{\omega} - \omega) \in B_\gamma$ for sufficiently small $\gamma > 0$. But, in view of the fact that $p_0^{\text{SP}}(\hat{\omega}, \bar{a}^*; \omega) = p^\dagger(\hat{\omega}, \bar{a}^*; \omega)$, we also have

$$p_0^{\text{SP}}(\hat{\omega}, a^*; \omega) = D_1(\hat{\omega}, \bar{a}^*) p^\dagger(\hat{\omega}; \omega | \bar{a}^*) \left\{ \prod_{q=1}^k \phi(a_q^*) \right\}.$$

In view of (5.12), we have

$$Lap_0 \int p^\dagger(\hat{\omega}; \omega | \bar{a}^*) D_1(\hat{\omega}, \bar{a}^*) d\hat{\omega} = 1.$$

The above result is precisely 1 because our definition of Lap_0 , given in Section 4, is such that all $O(n^{-1})$ and higher-order terms are excluded. Consequently, the $p_0^{\text{SP}}(\bar{a}^*; \omega)$ approximation to the marginal density $p(\bar{a}^*; \omega)$ is given by

$$p_0^{\text{SP}}(\bar{a}^*; \omega) = \prod_{q=1}^k \phi(a_q^*),$$

and the p_0^{SP} approximation to the conditional distribution $p(\hat{\omega}; \omega | \bar{a}^*)$ is given by

$$p_0^{\text{SP}}(\hat{\omega}; \omega | \bar{a}^*) = \frac{p_0^{\text{SP}}(\hat{\omega}, \bar{a}^*; \omega)}{p_0^{\text{SP}}(\bar{a}^*; \omega)} = p^\dagger(\hat{\omega}; \omega | \bar{a}^*) D_1(\hat{\omega}, \bar{a}^*). \quad (5.13)$$

Write $\tilde{c}_\omega^{\text{SP}}$ for the Lap_1 -normalizing factor for $p_0^{\text{SP}}(\hat{\omega}; \omega | \bar{a}^*)$ and let \tilde{c}_ω denote the Lap_1 -

normalizing factor for $p^\dagger(\hat{\omega}; \omega|\bar{a}^*)$. The relationship between $\tilde{c}_\omega^{\text{SP}}$ and \tilde{c}_ω is obtained by application of the Lap_1 approximation to the integral over $\hat{\omega}$ of (5.13), and is given by

$$\begin{aligned} \{\tilde{c}_\omega^{\text{SP}}\}^{-1} &= \tilde{c}_\omega^{-1} + n^{-1}S_2(\omega, \bar{a}^*) \\ &= \tilde{c}_\omega^{-1}D_1(\omega, \bar{a}^*)\{1 + O(n^{-1})\}, \end{aligned} \quad (5.14)$$

where S_2 and D_1 are the functions defined in (5.12), and we used the fact that $\tilde{c}_\omega = 1 + n^{-1}C_\omega$ for some C_ω . Since (5.12) implies that

$$D_1(\hat{\omega}, \bar{a}^*)/D_1(\omega, \bar{a}^*) = 1 + O(n^{-1}\|\hat{\omega} - \omega\|),$$

it follows from (5.13) and (5.14) that

$$\begin{aligned} p_0^{\text{NSP}}(\hat{\omega}; \omega|\bar{a}^*) &\equiv \tilde{c}_\omega^{\text{SP}} p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a}^*) \\ &= \tilde{c}_\omega D_1^{-1}(\omega, \bar{a}^*)\{1 + n^{-1}R(\bar{a}^*)\}^{-1}\{1 + O(n^{-2})\} \\ &\quad \times p^\dagger(\hat{\omega}; \omega|\bar{a}^*)D_1(\hat{\omega}, \bar{a}^*)\{1 + n^{-1}R(\bar{a}^*)\} \\ &= p^*(\hat{\omega}; \omega|\bar{a}^*)\{1 + O(n^{-1}\|\hat{\omega} - \omega\|) + O(n^{-2})\}, \end{aligned} \quad (5.15)$$

where $p_0^{\text{NSP}}(\hat{\omega}; \omega|\bar{a}^*)$ and $p^*(\hat{\omega}; \omega|\bar{a}^*)$ are the Lap_1 -normalized versions of $p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a}^*)$ and $p^\dagger(\hat{\omega}; \omega|\bar{a}^*)$, respectively.

Conclusion (3.2) follows directly from (5.10), with $\bar{a}^\dagger = \bar{a}^*$, and (5.15).

Proof of (3.3). Since the argument leading to (3.3) is a straightforward modification of the argument leading to (3.2), we only give brief details.

The analogue of (5.11) is

$$p^\dagger(\hat{\omega}, a^o; \omega) = p^\dagger(\hat{\omega}; \omega|\bar{a}^o)D_2(\hat{\omega}; \bar{a}^o) \left\{ \prod_{q=1}^k \phi(a_q^o) \right\},$$

where

$$D_2 = \prod_{j=1}^k \frac{a_j^o}{u_j},$$

in which the quantities a_j^o are the unmodified directed likelihoods defined in Section 2.4, but the u_j^o are calculated keeping the a_j^o , rather than the a_j^* , fixed in the required sample space differentiations. Following the reasoning in the proof of (3.2), we obtain

$$p_0^{\text{NSP}}(\hat{\omega}; \omega|\bar{a}^o) = p^*(\hat{\omega}; \omega|\bar{a}^o) \frac{D_2(\hat{\omega}, \bar{a}^o)}{D_2(\omega; \bar{a}^o)}. \quad (5.16)$$

The form of expansion (5.2) implies that

$$D_2(\hat{\omega}, \bar{a}^o) = D_2(\omega; \bar{a}^o)\{1 + O(\|\bar{a}^o\|)O(\|\hat{\omega} - \omega\|)\}. \quad (5.17)$$

Then (3.3) is obtained by combining (5.10) with $\bar{a}^\dagger = \bar{a}^o$, (5.16) and (5.17).

Proof of (3.4). To establish (3.4), it will be sufficient to compare $p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a})$ and $p^\dagger(\hat{\omega}; \omega|\bar{a})$ in a region of the form

$$\{\|\hat{\omega} - \omega\| < \gamma, \|\bar{a}\| < n^{-1/2}C\}, \quad (5.18)$$

with $\gamma > 0$ sufficiently small and $C > 0$ arbitrary. The desired conclusion will then follow an application of (5.10) with $\bar{a}^? = \bar{a}$. The first point to note is that we only need to study the exponents $\ell - \hat{\ell}$ and $n(m - \hat{m})$ in, respectively,

$$p^\dagger(\hat{\omega}; \omega|\bar{a}) = (2\pi)^{-d} |\hat{j}|^{1/2} e^{\ell - \hat{\ell}} \quad \text{and} \quad p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a}) = (2\pi)^{-d} J e^{n(m - \hat{m})},$$

since the factors $|\hat{j}|^{1/2}$ and J will be uniformly bounded away from zero and infinity in a region of the form (5.18). The quantity $m = m(\hat{\omega}, \bar{a})$ is defined in the next paragraph, and $\hat{m} = \hat{m}(\bar{a}) = \sup_{\hat{\omega}: \|\hat{\omega} - \omega\| < \gamma} m(\hat{\omega}, \bar{a})$.

Now consider the parametrization $(\hat{\omega}, \bar{a}^o)$ for the ambient (m, m) exponential model, where \bar{a}^o is an unmodified directed likelihood ancillary for which (2.13) is satisfied. The $(\hat{\omega}, \bar{a}^o)$ parametrization is smooth and one-one, at least in a region for which $\|\hat{\omega} - \omega, \bar{a}^o\| < \gamma$ with $\gamma > 0$ sufficiently small. Define

$$\begin{aligned} m^o(\hat{\omega}, \bar{a}^o) &= n^{-1}[\ell(\omega, 0; \hat{\omega}, \bar{a}^o) - \ell(\hat{\omega}_1, \bar{a}^o; \hat{\omega}, \bar{a}^o)] \\ &= n^{-1}[\ell(\omega, 0; \hat{\omega}, \bar{a}^o) - \ell(\hat{\omega}, 0; \hat{\omega}, \bar{a}^o)] - \frac{1}{2} \sum_{q=1}^k (\bar{a}_q^o)^2, \end{aligned}$$

where $\hat{\omega}$ is the maximum likelihood estimate of ω in the d -dimensional submodel \mathcal{M}_Ω as before, and $\hat{\omega}_1 = \hat{\omega}_1(\hat{\omega}, \bar{a}^o)$ is the maximum likelihood estimate of ω in the ambient (m, m) exponential model. The second line above follows from the definition of the unmodified directed likelihood ancillary $\bar{a}^o = (\bar{a}_1^o, \dots, \bar{a}_k^o)$. Recall from (2.13) that we may view \bar{a}^o as a function of $\hat{\omega}$ and \bar{a} , at least for those $(\hat{\omega}, \bar{a})$ which satisfy $\|\hat{\omega} - \omega, \bar{a}\| < \gamma_1$ with $\gamma_1 > 0$ sufficiently small. In such a region we may define

$$m(\hat{\omega}, \bar{a}) = m^o\{\hat{\omega}, \bar{a}^o(\hat{\omega}, \bar{a})\} = n^{-1}[\ell(\omega, 0; \hat{\omega}, \bar{a}^o) - \ell(\hat{\omega}, 0; \hat{\omega}, \bar{a}^o)] - \frac{1}{2} \sum_{q=1}^k (\bar{a}_q^o)^2,$$

with $\bar{a}^o = \bar{a}^o(\hat{\omega}, \bar{a})$. Ignoring terms which are $O(n^{-1})$, and using the definition of $p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a})$ implied by the discussion in Section 4, we have

$$\begin{aligned} n^{-1} \log \left\{ \frac{p^\dagger(\hat{\omega}; \omega|\bar{a})}{p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a})} \right\} &= n^{-1}[\ell(\omega, 0; \hat{\omega}, \bar{a}^o) - \ell(\hat{\omega}, 0; \hat{\omega}, \bar{a}^o)] - m(\hat{\omega}, \bar{a}) + \sup_{\hat{\omega}} m(\hat{\omega}, \bar{a}) \\ &= \frac{1}{2} \sum_{q=1}^k (\bar{a}_q^o)^2 + \sup_{\hat{\omega}} m(\hat{\omega}, \bar{a}). \end{aligned} \quad (5.19)$$

Now $\sup_{\hat{\omega}} m(\hat{\omega}, \bar{a}) = O(n^{-1})$ in a region of the form (5.18). Therefore, since (2.13) implies that

$$\begin{aligned} \frac{1}{2} \sum_{q=1}^k (\bar{a}_q^o)^2 &= \frac{1}{2} \sum_{q=1}^k \bar{a}_q^2 + O(n^{-3/2}) \\ &= O(n^{-1}), \end{aligned}$$

where the error terms are uniformly of the stated order for $(\hat{\omega}, \bar{a})$ in regions of the form (5.18), it follows that the absolute value of (5.19) is also uniformly of order $O(n^{-1})$. Hence the ratio $p^\dagger(\hat{\omega}; \omega|\bar{a})/p_0^{\text{SP}}(\hat{\omega}; \omega|\bar{a})$ stays bounded away from zero and infinity in regions of the form (5.18), and (3.4) follows. \square

5.3. Proof of Theorem 3.2

The set $G_\gamma(\bar{a}^*)$, referred to below, is defined analogously to the set $F_\gamma(\bar{a}^*)$ which appears at the beginning of the proof of Theorem 3.1. Under Assumptions 3.1 and 3.2, we obtain

$$\begin{aligned}
 p(r_\psi^*; \omega|\bar{a}^*, t \in \Xi_\gamma) &= \frac{p(r_\psi^*, a^*; \omega)}{\int_{G_\gamma(\bar{a}^*)} p(r_\psi^*, a^*; \omega) \mathrm{d}r_\psi^*} \\
 &= \frac{p_1^{\text{SP}}(r_\psi^*, a^*; \omega)\{1 + O(n^{-2})\}}{\{1 + O(n^{-2})\} \text{Lap}_1 \int_{G_\gamma(\bar{a}^*)} p(r_\psi^*, a^*; \omega) \mathrm{d}r_\psi^*} \\
 &= \frac{p_1^{\text{SP}}(r_\psi^*, a^*; \omega)\{1 + O(n^{-2})\}}{\text{Lap}_1 \int_{G_\gamma(\bar{a}^*)} p_1^{\text{SP}}(r_\psi^*, a^*; \omega)\{1 + O(n^{-2})\} \mathrm{d}r_\psi^*} \\
 &= \frac{p_1^{\text{SP}}(r_\psi^*, a^*; \omega)\{1 + O(n^{-2})\}}{p_1^{\text{SP}}(a^*; \omega)\{1 + O(n^{-2})\}} \\
 &= p_1^{\text{SP}}(r_\psi^*; \omega|\bar{a}^*)\{1 + O(n^{-2})\} \\
 &= p_0^{\text{NSP}}(r_\psi^*; \omega|\bar{a}^*)\{1 + O(n^{-1}\bar{r}_\psi^*) + O(n^{-2})\}. \tag{5.20}
 \end{aligned}$$

The steps leading to (5.20) are rather similar to the steps leading to (5.10), and the comments given below (5.10) are also relevant here.

The next step is to show that

$$p_0^{\text{NSP}}(r_\psi^*; \omega|\bar{a}^*) = \phi(r_\psi^*). \tag{5.21}$$

Using the argument leading to (5.11), we obtain

$$\begin{aligned}
 p_0^{\text{SP}}(r_\psi^*, a^*; \omega) &= \phi(r_\psi^*) \left\{ \prod_{q=1}^{m-d} \phi(a^*_q) \right\} \\
 &= \phi(r_\psi^*) p_0^{\text{SP}}(a^*; \omega).
 \end{aligned}$$

Thus $p_0^{\text{SP}}(r_\psi^*; \omega|\bar{a}^*) = \phi(r_\psi^*)$ and, since $\text{Lap}_1 \int_{\{0\}} \phi(r_\psi^*) \mathrm{d}r_\psi^* = 1$, (5.21) follows.

Let $z = z(\gamma, \bar{a}^*)$ denote the upper end point of the interval $G_\gamma(\bar{a}^*)$ multiplied by $n^{1/2}$. Suppose also that y is such that $n^{-1/2}y$ is bounded away from zero, but is ‘sufficiently small’. Then (3.5) is a consequence of the following:

$$\begin{aligned}
 P_\omega[r_\psi^* > y | \bar{a}^*, t \in \Xi_\gamma] &= \int_y^z p(r_\psi^*; \omega | \bar{a}^*, t \in \Xi_\gamma) dr_\psi^* \\
 &= \int_y^z p_0^{\text{NSP}}(r_\psi^*; \omega | \bar{a}^*) \{1 + O(n^{-1} \bar{r}_\psi^*) + O(n^{-2})\} dr_\psi^* \\
 &= \int_y^z \phi(r_\psi^*) \{1 + O(n^{-1} \bar{r}_\psi^*) + O(n^{-2})\} dr_\psi^* \\
 &= \{\Phi(z) - \Phi(y)\} \{1 + O(n^{-2})\} + O\left(n^{-3/2} \int_y^z |r_\psi^*| \phi(r_\psi^*) dr_\psi^*\right) \\
 &= \{1 - \Phi(y)\} \{1 + O(n^{-2})\} + \{\phi(y) - \phi(z)\} \times O(n^{-3/2}) \\
 &= \{1 - \Phi(y)\} \{1 + \{M(y)\}^{-1} O(n^{-3/2}) + O(n^{-2})\} \\
 &= \{1 - \Phi(y)\} \{1 + O(n^{-3/2} y) + O(n^{-3/2})\},
 \end{aligned}$$

since $M(y) = \{1 - \Phi(y)\} / \phi(y)$ is bounded away from zero and infinity on compact sets, and is $O(y^{-1})$ as $y \rightarrow \infty$. Note that the second and third lines of the above are consequences of (5.20) and (5.21), respectively. \square

5.4. Proof of Theorem 3.3

The proof of Theorem 3.3 boils down to a careful application of Lemma 5.4. Consider two alternative representations of the minimal sufficient statistic $(\hat{\omega}_{(k)}, \hat{\eta}_{(k)})$: $(\hat{\omega}, \bar{a}^*)$ and $(\hat{\omega}, \bar{a}^o)$. As before, $\hat{\omega}$ is the maximum likelihood estimator of ω under the hypothesis $H_0: \eta = \eta_{(0)}$, and \bar{a}^* , \bar{a}^o are the approximate ancillaries defined in (2.11) and (2.12), respectively. Assume that $\omega_{(1)}$ and $\omega_{(2)}$ are two values of ω such that

$$\ell(\omega_{(2)}, \eta_{(0)}; \hat{\omega}_{(k)}, \hat{\eta}_{(k)}) \geq \ell(\omega_{(1)}, \eta_{(0)}; \hat{\omega}_{(k)}, \hat{\eta}_{(k)}),$$

and define

$$\begin{aligned}
 R^*(\omega_{(1)}, \omega_{(2)}) &\equiv R^*(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^*) \\
 &\equiv \text{sign}(\omega_{(2)1} - \omega_{(1)1}) [2\{\ell(\omega_{(2)}, \eta_{(0)}) - \ell(\omega_{(1)}, \eta_{(0)})\}]^{1/2}. \quad (5.22)
 \end{aligned}$$

Observe that if we put $\omega_{(2)} = \hat{\omega}$ and $\omega_{(1)} = \hat{\omega}_\psi$, then R^* is a directed likelihood. Define R^o in exactly the same way, but as a function of \bar{a}^o rather than \bar{a}^* . Note that the right-hand side of (5.22) does not depend on the *specification* of the ancillary, and that $R^*(\omega_{(1)}, \omega_{(2)}) = R^o(\omega_{(1)}, \omega_{(2)})$ even though they are different functions. Consider also the ‘nuisance’ scores $\ell_\chi^*(\omega_{(1)}) = \ell_\chi^o(\omega_{(1)})$, viewed as depending on \bar{a}^* and \bar{a}^o , respectively, and put

$$v^*(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^*) \equiv (\ell_\chi^*(\omega_{(1)}), R^*(\omega_{(1)}, \omega_{(2)}))$$

and

$$\nu^o(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^o) \equiv (\mathcal{L}_\chi^o(\omega_{(1)}), R^o(\omega_{(1)}, \omega_{(2)})).$$

Then

$$\nu^*(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^*) = \nu^o(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^o).$$

Let

$$H^*(\omega_{(1)}, \omega_{(2)}) \equiv H^*(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^*)$$

and

$$H^o(\omega_{(1)}, \omega_{(2)}) \equiv H^o(\omega_{(1)}, \omega_{(2)}; \hat{\omega}, \bar{a}^o)$$

denote the $d \times d$ matrices of first-order derivatives of ν^* and ν^o with respect to $\hat{\omega}$, keeping respectively \bar{a}^* and \bar{a}^o fixed. Then it follows from the definitions of the quantities concerned that

$$\bar{u}_\psi^*/\bar{r}_\psi = |H^*(\hat{\omega}_\psi, \hat{\omega})|/\{|j_{\chi\chi}(\hat{\omega}_\psi)\|j(\hat{\omega})\}^{1/2} \quad (5.23)$$

and

$$\bar{u}_\psi^o/\bar{r}_\psi = |H^o(\hat{\omega}_\psi, \hat{\omega})|/\{|j_{\chi\chi}(\hat{\omega}_\psi)\|j(\hat{\omega})\}^{1/2}, \quad (5.24)$$

where \bar{u}_ψ^* is (2.8) based on \bar{a}^* , and \bar{u}_ψ^o is (2.8) based on \bar{a}^o . It follows from (5.23) and (5.24) that

$$\begin{aligned} \bar{r}_\psi^* - \bar{r}_\psi^o &= n^{-1}\bar{r}_\psi^{-1}\{\log(\bar{u}_\psi^*/\bar{r}_\psi) - \log(\bar{u}_\psi^o/\bar{r}_\psi)\} \\ &= n^{-1}\bar{r}_\psi^{-1}\log(\bar{u}_\psi^*/\bar{u}_\psi^o) \\ &= n^{-1}\bar{r}_\psi^{-1}\log\{|H^*(\hat{\omega}_\psi, \hat{\omega})|/|H^o(\hat{\omega}_\psi, \hat{\omega})|\}. \end{aligned} \quad (5.25)$$

The theorem will be proved once it has been established that (5.25) is $O(n^{-2})$ uniformly for $(\bar{r}_\psi^*, \bar{a}^*, \hat{\chi}_\psi - \chi) \in B(\gamma)$, for some $\gamma > 0$. Using Lemma 5.4, noting that repeated application of Lemma 5.1 implies that $\|\bar{a}^* - \bar{a}^o\| = O(n^{-1})$, we have

$$H^*(\hat{\omega}_\psi, \hat{\omega}) = H^o(\hat{\omega}_\psi, \hat{\omega}) + n^{-1}G(\psi)$$

and

$$H^*(\hat{\omega}, \hat{\omega}) = H^o(\hat{\omega}, \hat{\omega}) + n^{-1}G(\hat{\psi}),$$

where $G(\psi) \equiv G(\psi; \hat{\omega}, \bar{a}^o)$ is $O(1)$. But

$$H^*(\hat{\omega}, \hat{\omega}) = H^o(\hat{\omega}, \hat{\omega}) = \{|j_{\chi\chi}(\hat{\omega}_\psi)\|j(\hat{\omega})\}^{1/2},$$

which implies that $G(\hat{\psi}; \hat{\omega}, \bar{a}^o) = 0$. Consequently, differentiability implies that each element of $G(\psi)$ is $O(|\hat{\psi} - \psi|)$, and therefore

$$\begin{aligned} H^*(\hat{\omega}_\psi, \hat{\omega}) &= H^o(\hat{\omega}_\psi, \hat{\omega}) + n^{-1}G(\psi) \\ &= H^o(\hat{\omega}_\psi, \hat{\omega})\{I + n^{-1}(\psi - \hat{\psi})G_1(\psi)\}, \end{aligned} \quad (5.26)$$

where I is the identity matrix, each element of

$$G_1(\psi) \equiv (\psi - \hat{\psi})^{-1} \{H^o(\hat{\omega}_\psi, \hat{\omega})\}^{-1} G(\psi)$$

is $O(1)$, and we have used the fact that H^o has full rank for $(\bar{r}_\psi^*, \bar{a}^o, \hat{\lambda}_\psi - \chi) \in B(\gamma)$ and sufficiently small $\gamma > 0$. From (5.26), it follows that

$$\begin{aligned} \frac{|H^*(\hat{\omega}_\psi, \hat{\omega})|}{|H^o(\hat{\omega}_\psi, \hat{\omega})|} &= |I + n^{-1}(\psi - \hat{\psi})G_1(\psi)| \\ &= 1 + n^{-1}(\psi - \hat{\psi}) \operatorname{tr} \{G_1(\psi)\} + O\{n^{-2}(\psi - \hat{\psi})^2\}. \end{aligned} \quad (5.27)$$

The desired conclusion then follows after substituting (5.27) into (5.25), Taylor-expanding the log, and noting that, as $\psi \rightarrow \hat{\psi}$, $\bar{r}_\psi \sim (\psi - \hat{\psi})c$ for some fixed, non-zero c .

6. Discussion

We conclude this paper with a discussion of the relationship between the results in this paper and the work of Skovgaard (1990) and Jensen (1992).

Skovgaard (1990) approached the p^* approximation from a novel and in some respects rather illuminating perspective. At the end of that paper he gave a heuristic discussion of the relative error in p^* in large deviation regions when a general approximate ancillary \bar{a} is used. His main conclusion was that p^* has bounded relative error in a large deviation region for $\hat{\omega}$ and any normal deviation region for \bar{a} . His conclusion is confirmed by our Theorem 3.1; see (3.4).

However, Skovgaard's (1990) discussion does not allow us to draw clear conclusions about what happens to the relative error in p^* in large deviation regions for the ancillary. Our findings, given in statements (3.2) and (3.3) of Theorem 3.1, tell us that under Assumptions 3.1 and 3.2 the relative error in the p^* approximation stays bounded in a large deviation region when either \bar{a}^* or \bar{a}^o is used as the approximate ancillary. This still leaves open the question of what happens to the relative error of the p^* approximation in large deviation regions for a general approximate ancillary \bar{a} . We now show that, for any first-order ancillary \bar{a} in the class defined in Section 2.4, the p^* approximation typically (but not always, and in particular not if $\bar{a} = \bar{a}^o$) breaks down when \bar{a} , rather than \bar{a}^o or \bar{a}^* , is used. Due to the special cases which can occur, it is a little awkward to specify precise necessary and sufficient conditions for p^* to break down; for this reason, we shall be content with giving a simple sufficient condition.

Consider (5.19) with \bar{a}^o viewed as a function of $\hat{\omega}$ and \bar{a} . Since $\sup_{\hat{\omega}} m(\hat{\omega}, \bar{a})$ depends only on \bar{a} , it follows from (5.19) that the $p^*(\hat{\omega}; \omega | \bar{a})$ approximation to $p(\hat{\omega}; \omega | \bar{a})$ will have unbounded relative error in a large deviation region $\{\|\hat{\omega} - \omega\| < \gamma_1, \|\bar{a}\| < \gamma_2\}$ if, for some fixed \bar{a} in this region,

$$\sup_{\hat{\omega}: \|\hat{\omega} - \omega\| < \gamma_1} \sum_{q=1}^k (\bar{a}_q^o)^2 - \inf_{\hat{\omega}: \|\hat{\omega} - \omega\| < \gamma_1} \sum_{q=1}^k (\bar{a}_q^o)^2 \geq \varepsilon, \quad (6.1)$$

for some fixed number $\varepsilon > 0$. In view of (2.13) and the smoothness of the functions

concerned, a sufficient condition for (6.1) to occur is the following: for some $q \in \{1, \dots, k\}$ and for some fixed $C \in (0, \infty)$,

$$\left\| \frac{\partial \bar{a}_q^o}{\partial \hat{\omega}} \right\| > C \|\bar{a}\|^2 \quad (6.2)$$

for all $\hat{\omega}$ and \bar{a} which satisfy $\|\hat{\omega} - \omega\| < \gamma_1$ and $\|\bar{a}\| < \gamma_2$. Further calculations, which we do not present here, show that under condition (6.2) a sufficient condition for the $p^*(\hat{\omega}; \omega|\bar{a})$ approximation to have unbounded relative error in a region of the form

$$\{\|\hat{\omega} - \omega\| < n^{-\delta_1} C_3, \|\bar{a}\| < n^{-\delta_2} C_4\} \quad (6.3)$$

is

$$3\delta_1 + \delta_2 < 1. \quad (6.4)$$

So, to summarize: if (6.2) is satisfied then the relative error in the $p^*(\hat{\omega}; \omega|\bar{a})$ approximation to $p(\hat{\omega}; \omega|\bar{a})$ does not stay bounded in regions of the type indicated by (6.3) and (6.4).

Example 2.1 (continued). We now give an example in which (6.2) is satisfied by the Efron–Hinkley ancillary. Taylor expansion of (2.16) gives

$$\frac{1}{2}(\bar{a}^o)^2 = \frac{1}{2}\bar{a}^2 - \frac{1}{3}\bar{a}^3(\hat{\Psi}_+^3 + \hat{\Psi}_-^3) + O(\bar{a}^4),$$

from which we obtain

$$\bar{a}^o = \bar{a} - \frac{1}{3}\bar{a}^2(\hat{\Psi}_+^3 + \hat{\Psi}_-^3) + O(\bar{a}^3). \quad (6.5)$$

Finally, differentiating $\bar{a}^o(\hat{\psi}, \bar{a})$ with respect to $\hat{\psi}$, we obtain

$$\frac{\partial \bar{a}^o}{\partial \hat{\psi}} = -\hat{\psi}(1 + \hat{\psi})\hat{\Psi}_+^5 \bar{a}^2 + O(\bar{a}^3)$$

which establishes condition (6.2), since $\hat{\psi}$ and $\hat{\Psi}_+$ are both positive.

Pedersen (1981) compared the Efron–Hinkley ancillary, \bar{a} in (2.15), and the unmodified directed likelihood ancillary, \bar{a}^o in (2.16), in a detailed numerical study. His main finding was that \bar{a}^o is clearly superior to \bar{a} in this example, and he reported that \bar{a}^o ‘is almost exactly distribution constant, and has more stable conditional likelihood functions’. The results in this paper provide clear theoretical insight into Pedersen’s empirical findings.

We now discuss what happens in Corollary 3.4 if a general approximate ancillary \bar{a} , rather than \bar{a}^* or \bar{a}^o , is used. Once again, it is awkward to state precise conditions because of the special cases which arise. However, suppose that \bar{a} satisfies

$$\|\bar{a}^o - \bar{a}\| = O(\|\bar{a}\|^2)$$

uniformly on sufficiently small large deviation regions, as is generally the case for the Efron–Hinkley or score ancillary; see (2.13) and, for example, (6.5). Let r^a denote the modified directed likelihood obtained by keeping \bar{a} , rather than \bar{a}^* or \bar{a}^o , fixed in the required sample space differentiations. A reworking of Lemma 5.4 and Theorem 3.3 under Assumptions 3.1 and 3.2 and condition (6.2) shows that the analogue of Corollary 3.4 will ‘typically’ be

$$\begin{aligned}
 P_\omega[r_\psi^* > r_{\psi\text{obs}}^a | \bar{a}^* t \in \Xi_\gamma] &= P_\omega[r_{\psi\text{obs}}^a | \bar{a}^*, t \in \Xi_\gamma] \\
 &\quad \times \{1 + O(\|\bar{a}\|^2 \bar{r}_{\psi\text{obs}}^a) + O(n^{-1} \bar{r}_{\psi\text{obs}}^a) + O(n^{-3/2})\} \\
 &= \{1 - \Phi(r_{\psi\text{obs}}^a)\} \\
 &\quad \times \{1 + O(\|\bar{a}\|^2 \bar{r}_{\psi\text{obs}}^a) + O(n^{-1} \bar{r}_{\psi\text{obs}}^a) + O(n^{-3/2})\}.
 \end{aligned}$$

Thus the relative error is typically dominated by the term of order $O(\|\bar{a}\|^2 \bar{r}_\psi^a)$ on sufficiently small large deviation regions.

We now consider the relationship between the findings of this paper and those of Jensen (1992). His main result, stated in our notation, is that in curved exponential models

$$P_\omega(\rho_\psi^* > \rho_{\psi\text{obs}}^* | \bar{a}^o, t \in \Xi_\gamma) = \{1 - \Phi(\rho_{\psi\text{obs}}^*)\} \{1 + O(n^{-1} \bar{\rho}_{\psi\text{obs}}^*) + O(n^{-3/2})\}, \quad (6.6)$$

where we have written ρ_ψ^* for Jensen's (1992) modification of the directed likelihood. In general, ρ_ψ^* will be different to r_ψ^* and r_ψ^o . Jensen (1995) shows that they are different in Example 2.1 above.

From a theoretical point of view the standard normal approximation to the distributions of r_ψ^* , r_ψ^o and ρ_ψ^* will have similar accuracy in both normal and large deviation regions; compare (3.5), (3.6) and (6.6). Moreover, to establish the validity of (6.6) in settings where conditioning on an ancillary takes place, Jensen's (1992) approach also requires the introduction of a conditioning set Ξ_γ . (Note: we are not claiming that the introduction of Ξ_γ is necessary in a strictly logical sense, but without it there does not seem to be any prospect of providing fully rigorous justification of either (6.6) or our results stated in Section 3.) The main difference between the approaches is that in (3.5) and (3.6) the conditioning statistic is a^* , whereas in (6.6) it is a^o . Since a^* is ancillary to order $O(n^{-3/2})$ on normal deviation regions and to order $O(n^{-1})$ on large deviation regions, while a^o is only ancillary in a somewhat weaker sense, it follows that conditioning on a^* may be more satisfactory from the perspective of conditional inference. From the point of view of computation, r_ψ^o and ρ_ψ^* seem to be of roughly the same order of difficulty.

Finally, we mention two very recent papers which have points of contact with this paper. Skovgaard (1996) introduces a new approximation to the modified directed likelihood, the significant point being that calculation of this quantity is feasible in general parametric models. In the process, Skovgaard (1996) derives a result which is essentially equivalent to (3.3) above. In the other paper, Jensen (1996) derives large deviation properties of the statistic r_ψ^o via a different route to the one followed here.

Acknowledgements

We thank Jens Ledet Jensen and two referees for a careful reading of the paper and many helpful comments. ATAW is grateful to the Department of Theoretical Statistics at Aarhus University for hospitality and financial support in connection with a visit.

References

- Barndorff-Nielsen, O.E. (1980) Conditionality resolutions. *Biometrika*, **67**, 293–310.
- Barndorff-Nielsen, O.E. (1983) On a formula for the conditional distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
- Barndorff-Nielsen, O.E. (1984) On conditionality resolution and the likelihood ratio for curved exponential models. *Scand. J. Statist.*, **11**, 157–170. Corrigendum: *Scand. J. Statist.*, **12**(1985), 191.
- Barndorff-Nielsen, O.E. (1986) Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307–322.
- Barndorff-Nielsen, O.E. (1990) Approximate interval probabilities. *J. Roy. Statist. Soc. Ser. B*, **52**, 485–496.
- Barndorff-Nielsen, O.E. (1991) Modified signed log likelihood ratio. *Biometrika*, **78**, 557–563.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B*, **41**, 279–312.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*. London: Chapman & Hall.
- Barndorff-Nielsen, O.E. and Schmidli, H. (1995) Saddlepoint approximations for the probability of ruin in finite time. *Scand. Actuar. J.*, 169–186
- Barndorff-Nielsen, O.E. and Wood, A.T.A. (1995) On large deviations and choice of ancillary for p^* and the modified directed likelihood. Research Report 299, Department of Theoretical Statistics, University of Aarhus.
- Daniels, H.E. and Young, G.A. (1991) Saddlepoint approximations for the studentized mean, with an application to the bootstrap. *Biometrika*, **78**, 169–179.
- DiCiccio, T.J., Field, C.A. and Fraser, D.A.S. (1990) Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77–95.
- DiCiccio, T.J. and Martin, M.A. (1991) Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika*, **78**, 891–902.
- Fraser, D.A.S. (1988) Normed likelihood as saddlepoint approximation. *J. Multivariate Anal.*, **27**, 181–193.
- Fraser, D.A.S. (1990) Tail probabilities from observed likelihoods. *Biometrika*, **77**, 65–76.
- Jensen J.L. (1992) The modified signed likelihood statistic and saddlepoint approximations. *Biometrika*, **79**, 693–703.
- Jensen, J.L. (1995) *Saddlepoint Approximations*. Oxford: Oxford University Press.
- Jensen, J.L. (1997) A simple derivation of r^* for curved exponential families. *Scand. J. Statist.*, **24**, 33–46.
- Jing, B.-Y. and Robinson, J. (1994) Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Ann. Statist.* **22**, 1115–1132.
- McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman & Hall.
- Pedersen, B.V. (1981) A comparison of the Efron–Hinkley ancillary and the likelihood ratio ancillary in a particular example. *Ann. Statist.* **9**, 1328–1333.
- Reid, N. (1988) Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213–238.
- Shun, Z. and McCullagh, P. (1995) Laplace approximation of high dimensional integrals. *J. Roy. Statist. Soc. Ser. B*, **57**, 749–760.
- Skovgaard, I.M. (1985) Large deviation approximations for maximum likelihood estimators. *Probab. Math. Statist.*, **6**, 89–107.

- Skovgaard, I.M. (1990) On the density of minimum contrast estimators. *Ann. Statist.*, **18**, 779–789.
- Skovgaard, I.M. (1996) An explicit large deviation approximation to one-parameter tests. *Bernoulli*, **2**(2), 145–165.
- Temme, N.M. (1982) The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.*, **13**, 239–253.
- Tierney, L., Kass, R.E. and Kadane, J.B. (1989). Approximate marginal densities of non-linear functions. *Biometrika*, **76**, 425–433.

Received November 1995 and revised August 1996