Ancillaries and third order significance

D.A.S. Fraser
Department of Mathematics and Statistics
York University
North York, Canada, M3J 1P3

N. Reid
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

ABSTRACT. For a variable and parameter of the same dimension, the tangent exponential model (Fraser, 1988) approximates an asymptotic model to third order in a first derivative neighborhood of the data point and to second order otherwise. For the more usual case of a variable of larger dimension than the parameter, we obtain a unique expression for the third order ancillary distribution as projected to the observed maximum likelihood surface, obtain the tangent directions for a second order ancillary, and then show that third order inference needs only the observed likelihood and the tangent directions for a second order ancillary. These results are then combined and a unique third order distribution is obtained for testing a component parameter; for the case of a real parameter component a simple expression is obtained for the third order observed significance level.

1 Introduction

We consider the derivation and calculation of accurate approximations to significance probabilities for scalar and vector parameters. In the very specialized case where the dimension of the variable is the same as that of the parameter, possibly after a sufficiency or ancillaries reduction, approximations to significance probabilities having third order accuracy have been available. For a scalar parameter and scalar variable, see Lugannani & Rice (1980) for exponential models, DiCiccio, Field & Fraser (1990) for location

Utilitas Mathematica 47(1995), pp. 33-53

models, and Barndorff-Nielsen (1991) and Fraser & Reid (1993; also see a 1990 University of Toronto Technical Report) for general models. For a vector parameter with a scalar parameter of interest, Skovgaard (1987) and Fraser & Reid (1993) discuss approximations for exponential models in the canonical parametrization, in which nuisance parameters are eliminated by conditioning; DiCiccio, Field & Fraser (1990) discuss approximations for component canonical parameters in transformation models in which nuisance parameters are eliminated by marginalization; and Barndorff-Nielsen (1991) discusses inference for component parameters in general models assuming a preliminary reduction to the same dimension case by ancillaries is possible. These types of approximations are reviewed in Pierce & Peters (1992).

In the case when sufficiency and ancillarity do not reduce the dimension of the variable to that of the parameter some alternative reduction method such as approximate ancillarity seems needed in order to apply available methods. While some general constructions of approximate ancillaries have been suggested (Barndorff-Nielsen, 1980; McCullagh, 1987, Ch. 8) the issue is usually not addressed in the recent development of approximations to tail areas (see, for example, the reply to the discussion in Pierce and Peters, 1992) and feasible methods are lacking. For second order inference however the expansions in Section 2 indicate that exponential model theory without ancillary calculations can be used; see for example DiCiccio & Martin (1993) and Barndorff-Nielsen & Chamberlin (1994).

In this paper we consider a general construction of an approximate ancillary statistic and the subsequent derivation of significance probabilities having third order accuracy for scalar or vector parameters. We use two types of recently constructed approximating models; tangent exponential models (Section 2) and tangent location models (Section 5). The tangent location model provides a preliminary reduction by conditioning, and the tangent exponential model then gives significance for scalar parameters.

A general formula for calculating third order significance for a scalar parameter χ, based on a likelihood $\ell_*(\chi; y)$ relative to a scalar variable y is

$$p(\chi) = Pr(\hat{\chi} \leq \hat{\chi}^0) = \Phi(R) + \phi(R)(1/R - 1/Q) = \Phi(R, Q), \quad (1.1)$$

where R and Q here are the signed likelihood ratio r and maximum likelihood departure q,

$$r = sgn(\hat{\chi}^0 - \chi)[2\{\ell_*(\hat{\chi}^0) - \ell_*(\chi)\}]^{1/2}, \; q = (\hat{\beta}^0 - \beta)\hat{\jmath}_{\beta\beta}^{1/2}, \quad (1.2)$$

and β is a nominal parameter $(d/dy)\ell_*(\chi; y)|_{y^0}$ with corresponding observed information $\hat{\jmath}_{\beta\beta} = -(\partial^2/\partial\beta^2)\ell_*(\beta)|_{\hat{\beta}^0}$. Most available third order formulas can be expressed as special cases of this general formula; for some recent implementations of the general formula see Cheah, Fraser & Reid (1995).

A key formula for obtaining third order approximations to significance probabilities is the p^* formula of Barndorff-Nielsen (1980, 1983, 1988a); its simplicity belies its remarkable importance and usefulness in asymptotic inference. The formula intrinsically uses a likelihood function obtained from a sample space with dimension equal to that of the parameter, say d. For parameter θ and likelihood $\ell(\theta; y)$ the approximation is

$$f(\hat{\theta};\theta)d\hat{\theta} \doteq \frac{c}{(2\pi)^{d/2}} exp\{\ell(\theta;y) - \ell(\hat{\theta};y)\}|\hat{\jmath}_{\theta\theta}|^{1/2}d\hat{\theta} \qquad (1.3)$$

where c is a constant to $O(n^{-3/2})$ and is equal to 1 to $O(n^{-1})$. The formula applies in more general contexts conditional on an ancillary, and in these cases c is free of the ancillary to third order.

In some contexts we will encounter a modifed likelihood $\ell_*(\chi)$ for a scalar parameter χ and a derived density of the form

$$\frac{c}{(2\pi)^{/12}} exp\{\ell_*(\chi;y) - \ell_*(\hat{\chi};y)\}\delta(y;\chi)\hat{\jmath}_{\chi\chi}^{1/2}d\hat{\chi} \qquad (1.4)$$

where the adjustment factor $\delta(y; \chi)$ typically incorporates nuisance parameter and parameter curvature effects; it is assumed that the adjustment factor can be standardized so $\delta(y; \hat{\chi}) = 1$ and then that $\delta(y; \chi) = 1+O(n^{-1/2})$. A direct integration by parts gives (1.1) with relative error $O(n^{-3/2})$ where

$$R = r, \qquad Q = q/\delta,$$

and r, q given by (1.2): this uses $\hat{\jmath}_{\chi\chi}^{1/2} d\hat{\chi} = (r/q)dr$. This approximation places the adjustment δ in Q, and is analogous to the double saddlepoint approximation of Skovgaard (1987); for details see Cheah, Fraser & Reid (1995). An alternative approach incorporates the adjustment into the likelihood function, as in the sequential saddlepoint approximation of Fraser, Reid & Wong (1991). Numerical work mentioned in Pierce & Peters (1992), suggests that the sequential saddlepoint may be more accurate, especially if there are a large member of nuisance parameters, but this is still an open issue. Typically a numerical intergration of (1.4) is not possible as the needed values are available only at the observed data point.

The general formula (1.1), for an exponential model with a single canonical parameter, gives the Lugannani & Rice (1980) approximation, and for a location model gives the DiCiccio, Field & Fraser (1990) approximation. For general models it has the form proposed in Fraser (1988, 1990) and Fraser & Reid (1993) and corresponds closely to that proposed by Barndorff-Nielsen (1988b, 1990, 1991).

In Section 2 we provide background on the tangent exponential model and show that it is a generalization of the p^* formula of Barndorff-Nielsen (1983) and has an inverse form with special integration properties. This is used in

Section 3 to develop a unique distribution for a third order ancillary and in Section 6 to obtain significance levels for testing a component parameter.

In Section 4, it is shown that for third order inference the only needed information concerning the ancillary is the tangent directions to the ancillary surface at the data point; these will be called the ancillary directions and designated typically by an array $V = (v_1, \ldots, v_d)$ of tangent vectors. Also we show that it suffices to obtain the ancillary directions only for a second order ancillary.

In Seetion 5 we use location model theory to determine a first order ancillary at the data point, and then show that a second order ancillary has the same tangent directions at that data point; this gives the needed ancillary directions V for a second order ancillary. These can then be used with Sections 2-4 to show that simple likelihood and likelihood gradient information fully determines a tangent exponential model (2.3) for the conditional distribution given a third order ancillary. For inference concerning a component parameter of interest ψ, we derive in Section 6 a unique distribution that is free of the nuisance parameter; this uses ancillary results of Section 3 applied to the restricted model with fixed ψ. In the tangent exponential model this distribution manifests itself as a lower dimension exponential model with an adjustment factor as described in (1.4). This leads, for a scalar component parameter, to a simple expression for third order inference. For a vector interest parameter sequential analysis-of-variance type testing is mentioned briefly in the discussion at the end of Section 6.

2 Tangent exponential model

The tangent exponential model was developed in Fraser (1988, 1990) for a model with variable and parameter of the same dimension d. We will see that the tangent exponential model calculated for ancillary directions derived in Section 5 produces third order inference for component parameters. It is also used to obtain the ancillary distribution in Section 3.

Consider a continuous model $f(y;\theta)$ with variable y and parameter θ of dimension d: we assume that $\hat{\theta}(y)$ is unique and is $O_p(n^{-1/2})$ about the true θ, that $\ell(\theta;y) = \log f(y;\theta)$ is $O(n)$ for fixed θ and for fixed y, that $\ell(\theta;y)$ is repeatedly differentiable with respect to θ and y, and that $(n^{-1})\ell_{\theta\theta}$, is bounded from zero where the subscript θ denotes differentiation with respect to θ. We use column vectors for y and row vectors for θ.

Asymptotic properties of the tangent model approximation under these assumptions were recorded in Fraser & Reid (1993). We expand the log density in a Taylor series in y and θ about the observed data point y^0, and the corresponding maximum likelihood estimate $\hat{\theta}^0 = \hat{\theta}(y^0)$. There is then a transformation of the variable and a transformation of the parameter for which the expansion of the log-likelihood function in the new coordinates is nearly exponential and can be presented in a standardized form. For the case $d = 1$, this expansion of the log likelihood $\ell(\theta; y) + (1/2)\log(2\pi) = \Sigma a_{ij}(\theta - \hat{\theta}^0)^i(y - y^0)^j/(i!j!)$ has coefficients a_{ij}, $i, j = 0, \dots, 4$ given by the matrix

$$\begin{pmatrix} \frac{3\alpha_4 - 5\alpha_3^2 - 12c}{24n} & \frac{\alpha_3}{2n^{1/2}} & \left\{1 + \frac{\alpha_4 - 2\alpha_3^2 - 5c}{2n}\right\} & \frac{\alpha_3}{n^{1/2}} & \frac{\alpha_4 - 3\alpha_3^2 - 6c}{n} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \frac{c}{n} & - & - \\ -\frac{\alpha_3}{n^{1/2}} & 0 & - & - & - \\ -\frac{\alpha_4}{n} & - & - & - & - \end{pmatrix} \qquad (2.1)$$

where omitted elements are $O(n^{-3/2})$. We use c here for a particular standardized fourth derivative and the context will clarify the distinction from c as a normalizing constant.

Without the first row elements, the first column gives the likelihood function at the data point and the second column gives the sample space derivative of the likelihood function at the data point. The model expressed in (2.1) is an exponential model to order $O(n^{-1})$, and is an exponential model to order $O(n^{-3/2})$ in a first derivative neighborhood of the data point, except for the constant $-12c/24n$. The *tangent exponential model* is the exponential model obtained by setting $c = 0$ in (2.1); it is uniquely determined by the first two columns of (2.1) without the first row elements.

The expansion for $d > 1$ (Cakmak, Fraser & Reid, 1994) uses results from Fraser & Reid (1993) and is the obvious generalization of (2.1). The model is exponential to $O(n^{-1})$, and to $O(n^{-3/2})$ differs from an exponential model by terms of order $O(n^{-1})$: these terms are of the form $c_{ijab}y^iy^j\theta^a\theta^b/n$ plus terms that are constant, pure quadratic, and pure quartic in the variable y. Third and fourth order arrays replace α_3 and α_4 in (2.1).

Expansion (2.1) and its generalization to the multivariate case provides a simple and transparent proof of the p^* formula (1.3). For if the expansion is taken about a different point in the range of the standardized variable. it is seen that α_3^2, α_4, and c are constant to $O(n^{-1/2})$, and thus that the norming constant obtained in the top left corner is constant to $O(n^{-3/2})$. The expansion also shows that $dy = \hat{\jmath}^{1/2}d\theta$ at $y = y^0$. Thus (2.1) can be rewritten as the p^* formula.

Expansion (2.1) and its generalization to the multivariate case lead to what can be called an inverse p^* result. From (2.1) we see that the first row is uniquely determined by the remaining rows and presents a function of y that integrates to 1. Then from the preceding paragraph we have that the p^* expression (1.3) also uniquely determines the same first row. It follows that if a p^* formula is calculated from some asymptotic expression in y and θ that need not be a log-density then it produces a function that integrates to one. We need this inverse p^* result in Section 3. For details on the multivariate case see Cakmak, Fraser & Reid (1994).

The p^* formula (1.3) gives the probability density, including θ dependence, at a particular data point and must be recalculated for each data point of interest. The tangent exponential model gives the probability density in a neighborhood of the data point. It is convenient to reexpress (2.1) or its vector analog in a coordinate invariant manner. Let

$$\varphi = \varphi(\theta) = \ell_{;y}(\theta; y^0) = \frac{\partial}{\partial y'}\ell(\theta; y)|_{y^0},$$
$$s = \ell_{\varphi}(\varphi^0; y) = \frac{\partial}{\partial \varphi'}\ell(\theta(\varphi); y)|_{\hat{\theta}^0} \tag{2.2}$$

be a new parameter and corresponding score variable as calculated at the observed data point. By constructing the exponential model expressed by

$$\frac{c}{(2\pi)^{d/2}} exp\{\ell^0(\theta) - \ell^0(\hat{\theta}^0) + (\varphi - \hat{\varphi}^0)s\}|\tilde{\jmath}_{\varphi\varphi}|^{1/2} d\hat{\varphi}, \tag{2.3}$$

where $\tilde{\jmath}_{\varphi\varphi}$ is the observed information matrix for φ based on the exponent in (2.3), we obtain a model yielding the same first two columns as (2.1). Since an exponential model is uniquely determined by its likelihood function and sample space derivative of the likelihood function at the data point, i.e. by the first two columns of (2.1), (Fraser, 1990), it follows that the exponential model (2.3) reproduces the entries in (2.1) or its vector analog to order $O(n^{-1})$. When $d = 1$ it is shown in Fraser & Reid (1993) that the tangent exponential model gives $F(\hat{\theta}^0; \theta)$ to accuracy $O(n^{-3/2})$; i.e. that the integral up to the observed maximum likelihood point of the density in (2.1) is independent of c in (2.1). The resulting formula can be shown to be functionally equivalent to one in Barndorff-Nielsen (1988b).

Thus for an asymptotic context with variable and parameter of the same dimension (as obtained say from a third order ancillary as in Section 3) we have an exponential model that to accuracy $O(n^{-1})$ coincides with the given model, to accuracy $O(n^{-3/2})$ reproduces the original model in a first derivative neighborhood of the observed data point y^0, save a constant of order $O(n^{-1})$, and to accuracy $O(n^{-3/2})$ in the $d = 1$ case reproduces the distribution function value at the data point; we speak of agreeing in a first derivative neighborhood of a point to mean that the two functions have the same value and the same first derivative value at the point. From this we will see in Section 6 that the tangent exponential model can replace the given model for calculating third order significance levels.

3 Third order ancillaries and a unique distribution

Third order ancillaries are not unique but the distribution of such an ancillary projected to the observed maximum likelihood surface is unique to third order. This is developed now and expression (3.3) provides a simple method for checking ancillaries; the theory provides the background for later sections.

Consider a continuous model $f(y;\theta) = f(y_1, \ldots, y_n; \theta)$ with parameter θ of dimension d, satisfying the asymptotic assumptions in Section 2. We obtain an expression for a third order ancillary variable. Computing the total derivative for the score equation $\ell_\theta(\hat{\theta}; y) = 0$ gives the following expression for the differential for the maximum likelihood estimate:

$$d\hat{\theta} = \hat{\jmath}^{-1}\ell_{\theta;y}(\hat{\theta}; y)dy, \tag{3.1}$$

where $\hat{\jmath} = -\ell_{\theta\theta'}(\hat{\theta}; y)$ is the observed information matrix. We then have that volume perpendicular to the surface $\hat{\theta} = \mathit{constant}$ is given by

$$dy_p = |\ell_{\theta;y}(\hat{\theta}; y)|^{-1}|\hat{\jmath}|d\hat{\theta}; \tag{3.2}$$

where $|\ell_{\theta;y}| = |\ell_{\theta;y}(\hat{\theta}; y)\ell'_{\theta;y}(\hat{\theta}; y)|^{1/2}$ is the nominal volume of the p row vectors in the matrix $\ell_{\theta;y}(\hat{\theta}; y)$; also let dy_c be the complementing volume on the surface $\hat{\theta} = \mathit{constant}$.

The statistical model can then be expressd in terms of the new variables,

$$\begin{aligned} f(y;\theta)dy &= exp\{\ell(\theta; y)\}|\ell_{\theta;y}(\hat{\theta}; y)|^{-1}|\hat{\jmath}|dy_c d\hat{\theta} \\ &= g(y;\hat{\theta})dy_c \cdot \frac{c}{(2\pi)^{d/2}} exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\}|\hat{\jmath}|^{1/2}d\hat{\theta}, \end{aligned} \tag{3.3}$$

$$g(y;\hat{\theta})dy_c = \frac{(2\pi)^{d/2}}{c} exp\{\ell(\hat{\theta}; y)\}|\ell_{\theta;y}(\hat{\theta}; y)|^{-1}|\hat{\jmath}|^{1/2}dy_c, \tag{3.4}$$

and tentatively factored as shown. We will see that $g(y;\hat{\theta})$ gives a density on the observed maximum likelihood surface which records the marginal density for any third order ancillary; (3.3) then provides a simple method for checking ancillarity. It is of related interest that (3.4) can also be viewed as a nominal conditional distribution on the observed maximum likelihood surface given the constant information metric $|\hat{\jmath}|^{1/2}d\hat{\theta}$ at $\theta = \hat{\theta}$. These results provide the basis for the ancillaries derived in Section 5.

We now develop properties of this factorization. First, we will be interested in how the likelihood function changes on the surface $\hat{\theta} = \hat{\theta}^0$. From the Taylor expansion methods used for the model (2.1) and its vector version, we note as discussed in Section 2 that the components of the standardized fourth derivative tensor array $\alpha_4(\hat{\theta}^0)$ are constant to order $O(n^{-1/2})$. Thus the change in likelihood on the surface can be described by the second derivative matrix $j(\hat{\theta}^0)$ and the standardized third derivative tensor array $\alpha_3(\hat{\theta}^0)$. We assume that these are expressed in terms of some choice of s scalar variables $a_1, \ldots, a_s$ and then let $a_{s+1}, \ldots, a_{n-d}$ be $n-d-s$ complementing variables, giving a full variable $a(y)$ complementary to $\hat{\theta}$:

$$a(y) = (a_1, \ldots, a_s; a_{s+1}, \ldots, a_{n-d}) = (a_{(1)}; a_{(2)}). \tag{3.5}$$

Now consider $g(y; \hat{\theta})dy_c$ as a nominal relative density for y on the surface with given $\hat{\theta}$ and let $A(\hat{\theta})$ be the corresponding norming constant. Let G be some probability integral transformation, say the coordinate-by-coordinate $G = \{G_1(a_1; \hat{\theta}), \ldots, G_{n-d}(a_{n-d}|a_1, \ldots, a_{n-d-1}; \hat{\theta})\}'$, and $u = \{A(\hat{\theta})G_1, G_2, \ldots, G_{n-d}\}$ be a transformation to a new variable u from the a's, thus giving $g(y; \hat{\theta})dy_c = du$ on $(0, A(\hat{\theta})) \times (0,1)^{n-d-1}$. The model can then be written as

$$f(y; \theta)dy = du \cdot \frac{c}{(2\pi)^{d/2}} exp(\ell - \hat{\ell})|\hat{\jmath}|^{1/2} d\hat{\theta} \tag{3.6}$$

on $\cup[\{0, A(\hat{\theta})\} \times (0,1)^{n-d-1} \times \{\hat{\theta}\}]$ where the union is over values for $\hat{\theta}$.

At this point we wish to prove that $u(y)$ has ancillary properties and that $\ell(\theta; y)$ gives a conditional likelihood, or equivalently that $c(2\pi)^{-1/2} exp(\ell - \hat{\ell})|\hat{\jmath}|^{1/2}$ is a conditional density for $\hat{\theta}$ given u. None of these are available immediately from (3.6). However, from the assumptions that $\ell(\theta; y)$ and its derivatives are $O(n)$ and the particular choice (3.5) for the complementary variable, we do know that on a contour $u(y) = constant$, $\hat{\theta}$ retains the property of being the maximizing θ value. Then from the inverse p^* result it follows that the second factor in (3.3) is a density that integrates to 1 to order $O(n^{-3/2})$ and that the normalizing constant c is constant to the same order.

For any particular θ value we can then integrate $\hat{\theta}$ out of (3.6) obtaining the marginal probability element du for $u = u(y)$. It follows that $A(\hat{\theta}) = 1$, and that u is uniform on $(0,1)^{n-d}$. In succession we then have that $u(y)$ is ancillary of order $O(n^{-3/2})$, that the second factor in (3.6) is the conditional density of $\hat{\theta}$ for given u, and that $\ell(\theta; y)$ is a conditional likelihood to order $O(n^{-3/2})$. Various third order ancillaries are obtained with various probability integral transformations.

4 Needed ancillary information

In this section we investigate the information that is needed for third order inference using the conditioning associated with a third order ancillary. We will see that only the likelihood at the data point and its gradient at the data taken in the ancillary directions are needed. Also we will see that the ancillary directions need to be determined only for a second order ancillary.

For the case that the variable and parameter have the same dimension, we know that the tangent exponential model reproduces the full model to order $O(n^{-1})$ and the local model save a constant to order $O(n^{-3/2})$ and that it provides third order significance for appropriate scalar parameters. To calculate this tangent model we need only the observed likelihood and the likelihood gradient at the data point. If the model is conditional within some third order ancillary. then the conditional likelihood gradient becomes the full likelihood gradient taken tangent to the ancillary surface. To describe this let $V = (v_1, \ldots, v_d)$ be an array of d vectors tangent to a third order ancillary at the data point. Then

$$\ell^0(\theta; y^0), \qquad \varphi = \ell^0_{;V}(\theta; y^0) \tag{4.1}$$

fully determine the tangent exponential model within the conditioning of the ancillary.

We now show that for third order accuracy it suffices to have vectors V only for a second order ancillary. For this we use coordinates $(\hat{\theta}, a_{(1)})$ standardized as for (2.1), and ignore $a_{(2)}$ as $(\hat{\theta}, a_{(1)})$ forms a sufficient statistic to third order. Thus V here becomes an array of vectors each with just s coordinates. For the calculations based on (4.1), the length of a vector is unessential, as its effect appears as a scale factor that cancels out from the tangent exponential model (2.3); in fact all that is required is a set of generators for $\mathcal{L}(V)$ (Fraser, 1990). It is convenient to take an initial V to be an orthonormal set.

Now let an initial array V be tangent to an ancillary of order say just $O(n^{-1})$. Such an ancillary can be upgraded to an ancillary of order $O(n^{-3/2})$ by an $O(n^{-1})$ adjustment (Skovgaard, 1986). Let $W = (w_1, \ldots, w_d)$ be a tangent array at the data point for the upgraded ancillary. For a particular component vector we can then write $v = w + \delta$ with δ parallel to the maximum likelihood surface, and from Skovgaard (1986) have that the length of δ is $O(n^{-1})$. For notation let $w = w_0\tilde{w}$, $\delta = \delta_0\tilde{\delta}$ where for example δ_0 is the length of δ and $\tilde{\delta}$ is the corresponding unit vector. We then have

$$\ell_{;v} = \ell_{;\tilde{w}} w_0 + \ell_{;\tilde{\delta}}\delta_0$$

where $\delta_0 = O(n^{-1})$ and $w_0 = 1$ to order $O(n^{-1})$.

Consider the discrepancy in using V rather than W, as given by the second term on the right. We have that $\hat{\theta}$ is constant parallel to the $\hat{\theta} =$ *constant* surface and thus to order $O(n^{-1})$ the likelihood function varies on the $\hat{\theta} =$ *constant* surface only with respect to $\hat{\jmath}$ and thus does so only to order $O(n^{-1/2})$. It follows that

$$\ell_{;\tilde{\delta}}\delta_0 = O(n^{-1/2})O(n^{-1}) = O(n^{-3/2})$$

We thus have that the needed likelihood gradient on an $O(n^{-3/2})$ ancillary can be obtained by using the gradient in an $O(n^{-1})$ ancillary direction.

5 Determining the ancillary directions

5.1 For a real parameter

Fraser (1964) derived a local ancillary based on an approximating location model for a distribution with a scalar parameter. In this section we describe the approximating location model and first order ancillary statistic. We then show that there is a second order ancillary with the same tangent direction at the data point. Combined with the result from Section 4 this gives the ingredients for third order inference for the parameter. We thus have a second order ancillary and third order inference from a particularly simple and easily calculated first derivative ancillary. For applications the second order ancillary itself does not need to be calculated.

Consider a continuous model $f(y;\theta) = \Pi f_i(y_i;\theta)$ having independent scalar coordinates and scalar parameter θ, together with a data value y^0 and corresponding $\hat{\theta}(y^0) = \hat{\theta}^0$ which we write as θ_0 for simplicity. We assume that the parameter θ is effective: for each i, the distribution function $F_i(y_i;\theta)$ has a nonzero derivative with respect to θ and thus each y_i is stochastically monotone. For convenience we assume that directions for the variables have been reversed as needed, so that the variables are stochastically increasing in θ.

For the ith coordinate with density f and distribution function F, define a new variable by

$$x_i = \int^{y_i} \{-F_{y;}(y;\theta_0)\}/F_{;\theta}(y;\theta_0)\}dy. \tag{5.1}$$

As shown in Fraser (1964), the density function for x satisfies $\partial f(x;\theta)/\partial\theta = -\partial f(x;\theta)/\partial x$, at $\theta = \theta_0$; i.e. the model for x is a location model to first derivative at $\theta = \theta_0$.

We will use y_i again for the modified variable and denote the corresponding location model by $f(y-\Delta;\theta_0)$ where $\Delta = \theta - \theta_0$. Expanding $\log f(y;\theta) - \log f(y-\Delta;\theta_0)$ in Δ about $\Delta = 0$ gives

$$\begin{aligned}
\ell(\theta;y) - \ell(\theta_0;y-\Delta) &= \{\ell_{\theta\theta}(\theta_0;y) - \ell_{yy}(\theta_0;y)\}\frac{\Delta^2}{2} \\
&\quad + \{\ell_{\theta\theta\theta}(\theta_0;y) + \ell_{yyy}(\theta_0;y)\}\frac{\Delta^3}{6} + O(\Delta^4), \\
f(y;\theta) &= f(y-\Delta;\theta_0)exp\{r(y)\Delta^2/2 + s(y)\Delta^3/6 + O(\Delta^4)\} \\
&= f(y-\Delta;\theta_0)\{1 + r(y)\Delta^2/2 + s(y)\Delta^3/6 + O(\Delta^4)\}
\end{aligned} \tag{5.2}$$

from which we see that $E\{r(y);\theta_0\} = 0$.

For the full model on R^n, we write $\theta = \theta_0 + \delta n^{-1/2}$. Then $f(y;\theta)$ can be written

$$
\begin{aligned}
&\Pi[f_i(y_i - \delta n^{1/2};\theta_0) exp\{\frac{1}{2n}\delta^2 r_i(y_i) + \frac{1}{n^{3/2}}\delta^3 s_i(y_i) + O(n^{-2})\}] \\
&= \Pi\{f_i(y_i - \delta n^{-1/2};\theta_0)\} \\
&\quad \{1 + n^{-1}\sum r_i(y_i)\delta^2/2 + n^{-3/2}\sum s_i(y_i)\delta^3/6 + O(n^{-2})\}.
\end{aligned}
\tag{5.3}
$$

The location model that is the first factor in (5.3) has an exact ancillary, say d, so this factor can be expressed as the product of the marginal density for d, which is free of δ and the conditional density given d, which is a one-dimensional distribution depending on δ. For an arbitrary point y, let $y_0 = y_0(y)$ be the point on the observed maximum likelihood surface with the same value for the location model ancillary $d(y)$; that is, $\hat\theta(y_0) = \hat\theta^0 = \theta_0$ and $y - y_0 = (z/\sqrt{n})\cdot 1$ where 1 designates the 1-vector. The location model factorization can be written as

$$
\Pi f_i\{y_{0i} + (z-\delta)n^{-1/2};\theta_0\}dy = exp\{\ell(\delta - z; y_0)\}dzh(y_0)dy_0, \tag{5.4}
$$

where the conditional density along the orbit has the location model form and the marginal density for the orbits on the observed maximum likelihood surface has density $h(y_0)$ with respect to volume dy_0 orthogonal to $\mathcal{L}(1)$.

We now wish to find the corresponding marginal and conditional distributions in the full model $f(y;\theta)$. The non-location factor for the full density in (5.3) can be examined along the observed maximum likelihood surface and along the orbits; we write

$$
\frac{1}{n}\sum r_i(y_i) = \frac{1}{\sqrt{n}}u(y) = \frac{1}{\sqrt{n}}(w + fz)
$$

where $u(y)$ has mean 0 and variance $O(1)$ at $\delta = 0$, $w(y) = u(y_0(y))$, and $u(y) - u(y_0)$ is expanded in terms of z along the orbits $y_0 + \mathcal{L}(1)$. For convenience we assume w has been scaled so that it has variance 1; the log density can then be written

$$
\ell(\theta; y) = a(y_0) + \ell(\delta - z; y_0) + ew\delta^2/2n^{1/2} + fz\delta^2/2n^{1/2} + g\delta^3/6n^{1/2} \tag{5.5}
$$

and the departures from location model form appear in the terms involving e, f and g. These departures depend on y only through w and z.

In the Appendix we show that the orbits of the location model ancillary defined by z can be adusted by introducing a curvature term, and that the marginal density of the new orbits is free of δ to $O(n^{-1})$; i.e. the new orbits define a second order ancillary statistic. Since the new orbits are obtaind by adding a curvature term at the data point the tangent direction of the second order ancillary at the data point is the same as that of the first order location model ancillary. Thus in applications it is sufficient to work with the ancillary for the local location model defined in (5.1). For later use it is convenient to record the $O(n^{-1})$ ancillary direction $v = v(y) = \{v_1(y), \ldots, v_n(y)\}'$ in terms of the original coordinates:

$$v_i(y) = \frac{\partial F_i(y;\theta)/\partial\theta}{\partial F_i(y;\theta)/\partial y}|_{\hat{\theta}}. \qquad (5.6)$$

5.2 For a vector parameter

For a vector parameter we restrict our attention here to the case of independent coordinates, each with a single parameter θ_i which in turn is linearly dependent on a common parameter $\beta = (\beta_1, \ldots, \beta_d)$. We obtain an approximating location model for β with a first order ancillary and then find a second order ancillary with the same tangent directions. We assume that the location adjustment described by (5.1) has already been made.

For the ith component y_i we use (5.2) with $\Delta_i = \theta_i - \theta_{i0} = X_i(\beta - \beta_0)$ where $\beta_0 = \hat{\beta}_0$ is the overall maximum likelihood estimate, $\theta_{i0} = X_i\beta_0$; and $X_i = \partial\theta_i/\partial\beta'$ at β_0. It would be sufficient to assume local linearity of $\theta_i = h_i(\beta)$, with a bound on the second derivatives. For the full model we use $\beta = \beta_0 + \delta n^{-1/2}$ and obtain

$$\Pi\{f_i(y_i - X_i\delta n^{-1/2}; \theta_{i0})\}\{1 + n^{-1}\delta' R\delta/2 + n^{-3/2}S\delta^3/6\} \qquad (5.7)$$

where $R = \sum r_i(y_i)X_i'X_i$ and $S\delta^3$ is in fact a third order array.

The location model given by the first expression in (5.2) has an exact ancillary. For this we write $y_0 = y_0(y)$ where y_0 is on the observed maximum likelihood surface and has the same value for the location model ancillary, and $y - y_0 = Xzn^{-1/2}$, where X is the $n \times d$ matrix with rows X_i. The location model factorization is then

$$\Pi f_i\{y_{0i} + X_i(z - \delta)n^{-1/2}; \theta_0\}dy = exp\{\ell(\delta - z; y_0)\}dzh(y_0)dy_0 \quad (5.8)$$

where the conditional density has d dimension location form and the marginal density $h(y_0)$ on the observed maximum likelihood surface uses volume orthogonal to $\mathcal{L}(X)$.

Now consider the full model $f(y;\beta)$. The non-location factor can be examined in terms of the observed maximum likelihood surface and the location orbits, and we have

$$\frac{R}{n} = \frac{W}{\sqrt{n}} + \frac{\sum F_i z_i}{\sqrt{n}} \qquad (5.9)$$

where R has mean zero at $\delta = 0$, $W = n^{-1/2}R(y_0)$, and $n^{-1/2}\{R(y) - R(y_0)\}$ has been expanded in the coordinates of z along $\mathcal{L}(X)$. The full log density is then

$$\ell(\beta; y_0) = a(y_0) + \ell(\delta - z; y_0) + \frac{\delta' W \delta}{2n^{1/2}} + \frac{\sum \delta F_i \delta z_i}{2n^{1/2}} + \frac{g\delta^3}{6n^{1/2}} \quad (5.10)$$

and the departure from location model form appears in the last three terms, and depends on y only through the coordinates of W and of z.

As in Section 5.1 the orbits can be adjusted by a curvature term that does not alter the direction at the observed maximal likelihood surface. Thus we can work with the location model ancillary implicit in (5.8). For later use it is convenient to record the $O(n^{-1})$ ancillary directions V in terms of the original coordinates: $V = DX$ where D is diagonal with

$$d_{ii}(y) = -\frac{\partial F_i(y;\theta)/\partial\theta}{\partial F_i(y;\theta)/\partial y}\Big|_{\{y;\theta(\hat\beta)\}}$$

6 Testing an interest parameter

In this Section we assume that the parameter θ is partitioned as (λ, ψ) where ψ is the parameter of interest, of dimension d_ψ and λ is a nuisance parameter of dimension $d_\lambda = d - d_\psi$. For the full parameter θ the ancillary directions V are determined in Sections 4 and 5. This gives a tangent exponential model for a d-dimensional variable as at (2.2) and (2.3):

$$\varphi = \ell_{;V}(\theta; y)|_{y^0}, \quad s = \ell_{\varphi;}(\theta; y)|_{\hat\theta^0}, \quad (6.1)$$

$$\frac{c}{(2\pi)^{d/2}} exp\{\ell^0(\varphi) - \ell^0(\hat\varphi^0) + (\varphi - \hat\varphi^0)s\} |\tilde{\jmath}_{\varphi\varphi}|^{-1/2} ds, \quad (6.2)$$

where we write $\ell^0(\varphi)$ for $\ell^0(\theta)$ with θ expressed in terms of φ by (6.1).

We now restrict attention to a particular value for ψ, writing $\theta_{\psi_0} = (\lambda, \psi_0)$. As in Sections 3 and 4, the submodel of (6.2) obtained by fixing $\psi = \psi_0$ can be factored into a conditional distribution given some ancillary directions and a marginal distribution for the ancillary, which is free of λ. The marginal distribution, which is unique when projected to the surface $\hat\theta_{\psi_0} = \hat\theta^0_{\psi_0}$, provides pivotal assessment of the parameter value ψ_0. The marginal distribution can be obtained from the ratio of (6.2) to the conditional distribution given the ancillary directions for λ taken at the observed maximum likelihood point.

We first need to establish a link between the partition of θ and a partition of the canonical parameter φ. We write $\varphi = (\varphi_1, \varphi_2)$ and $s' = (s_1', s_2')$ and assume the vectors V have been chosen so that the surface $\psi = \psi_0$ on the full parameter space is tangent to the surface $\varphi_2 = \varphi_{20}$ at $\hat\theta = \hat\theta_{\varphi_0}$ and increments in the same directions. It follows that the maximum likelihood surfaces $\hat{\theta}_{\psi_0} = \hat{\theta}^0_{\psi_0}$ and $\hat{\theta}_{\varphi_{20}} = \hat{\theta}^0_{\varphi_{20}}$ are the same.

We now follow the construction in Section 3 within the conditional model (6.2) with $\varphi = (\varphi_1, \varphi_{20})$. The conditional density for the nuisance parameter score at the observed maximum likelihood point is given by the analog of the second factor in (3.3):

$$\frac{c'}{(2\pi)^{(d-d_\varphi)/2}} exp\{\ell^0(\varphi_1, \varphi_{20}) - \ell^0(\hat{\varphi}^0_{1\varphi_{20}}, \varphi_{20})\} |\tilde{\jmath}_{(\varphi_1\varphi_1)}(\hat{\theta}^0_{\varphi_{20}})|^{-1/2} ds_1,$$

where the conditional score is just s_1 and the information is calculated for φ_1 with ψ fixed at ψ_0,

$$|\tilde{\jmath}_{(\varphi_1\varphi_1)}(\hat{\theta}^0_{\varphi_{20}})| = |\jmath_{\lambda\lambda}(\hat{\theta}^0_\psi)||\varphi_\lambda(\hat{\theta}^0_\psi)|^{-2}.$$

The marginal ancillary density for s_2 as projected to the maximum likelihood surface $\hat{\theta}_{\psi_0} = \hat{\theta}^0_{\psi_0}$, $\hat{\varphi}_2 = \hat{\varphi}^0_{20}$ or $s_1 = 0$ can then be obtained by dividing the joint density (6.2) by the preceding conditional density:

$$\begin{aligned} &\frac{c''}{(2\pi)^{d_\psi/2}} exp\{\ell^0(\hat{\varphi}^0_{1\varphi_{20}}, \varphi_{20}) - \ell^0(\hat{\varphi}^0) \\ &\quad + (\varphi_2 - \hat{\varphi}_2)s_2\} |\tilde{\jmath}_{\varphi\varphi}|^{-1/2} |\tilde{\jmath}_{(\varphi_1\varphi_1)}(\hat{\theta}^0_{\varphi_{20}})|^{1/2} ds_2. \end{aligned} \tag{6.3}$$

This presents itself as an exponential model with modulating factor as in (1.5). Note that the parameterization is special to the value ψ_0 or to φ_{20}.

In the case $d_\psi = 1$ the left tail probability for testing the value ψ (for convenience we now write ψ for ψ_0) is given by (1.1) using the signed profile likelihood ratio corresponding to the data point $s_2 = 0$ given by

$$\begin{aligned} R &= sgn(\hat{\varphi}^0_2 - \varphi_2)[2\{\ell^0(\hat{\varphi}^0) - \ell^0(\hat{\varphi}^0_{1\varphi_2}, \varphi_2)\}]^{1/2} \\ &= sgn(\hat{\psi}^0 - \psi)[2\{\ell^0(\hat{\theta}^0) - \ell^0(\hat{\theta}^0_\psi)\}]^{1/2} \end{aligned} \tag{6.4}$$

together with the standardized nuisance-adjusted departure $Q = MJ_2^{1/2}/J_1^{1/2}$. Here M is the maximum likelihood departure for the coordinate φ_2, and J_2 and J_1 are full and nuisance parameter information determinants. In terms of the original φ coordinates

$$M = \hat{\varphi}^0_2 - \varphi_2 = sgn(\hat{\psi} - \psi)|\{\hat{\varphi}^0 - \varphi(\hat{\theta}^0_\psi)\}s_\psi|/|s_\psi|,$$

where $s_\psi = \varphi^\psi(\hat{\theta}^0_\psi)$ gives in φ coordinates the vector perpendicular to the ψ curve at $\hat{\theta}^0_\psi$

$$d\varphi = (d\lambda, d\psi)\varphi_{\theta'}(\hat{\theta}^0_\psi), \quad d\psi = d\varphi\ \varphi^\psi(\hat{\theta}^0_\psi),$$

and $\{\varphi_{\theta'}(\hat\theta^0_\psi)\}^{-1} = \{\varphi^\lambda(\hat\theta^0_\psi), \varphi^\psi(\hat\theta^0_\psi)\}$ is the inverse of the θ to φ Jacobian. Similarly

$$J_2 = |\tilde{\jmath}_{\varphi\varphi}| = |\jmath(\hat\theta^0)||\varphi_{\theta'}(\hat\theta^0)|^{-2},$$
$$J_1 = |\tilde{\jmath}_{(\varphi_1\varphi_1)}(\hat\theta^0_{\varphi_2})| = |\jmath_{\lambda\lambda}(\hat\theta^0_\psi)||\varphi_{\lambda'}(\hat\theta^0_\psi)|^{-2}, \tag{6.5}$$

recalibrated on the φ scale. We see that the maximum likelihood departure is standardized by a quotient involving full and nuisance information in appropriate coordinates and can be viewed as applying an appropriately calculated marginal standard deviation for the parameter replacing ψ. These calculations use values within a first derivative neighborhood of the observed data, where the approximate model provides $O(n^{-3/2})$ accuracy for a tail probability. Alternative formulas using adjusted profile likelihoods are discussed in Cheah, Fraser & Reid (1995).

In the effective context of the same dimension for variable and parameter, Barndorff-Nielsen (1991) obtained a third order formula for a scalar component parameter. In the submodel obtained by fixing ψ, Barndorff-Nielsen (1991) and (1986) shows that the likelihood ratio statistic for testing the submodel, designated R in (6.4), can be adjusted to be ancillary to order $O(n^{-3/2})$ with respect to the nuisance parameter of the submodel. Thus the adjusted version Barndorff-Nielsen's r^*, provides a pivotal assessment for ψ, as does the ancillary derived above. Our Q is similar to the adjustment from R to r^*: the exact form of the adjustment is given in (3.3) in Barndorff-Nielsen (1991). If this assessment of ψ is combined with the ancillary from Section 5, we obtain an alternative to the preceding tail probability.

If the parameter of interest is a vector, one approach is to apply the methods of Section 6 to successive components of the parameter vector. The general approach is described in Fraser & McKay (1975, 1976), Fraser (1987), and Fraser & McDunnough (1988). Barndorff-Nielsen (1986) describes successive testing of parameter components using r^*.

For example writing the parameter of interest as $\theta = (\theta_1, \dots, \theta_d)$ the ith hypothesis $H_i\colon \theta_i = \theta_{i0}$ would have parameter $\theta_{(i)} = (\theta_1, \dots, \theta_i, \theta_{(i+1)0}, \dots, \theta_{d0})$. The value for the ith hypothesis is given by $p_i = \Phi(R_i, M_i, J_i^{1/2}/J_{i-1}^{1/2})$ where

$$R_i = sgn_i(\hat\theta^0_{(i)} - \hat\theta^0_{(i-1)})[2\{\ell^0(\hat\theta^0_{(i)}) - \ell^0(\hat\theta^0_{(i-1)})\}]^{1/2}$$
$$M_i = \varphi_1(\hat\theta^0_{(i)}) - \varphi_1(\hat\theta^0_{(i-1)})$$
$$J_i = |\jmath_{\theta_{(i)}\theta_{(i)}}(\hat\theta^0_{(i)})||\varphi_{\theta_{(i)}}(\hat\theta^0_{(i)})|^{-2}$$

where sgn_i gives the sign of the ith coordinate and the φ coordinates are successively defined. Details on constructing the successive ancillary directions are outlined in an unpublished technical report available from the authors.

7 Example: Exponential regression

We write the model for the ith observation as $f_i(y_i) = exp\{-\theta_i y_i + \log(\theta_i)\}$ with $E(y_i) = \theta_i^{-1} = exp\{\alpha + \beta(x_i - \bar{x})\} = exp\{X_i(\alpha, \beta)'\}$. This is an example of a generalized linear model with link function

$$\theta_i = g\{X_i(\alpha, \beta)'\} = exp\{-(1, x_i - \bar{x})(\alpha, \beta)'\}, \tag{7.1}$$

which is nonlinear in the canonical parameter θ_i.

First we follow the methods in Section 5 and determine how the parameters α, β move the data point in a translation manner. Writing $dy_i = d(y_i; \theta_i)d\theta_i$ we obtain from Section 5.1 that $d(y_i; \theta_i) = -y_i/\theta_i$. Then introducing the dependence of θ_i on α, β we obtain

$$dy_i = d(y_i; \theta_i) g_i'\{X_i(\alpha, \beta)'\}\{-1 \cdot d\alpha - (x_i - \bar{x})d\beta\}$$

which when evaluated at $(\hat{\alpha}^0, \hat{\beta}^0)$ gives the ancillary vectors

$$v_1 = (y_1^0, \ldots, y_n^0), \quad v_2 = \{y_1^0(x_1 - \bar{x}), \ldots, y_n^0(x_n - \bar{x})\}.$$

Next we record the tangent exponential model (Section 2) that approximates the conditional model given the ancillary with tangent vectors v_1, v_2 at the observed data:

$$g(s_1, s_2) = exp\{\ell^0(\alpha, \beta) + \varphi_1 s_1 + \varphi_2 s_2\} |\tilde{\jmath}_{(\alpha,\beta),(\alpha,\beta)}|^{1/2} ds_1 ds_2 \tag{7.2}$$

where

$$\ell^0(\alpha, \beta) = n\alpha - \sum y_i^0 exp\{-\alpha - \beta(x_i - \bar{x})\} \tag{7.3}$$

$$\begin{aligned} \varphi_1 &= \frac{d}{dv_1}\ell(\alpha, \beta; y)|_{y^0} = \sum exp\{-\alpha - \beta(x_i - \bar{x})\} y_i^0 \\ \varphi_2 &= \frac{d}{dv_2}\ell(\alpha, \beta; y)|_{y^0} = \sum exp\{-\alpha - \beta(x_i - \bar{x})\} y_i^0 (x_i - \bar{x}). \end{aligned} \tag{7.4}$$

Then to obtain significance for the parameter β we follow Section 6. The signed likelihood ratio for β is

$$R = sgn(\hat{\beta}^0 - \beta)[2\{\ell^0(\hat{\alpha}, \hat{\beta}) - \ell^0(\hat{\alpha}_\beta, \beta)\}]^{1/2} \tag{7.5}$$

and by itself provides significance $\Phi(\beta)$ to order $O(n^{-1/2})$. To calculate the particular canonical parameter component of the exponential model that coincides with β at $(\hat{\alpha}_\beta, \beta)$ we can calculate the gradient $(\ell_\alpha^0, \ell_\beta^0)$ of the observed likelihood at $(\hat{\alpha}_\beta, \beta)$ expressed in terms of the canonical parameters (7.4) and evaluated at $(\hat{\alpha}_\beta, \beta)$; this is an alternative to the calculation preceding (6.5):

$$s_\beta = \begin{pmatrix} \partial\varphi_1/\partial\alpha & \partial\varphi_1/\partial\alpha \\ \partial\varphi_1/\partial\beta & \partial\varphi_2/\partial\beta \end{pmatrix}^{-1} \begin{pmatrix} \ell^0_\alpha \\ \ell^0_\beta \end{pmatrix} |_{(\hat{\alpha}_\beta, \beta)}. \tag{7.6}$$

The standardized maximum likelihood departure $Q = MJ_2^{1/2}/J_1^{1/2}$ is then obtained from

$$M = sgn(\hat{\beta}^0 - \beta)\{(\varphi_1(\hat{\alpha}, \hat{\beta}) - \varphi_1(\hat{\alpha}_\beta, \beta), \varphi_2(\hat{\alpha}, \hat{\beta}) \\ - \varphi_2(\hat{\alpha}_\beta, \beta)\} \frac{s_\beta}{|s_\beta|} \tag{7.7}$$

together with (6.5). The observed significance is given by (1.1) using R and Q as just defined.

This model is a location model on the log scale, so an exact calculation of significance for β is available. The location model pivotal $t_i = \log y_i - \alpha - \beta(x_i - \overline{x})$ has the extreme value distribution. The joint distribution of $(\hat{\alpha}, \hat{\beta})$ conditional on the ancillary of the location model is obtained from the likelihood function

$$f(\hat{\alpha}, \hat{\beta}; \alpha, \beta) = c\ exp\{\ell^0(\alpha - \hat{\alpha} + \hat{\alpha}^0, \beta - \hat{\beta} - \hat{\beta}^0)\}$$

and the significance for β with α unknown is

$$P(\hat{\beta} \le \hat{\beta}^0; \beta) = \int_\beta^\infty \int_{-\infty}^\infty c\ exp\{\ell^0(\alpha, t)\} d\alpha dt.$$

This model was used to fit data from Example U in Cox & Snell (1981) in Fraser, Monette, Ng, and Wong (1995). The data consists of 17 observations on lifetime (in weeks) of leukemia patients, and the concomitant variable is the log of the initial white blood cell count. The 95% confidence intervals for β are

first order	-1.9153	-0.2835
third order	-1.9144	-0.2729
exact	-1.9145	-0.2726.

This model is an example of a generalized linear model with non-canonical link. Such models are considered in general in Fraser, Monette, Ng, and Wong (1995). When the link function of a generalized linear model is canonical, then the usual third order methods for exponential families can be applied: the approach described in this paper then reduces to Skovgaard's (1987) double saddlepoint approximation. Numerical results described in Pierce & Peters (1992) suggest that the sequential saddlepoint approximation of Fraser, Reid & Wong (1991) can be more accurate. Both of these approximations are asymptotically equivalent to Barndorff-Nielsen's r^* approximation (1986) in exponential models.

Appendix

The location model included in (5.5) has marginal density $h(y_0) = e^{a(y_0)}$ on the observed maximum likelihood surface and conditional location density $\ell(\delta - z; y_0)$ along the orbit $y_0 + \mathcal{L}(1)$. The adjustments to this location model are $ew\delta^2/2n^{1/2}$, which depends on a scalar variable w on the maximum likelihood surface, $fz\delta^2/2n^{1/2}$, which depends on z along the location orbit, and $q\delta^3/6n^{1/2}$, which is constant.

The first order limiting distribution of w is normal with mean 0 and variance 1. We can obtain an ancillary $A(y_0)$ following the methods in Section 5.1 or using the conditioning as with $a_{(2)}$ in Section 3 with respect to the $ew\delta^2$ modulation on the maximum likelihood surface, such that w has the same limiting normal distribution given that ancillary. Now the δ dependence can be examined in the joint density for z and w, conditional on this ancillary A. We now modify w so it becomes ancillary with respect to δ.

Define $\tilde{w} = w + az^2/2n^{1/2}$. Substituting this into the location model for w and z gives the joint density for $\tilde{w}$ and z:

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}} exp\{-(\tilde{w} - az^2/2n^{1/2})^2/2\} exp\ \ell(\delta - z, y_0) \\
=&\frac{1}{\sqrt{2\pi}} exp\{-\tilde{w}^2/2\}(1 + \tilde{w}az^2/2n^{1/2}) exp\ \ell(\delta - z; y_0).
\end{aligned}
$$

To first order z is normal with mean δ and variance v, say. Thus the derived marginal density for $\tilde{w}$ is $N(0,1)\ \{1 + \tilde{w}a(\delta^2 + v)/(2n^{1/2})\}$. The bending of the location orbits by going from w to $\tilde{w}$ introduces a factor asymptotically equivalent to $exp(aw\delta^2/2n^{1/2})$ which by choice of a can eliminate the existing modulating factor $exp(ew\delta^2/2n^{1/2})$. Thus by projecting along the modified orbits $\tilde{w} = constant$ we obtain a new marginal distribution on the maximum likelihood surface and this distribution is free of δ to order $O(n^{-1})$. This adjustment to the orbits does not alter the tangent directions at the maximum likelihood surface.

For the vector parameter case we have a factor $exp(\delta' W\delta/2n^{1/2})$ which modulates the density of the location model ancillary. Consider any component, say w_{ij} in the matrix W. As in the preceding case we can find a corresponding ancillary and, conditional on this ancillary, examine the $d+1$-dimensional distribution for z and w_{ij}. A curvature term of the form $a_{ij}z_iz_j/2n^{1/2}$ can bend the location orbit in the direction prescribed by the ancillary; and an appropriate choice of a_{ij} then eliminates the modulation above. Each of the components that modulates the location ancillary density can be eliminated this way without modifying the location ancillary directions at the maximum likelihood surface.

References

[1] Barndorff-Nielsen, O.E. (1980). Conditionality resolutions, *Biometrika* **67** 293–310.

[2] Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–65.

[3] Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized, signed log likelihood ratio. *Biometrika* **73** 307–22.

[4] Barndorff-Nielsen, O.E. (1988a). *Parametric Statistical Models and Likelihood*, Lecture Notes in Statistics **50**, Heidelberg: Springer.

[5] Barndorff-Nielsen, O.E. (1988b). Contribution to discussion of Reid (1988).

[6] Barndorff-Nielsen, O.E. (1990). Approximate internal probabilities. *J. R. Statist. Soc. B* **52** 485–96.

[7] Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio, *Biometrika* **78** 557–563.

[8] Barndorff-Nielsen, O.E. Chamberlin, S. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika* **81** 485–499.

[9] Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddle-point approximations with statistical applications. *J. R. Statist. Soc. B* **41** 279–312.

[10] Bartlett, M.S. (1955). Approximate confidence intervals: III, a bias correction. *Biometrika* **42** 201–204.

[11] Cakmak, S., Fraser, D.A.S. and Reid, N. (1994). Multivariate asymptotic model, exponential and location approximations. *Utilitas Math.*, **46** 21–31.

[12] Cheah, P.K., Fraser, D.A.S., and Reid, N. (1994). Adjustments to likelihoods and densities; calculating significance. *Journal of Statistical Research*, to appear.

[13] Cox, D.R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.

[14] Cox, D.R. (1971). The choice between alternative ancillary statistics. *J. R. Statist. Soc. B* **33** 251–262.

[15] Cox, D.R. and Snell, E.J. (1981). "Applied Statistics", Chapman and Hall, London.

[16] Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.

[17] DiCiccio, T.J., Field, C.A. and Fraser, D.A.S. (1990). Approximation of marginal tail probabilities and inference for scalar parameters. *Biometika* **77** 77–95.

[18] DiCiccio, T.J. and Martin, M.A. (1993) Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc. B* **55** 305–316.

[19] Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected infomation. *Biometrika* **65** 457–487.

[20] Evans, M., Fraser, D.A.S. and Monette, G. (198G). On principles and arguments to likelihood. *Canad. J. Statist.* **14** 181–200.

[21] Fisher. R.A. (19:31). Two new properties of mathematical likelihood. *Proc. R. Soc. A* **144** 285–307.

[22] Fraser, D.A.S. (1964). Local conditional sufficiency. *J. R. Statist. Soc. B* **26** 52–62.

[23] Fraser, D.A.S. (1987). Sequential parameter structure, conditional inference, and likelihood drop. *Statistical Papers* **28** 27–52.

[24] Fraser, D.A.S. (1988). Normed likelihood as saddlepoint approximation. *J. Mult. Anal.* **27** 181–193.

[25] Fraser, DA.S. (1990). Tail probabilities from observed likelihood. *Biometrika* **77** 65–76.

[26] Fraser, D.A.S. and McKay, J. (1975). Parameter factorization and inference based on significance, likelihood, and objective posterior. *Ann. Statist.* **3** 559–572.

[27] Fraser, D.A.S. and McKay, J. (1976). On the equivalence of standard inference procedure. *Foundations of Probability Theory, Statistical Inference* W.L. Harper and C.A. Hooker (eds.), in "Statistical Theories of Science", **Vol II** 47–62.

[28] Fraser, D.A.S. and McDunnough P. (1988). On generalization of the analysis of variance. *Ann. Inst. Statist. Math* **40** 353–366.

[29] Fraser, D.A.S.. Monette, G., Ng. K.W., Wong, A. (1995). Higher order approximations with generalized linear models. *Proceedings of the Symposium on Multivariate Analysis*, Hong Kong, to appear.

[30] Fraser, D.A.S. and Reid, N. (1993). Simple asymptotic connections between densities and cumulant generating function leading to accurate approximations for distribution functions. *Statist. Sinica* **3** 67–82.

[31] Fraser, D.A.S., Reid, N. and Wong, A. (1991). From observed likelihood to tail area: a two pass procedure. *J. R. Statist. Soc. B* **52** 483–492.

[32] Lugannani, R. and Rice, S.O (1980). Saddlepoint approximation for the distribution of the sums of independent random variables. *Adv. Appl. Prob.* **12** 475–490.

[33] McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman & Hall.

[34] Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families, *J. Royal Statist. Soc. B* **54**.

[35] Reid. N. (1988) Saddlepoint methods and statistical inference. *Stat. Sci.* **3** 213–237.

[36] Skovgaard, I.M. (1986). Successive improvements of the order of ancillarity. *Biometrika* **3** 516–519.

[37] Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions, *J. App. Prob.* **24** 875–887.

[38] Skovgaard, I.M. (1990). On the density of minimum contrast estimators. *Ann. Statist.* **18** 779–789.