Multiparameter testing in exponential models: third order approximations from likelihood

P.K. Cheah
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

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

Summary

For composite testing of several canonical components in an exponential model, a conical or directional test provides a one-dimensional measure of departure. We approximate the distribution of this test statistic to third order by numerical inversion of a profile likelihood. The method extends the Lugannani and Rice type formula to kernels other than the normal.

Some keywords: Approximate distributions; Asymptotic expansions; Conditional inference; Conditional likelihood; Exponential families; Lugannani and Rice approximation; Saddlepoint approximation; Tail probability.

1. Introduction

Consider an exponential family model with p dimensional parameter φ,

$$\exp\{\varphi' y - c(\varphi)\} h(w) \tag{1.1}$$

where y is a p dimensional statistic and the density is for a variable w on some original sample space of dimension n. The log-likelihood function is $l(\varphi; y) = \varphi' y - c(\varphi)$. The saddlepoint approximation to the marginal density of y (Daniels, 1954; Barndorff-Nielsen & Cox, 1979) is given by

$$f(y; \varphi) = (2\pi)^{-p/2} \exp\left(-\frac{\delta}{2} - \frac{r^2}{2}\right) |\hat{j}|^{-1/2}\{1 + O(n^{-\frac{3}{2}})\}\ , \tag{1.2}$$

where

$$r^2/2 = r^2(\hat{\varphi}; \varphi)/2 = l(\hat{\varphi}; y) - l(\varphi; y), \tag{1.3}$$

$\hat{j} = -l_{\varphi\varphi'}(\hat{\varphi}; y)$ is the observed information matrix, and δ is a constant of order $O(n^{-1})$ involving standardized third and fourth derivatives of $c(\varphi)$. For the $p = 1$ case an approximation to the distribution function that is accurate to $O(n^{-\frac{3}{2}})$ was derived by Lugannani and Rice (1980). We develop a related result for the multidimensional case.

Now suppose that $\varphi' = (\theta', \kappa')$ and $y' = (y_1', y_2')$ with θ and y_1 of dimension k. The conditional distribution of y_1 given y_2 is an exponential family with canonical parameter θ and is the natural distribution for inference concerning θ as it is free of the nuisance parameter κ. Then taking the quotient of (1.2) for (y_1, y_2) and for y_2 gives

$$f(y_1|y_2; \theta) = (2\pi)^{-k/2} \exp\left\{-\frac{\delta - \delta_2(\theta)}{2} - \frac{r_p^2}{2}\right\} |\hat{j}_p|^{-1/2} \left\{\frac{|j_\kappa(\hat{\theta}, \hat{\kappa})|}{|j_\kappa(\theta, \hat{\kappa}_\theta)|}\right\}^{-1/2} \tag{1.4}$$

to order $O(n^{-\frac{3}{2}})$ where $\hat{j}_p$ is the nominal observed information obtained from the profile log-likelihood $l_p(\theta) = l(\theta, \hat{\kappa}_\theta)$, $\hat{\kappa}_\theta$ is the restricted maximum likelihood estimate of κ for fixed θ, $r_p^2/2 = l_p(\hat{\theta}) - l_p(\theta)$ is the corresponding log-likelihood ratio, and j_κ is the nuisance parameter submatrix of the information matrix from the full log-likelihood $l(\theta, \kappa)$. As this conditional model is exponential family it follows easily that

$$l_p(\theta; y_1, y_2) = \theta'(y_1 - y_1^0) + l_p(\theta; y_1^0, y_2), \tag{1.5}$$

where the observed data point $y^{0'}$ is partitioned as $(y_1^{0'}, y_2^{0'})$.

Approximation (1.4) is the same as (3.15) in Barndorff-Nielsen and Cox (1979), but is expressed here entirely in terms of the log-likelihood function. The likelihood formulation has been very fruitful as a method for generating new approximations, in some of the recent literature. See for example Barndorff-Nielsen and Cox (1989, Ch. 7), Barndorff-Nielsen (1990, 1991), Fraser (1990) and references therein. By examining general asymptotic expansions, Fraser and Reid (1993) show that $\delta_2(\theta)$ in (1.4) is constant to order $O(n^{-\frac{3}{2}})$ and thus that the conditional log-likelihood for the density in (1.4) can be approximated by

$$l_p(\theta) + \frac{1}{2} \log |j_\kappa(\theta, \hat{\kappa}_\theta)|. \tag{1.6}$$

Saddlepoint inversion of this conditional log-likelihood gives density and distribution function approximations to order $O(n^{-\frac{3}{2}})$. The same reference can also be used to show that the inversion of the profile log-likelihood $l_p(\theta)$ defines a nominal profile density to order $O(n^{-\frac{3}{2}})$:

$$f(y_1; \theta | l_p) = \frac{c}{(2\pi)^{k/2}} \exp\left(-\frac{r_p^2}{2}\right) |\hat{j}_p|^{-1/2}. \tag{1.7}$$

The numerical implementation of these inversions as a filter transforming likelihood to density and significance is discussed in Fraser, Reid, and Wong (1991) and Fraser (1991).

2. Multiparameter significance for exponential models

2.1 Preliminaries

We assume that we have a sample of size n from the initial exponential family model. We investigate the hypothesis $\theta = \theta_0$ where θ is of dimension k, using the conditional model (1.4) for y_1 given y_2; the corresponding log-likelihood to order $O(n^{-\frac{3}{2}})$ is given by (1.6). Since the nuisance parameter is eliminated in the conditional density, we now take the effective model to be (1.4) with log-likelihood (1.6), and we will use the notation $f(y;\theta)$ and $l(\theta;y)$ where y now is the original y_1 of dimension k.

The usual likelihood based test statistics for testing $\theta = \theta_0$ are the log-likelihood ratio statistic r^2 and the maximum likelihood and score statistics

$$q^2 = (\hat{\theta} - \theta_0)'\hat{j}(\hat{\theta} - \theta_0) \ , \quad t^2 = s(\theta_0;y)'\hat{j}^{-1}s(\theta_0;y) \ , \tag{2.1}$$

where $s(\theta;y)$ is the score vector $(\partial/\partial\theta)l(\theta;y)$. Significance probabilities for testing θ_0 can be assessed , on the basis of first order asymptotic theory, by referring r^2 , q^2 , or t^2 to a χ^2_k distribution. The accuracy of the approximation to the significance probability is likely to be better for r^2 than for q^2 or t^2 , if indications from the one-dimensional case hold (for example, Cox, 1988; Fraser, 1991). In the multiparameter setting the three statistics can be measuring quite different versions of departure, thus raising a deeper issue separate from the computation of the tail probability. The differences between the statistics largely disappear, however, when considered from the conditional viewpoint discussed next.

Let y_0 be defined by $\hat{\theta}(y_0) = \theta_0$. Fraser and Massam (1985) proposed a conditional assessment of the magnitude of departure $|y - y_0|$ given the direction of departure $(y - y_0)/|y - y_0|$. The test was called a conical test, although the more descriptive term directional test used in Skovgaard (1987, 1988) seems more appropriate. Directional tests can be defined for a continuous statistical model $f(y;\theta)$ with variable and parameter of dimension k and are particularly appropriate when the variable y has some natural vector space properties, as is the case with the exponential family model. Directional tests for the score space of a general model were proposed in Fraser (1987). An example of the use of directional tests in the multivariate linear model is given in Fraser, Guttman and Srivastava (1991). For an exponential family model Skovgaard (1988) proposed a directional test based on the saddlepoint inversion (1.4), and derived a significance probability function with accuracy $O(n^{-1})$.

2.2 The conditional directional distribution for the exponential model

Now consider the derived exponential family model $f(y;\theta)$ given by (1.4) with log-likelihood $l(\theta;y)$ given by (1.6). We derive a directional test for testing $\theta = \theta_0$ in the model $f(y;\theta)$ given by (1.4) and log-likelihood (1.6). The test depends only on the observed log-likelihood and values for an appropriately defined score variable.

For the hypothesized θ_0 with data y^0 we will condition on the direction defined by $d = d^0$, where

$$d = \frac{y - y_0}{|y - y_0|}, \quad d^0 = \frac{y^0 - y_0}{|y^0 - y_0|}, \tag{2.2}$$

and y_0 is the data point with maximum likelihood value θ_0. Let $u = d^0 \cdot (y - y_0)$ be the signed departure along the line L through y_0 and y^0 and $\psi = d^0 \cdot \theta = \sum d_i^0 \theta_i$ be a corresponding parameter. Let $A = (d^0 : A_2)$ be a $k \times k$ orthogonal matrix and consider the change of variable and parameter

$$\begin{pmatrix} u \\ v \end{pmatrix} = A'(y - y_0) \ , \ \begin{pmatrix} \psi \\ \lambda \end{pmatrix} = A'\theta. \tag{2.3}$$

The underlying distribution along the line L can again be approximated by the saddlepoint method, as at (1.4), (1.5), and (1.7). Write $l^*(\psi, \lambda) = l(\theta)$ for the log-likelihood function in the new parameterization. The profile log-likelihood for ψ, $l_p(\psi) = l^*(\psi, \hat{\lambda}_\psi)$, satisfies

$$l_p(\psi; u) = (u - u^0)\psi + l_p(\psi; u^0) \ , \tag{2.4}$$

where $u^0 = d^0 \cdot (y^0 - y_0) = |y^0 - y_0|$. The approximation to the conditional density of u along the line is thus

$$f(u|L) \doteq \frac{c}{(2\pi)^{1/2}} \exp(-r_p^2/2) \hat{j}_p^{-1/2} \left\{ \frac{|j_\lambda(\hat{\psi}, \hat{\lambda})|}{|j_\lambda(\psi, \hat{\lambda}_\psi)|} \right\}^{-1/2} \tag{2.5}$$

where the constant c differs from 1 by $O(n^{-1})$ and $r_p^2/2$ is obtained from the tilted profile (2.4). The right hand side of (2.5) is the inversion of the profile log-likelihood times an information adjustment, and in the notation of (1.7) can be expressed as

$$cf(u; \psi | l_p) \left\{ \frac{|j_\lambda(\hat{\psi}, \hat{\lambda})|}{|j_\lambda(\psi, \hat{\lambda}_\psi)|} \right\}^{-1/2} . \tag{2.6}$$

The conditional distribution of u given the radial direction d^0 of departure is obtained by multiplying the approximate density in (2.5) by the conical expansion factor u^{k-1} which provides the Jacobian effect of the transformation to radial coordinates (Fraser & Massam, 1985):

$$f(u \mid d^0) \ = \ \frac{c_1}{(2\pi)^{1/2}} \exp(-r_p^2/2) \hat{j}_p^{-1/2} \left\{ \frac{|j_\lambda(\hat{\psi}, \hat{\lambda})|}{|j_\lambda(\psi, \hat{\lambda}_\psi)|} \right\}^{-1/2} u^{k-1}; \qquad u > 0, \tag{2.7}$$

where c_1 is a normalizing constant to be determined.

2.3 The significance function

The conditional significance is then given by one dimensional integration of (2.7):

$$p(\theta_0) = pr(u \geq u^0 \mid d^0; \theta_0) = \frac{\int_{u^0}^{\infty} e^{-r_p^2/2} \hat{j}_p^{-1/2} \left\{ \frac{|j_\lambda(\hat{\psi},\hat{\lambda})|}{|j_\lambda(\psi_0,\hat{\lambda}_{\psi_o})|} \right\}^{-1/2} u^{k-1} du}{\int_0^{\infty} e^{-r_p^2/2} \hat{j}_p^{-1/2} \left\{ \frac{|j_\lambda(\hat{\psi},\hat{\lambda})|}{|j_\lambda(\psi_0,\hat{\lambda}_{\psi_0})|} \right\}^{-1/2} u^{k-1} du} \tag{2.8}$$

where r_p^2 gives log-likelihood ratio relative to ψ_0 and the related information expressions are calculated along the line L with positive values for u.

For the approximation, we choose a chi distribution with k degrees of freedom as the reference distribution for the calculations. The survivor and density functions are

$$\Xi_k(r) = \int_r^{\infty} \chi_k(r) dr \ , \quad \chi_k(r) = a_k r^{k-1} e^{-r^2/2} \tag{2.9}$$

where $a_k = 1/\{2^{(k/2)-1}\Gamma(k/2)\}$; we also need the residual life function for Ξ_k :

$$\Lambda_k(r) = \frac{\int_r^{\infty} \Xi_k(r) dr}{\int_0^{\infty} \Xi_k(r) dr} = \frac{\int_r^{\infty} \Xi_k(r) dr}{A_k}. \tag{2.10}$$

Calculation of the integral is simplified by replacing u by the standardized score $z = u\left(1 - \frac{\alpha_4 - \alpha_3^2}{4n}\right)$. This is a linear transformation of u and has the same first derivative as r at $r = 0$. The coefficients α_3 and α_4 are the standardized third and fourth order coefficients in the expansion of the negative of the profile $l_p(\psi)$. The score z is standardized by expected information so that it is just a rescaling of u; thus the constant factor divides out in (2.8) and so z can directly replace u in that expression.

For the integration in (2.8) we let a_k be an approximation to the norming constant and obtain the correct constant a later; thus

$$ap(\theta_0) = \int_{r^0}^{\infty} \chi_k(r) \left\{ \frac{|j_\lambda(\hat{\psi},\hat{\lambda})|}{|j_\lambda(\psi_0,\hat{\lambda}_{\psi_0})|} \right\}^{-1/2} \frac{r}{q} \left(\frac{z}{r}\right)^{k-1} \mathrm{d}r + O(n^{-3/2})$$

where $r = r_p$ is the log-likelihood ratio for ψ_0 obtained from the profile log-likelihood (2.4), $q = q_p$ is the maximum likelihood departure from ψ_0 standardized by observed information from $l_p(\psi; u^0)$, and $\hat{j}_p^{-1/2}\ du = (r/q)dr$ is obtained by differentiating $r^2/2$.

We then integrate by parts the discrepancy from the reference distribution

$$\begin{aligned} ap(\theta_0) &= \int_r^\infty \chi_k(r)\mathrm{d}r + \int_r^\infty r\left[\left(\frac{z}{r}\right)^{k-1}\left\{\frac{|j_\lambda(\hat{\psi},\hat{\lambda})|}{|j_\lambda(\psi_0,\hat{\lambda}_{\psi_0})|}\right\}^{-1/2}\frac{1}{q}-\frac{1}{r}\right]\chi_k(r)\mathrm{d}r \\ &\quad + O(n^{-3/2}) \\ &= \Xi_k(r) + \int_r^\infty \alpha_k(r)\chi_{k+1}(r)\mathrm{d}r + O(n^{-3/2}) \\ &= \Xi_k(r) + \alpha_k(r)\Xi_{k+1}(r) + \int_r^\infty \Xi_{k+1}(r)\mathrm{d}\alpha_k(r) + O(n^{-3/2}) \end{aligned} \tag{2.11}$$

where $\alpha_k(r)$ is a discrepancy function,

$$\alpha_k(r) = \frac{a_k}{a_{k+1}}\left\{\left(\frac{z}{r}\right)^{k-1}\left\{\frac{|j_\lambda(\hat{\psi},\hat{\lambda})|}{|j_\lambda(\psi_0,\hat{\lambda}_{\psi_0})|}\right\}^{-1/2}\frac{1}{q}-\frac{1}{r}\right\}, \tag{2.12}$$

and the unexpected sign is due to the use of the survivor function. Then using

$$\alpha_k(r) = b_k n^{-1/2} + c_k r n^{-1} + O(n^{-3/2}) \tag{2.13}$$

which is derived in Section 2.4 we obtain

$$\begin{aligned} p(\theta_0) &= \Xi_k(r)\{1 - b_k/n^{1/2} - c_k A_{k+1}/n + b_k^2/n\} \\ &\qquad + \Xi_{k+1}(r)\alpha_k(r)\{1 - b_k/n^{1/2}\} + \Lambda_{k+1}(r)A_{k+1}c_k/n + O(n^{-3/2}) \end{aligned} \tag{2.14}$$

where $a = 1 + \alpha_k(0) + c_k A_{k+1}/n$ is expanded on the right-hand side.

Wood, Booth and Butler (1993) develop a Lugannani and Rice type approximation with a chi kernel from a different point of view. Their approximation adjusts the values of r and q, rather than adjusting the integrand using $\alpha_k(r)$. They also mention the possibility of using the approximate conditional likelihood in their procedure.

The integration by parts in (2.13) is the same as that used to derive the Lugannani and Rice formula in the one dimensional case; see e.g. Barndorff-Nielsen and Cox (1989, Ch.4) and DiCiccio, Field and Fraser (1990). The normal kernel has the special feature that the first and third terms in (2.11) can be combined, making the resulting expression much simpler.

2.4 The adjustment constants

The constants $b_k n^{-1/2}$ and $c_k n^{-1}$ are the limiting value and first derivative of the adjustment factor $\alpha_k(r)$ in (2.12) at $r = 0$. Let $\alpha_3 n^{-1/2}$ and $\alpha_4 n^{-1}$ be the third and fourth standardized derivatives of the negative profile log-likelihood $l_p(\psi)$ at the tested value ψ_0. Also let $bn^{-1/2}$ and cn^{-1} be the first and second derivatives of $|j_\lambda(\hat{\psi}, \hat{\lambda})|/|j_\lambda(\psi, \hat{\lambda}_\psi)|$ at ψ_0. Then applying the information expansions given in Fraser & Reid (1993) to the profile log-likelihood, we obtain

$$\frac{|j_\lambda(\hat{\psi}, \hat{\lambda})|}{|j_\lambda(\psi, \hat{\lambda}_\psi)|} = 1 + \frac{b}{n^{1/2}}(\hat{\psi} - \psi_0) + \frac{c}{2n}(\hat{\psi} - \psi_0)^2 = 1 + \frac{B}{n^{1/2}} r + \frac{C}{2n} r^2$$

with $B = b$ and $C = c - 2b\alpha_3/3$. The intercept $b_k n^{-1/2}$ and slope $c_k n^{-1}$ for the expansion $\alpha_k(r) = b_k n^{-1/2} + c_k r n^{-1}$ are then given respectively by

$$(a_k/a_{k+1})\{(k-2)\alpha_3 - 3B\}/6n^{1/2},$$

$$(a_k/a_{k+1})[\{15\alpha_3^2 - 9\alpha_4 + (k-1)(3\alpha_4 - 7\alpha_3^2) + (k-1)^2\alpha_3^2\} - 6(k-2)\alpha_3 B + (27B^2 - 18C)]/72n.$$

Unless $l^*(\psi, \lambda)$ happens to be rather simple, the various quantities required for calculating $p(\theta_0)$ in (2.14) have to be determined numerically. We used a Newton-Raphson algorithm to find $\hat{\lambda}_\psi$, which in turn gives $l_p(\psi)$ and $|j_\lambda(\psi, \hat{\lambda}_\psi)|$. The derivatives of $l_p(\psi)$

and $|j_\lambda(\psi, \hat{\lambda}_\psi)|$ at $\hat{\psi}$ and ψ_0 needed for computing the adjustment factor $\alpha_k(r)$ can be obtained numerically by smoothing, or the intercept and slope of $\alpha_k(r)$ at $r = 0$ can be approximated from a chord at r-values greater than zero. The approximation is harder to calculate when there are many nuisance parameters to be eliminated in the original model, as (1.6) cannot usually be obtained explicitly. Skovgaard's approximation is easier to obtain, as all computations are at the null value θ_0, and a likelihood function for θ is not needed. His method seems to break down in the extreme case of Example 3.1 below, but the extra order of accuracy provided by our method is not practically important in Example 3.2.

3. Examples

3.1 Blend of long and short tails

Consider a two dimensional exponential model with very short tails on one axis and very long tails on the other: $f(y_1, y_2; \theta_1, \theta_2) = f_1(y_1; \theta_1) f_2(y_2; \theta_2)$ where

$$f_1(y;\theta) = (1/2)\exp\{y\theta - c_1(\theta)\}\ , \qquad c_1(\theta) = \log\left(\frac{\sinh\,\theta}{\theta}\right)$$

is an exponential tilt of the uniform distribution on $(-1, +1)$, and

$$f_2(y;\theta) = (2\pi)^{-1/2} y^{-3/2} \exp\{y\theta - c_2(\theta) - (y-1)^2/(2y)\}\ , \qquad c_2(\theta) = 1 - (1-2\theta)^{1/2}$$

is an exponential tilt of the inverse gaussian distribution on $(0, \infty)$. We test the hypothesis $(\theta_1, \theta_2) = (0, 0)$, which corresponds to the uniform for y_1 and the inverse gaussian for y_2; the point $y_0 = (0, 1)$ has maximum likelihood estimate equal to the hypothesized value $(0, 0)$.

We examine significance approximations for data points in various directions from the null point $y_0 = (0,1)$: $0°, 22.5°, 45°, 67.5°, 90°$ with $0°$ corresponding to the uniform direction, i.e. the y_1 axis, and $90°$ corresponding to the inverse gaussian direction, i.e. the y_2 axis. In the five directions we choose observed data points having the same value, 9.97 for the log-likelihood ratio r^2: the values of the three standard statistics are given in Table 1.

Table 1. Values for the test statistics in five directions

Direction	0°	22.5°	45°	67.5°	90°
t^2	39090.0	34610.0	23740.0	7176.0	0.0706
q^2	1.00	1.17	2.11	9.09	413.87
r^2	9.97	9.97	9.97	9.97	9.97

In Table 2 we record the significance probability for the 5 points by the first order chi-squared approximations for t^2, q^2, r^2, together with the exact marginal probability for r^2, the second order approximation of Skovgaard (1988), the third order chi-based approximation (2.14), and the exact probability obtained by numerical integration. We also examine the particular intermediate direction ($45°$) where there is a nuisance parameter effect, and record the approximations and exact values at five points at various distances from the null point in Table 3. In this direction, an improvement in the use of the approximation (2.14) for significance levels above 20% was obtained by using the asymptotic version (2.13) in place of the exact formula (2.12) for the discrepancy function $\alpha_k(r)$.

Table 2. Tail probability approximations in several directions

Direction	0°	22.5°	45°	67.5°	90°
χ^2 approx. for t^2	0.0	0.0	0.0	0.0	0.9653
χ^2 approx. for q^2	0.6065	0.5557	0.3482	0.0106	0.0
χ^2 approx. for r^2	0.0068	0.0068	0.0068	0.0068	0.0068
Marginal r^2	0.0144	0.0144	0.0144	0.0144	0.0144
Skovgaard	0.0122	0.0127	0.0185	0.0111	0.0034
Third order	0.0116	0.0090	0.0085	0.0089	0.0042
Conditional exact	0.0100	0.0088	0.0083	0.0087	0.0041

Table 3. Tail probabilities along the 45° *line*

r^2	1.508	3.502	8.196	10.977	12.761
Marginal r^2	0.5693	0.2543	0.0324	0.0091	0.0039
Skovgaard	0.8314	0.3981	0.0443	0.0112	0.0046
Third order	0.4096	0.1871	0.0204	0.0051	0.0021
Conditional exact	0.5005	0.1998	0.0200	0.0050	0.0020

The first order approximations give nominal marginal probabilities and are thus strongly influenced by the different kinds of departure being measured. The second and third order approximations give conditional probabilities and are independent of the kind of departure measure. The third order approximation does not seem to have the extreme accuracy usually found with the Lugannani and Rice type approximation; this may be due to the increased weight on the tail of the distribution given by the Jacobian factor u^{k-1}.

3.2 Equality of exponential rates

In this example we consider comparing three exponential rates to a common known value. We use the data on three of the aircraft from Example T in Cox & Snell (1981), and compare their rates to the value $\lambda_0 = 199/18084$, which is the maximum likelihood estimate based on the data from all ten aircraft. The aircraft we chose were numbers 1, 3 and 7, with 23, 15 and 24 observations respectively. The observed values of the statistics r, q and z needed for evaluation of $\alpha_k(r)$ are 2.011527, 1.874821, and 1.954995, respectively. The coefficient b_k and c_k were obtained from a chord passing through points at $r_1 = 0.09$ and $r_2 = 1.58$ of $\alpha_k(r)$, which is quite linear up to $r = 2$. The significance probabilities for the first order normal theory method, Skovgaard's second order method, and our third order method from (2.14) are compared to the exact conditional p-value in Table 4. The exact value was obtained by numerical integration of the exact conditional density, given in Fraser and Massam (1985).

Table 4. Tail probabilities for testing equality of exponential rates

χ^2_3 for r^2	0.2565
Skovgaard	0.2393
Third order	0.2405
Conditional exact	0.2407

A more practical treatment of this example would test equality of rates treating the common value as a nuisance parameter. In principle this poses no further difficulty, nor indeed does extension to a gamma model with common shape parameter, but expression (1.6) is only available indirectly. The restricted maximum likelihood estimate, $\hat{\kappa}_\theta$ in the notation of Section 2, is a solution of a cubic equation, or in the case of comparing d rates, is a root of a polynomial equation of degree d. We would expect that the linear approximation to $\alpha_k(r)$ would break down with increasing dimension of the parameter of interest, although numerical integration also becomes more difficult in this case. A reliable and robust computational procedure for problems with large dimension would require more attention to numerical details than we have provided here.

Acknowledgements We would like to thank the referees and Associate Editor for very

helpful comments.

References

Barndorff-Nielsen, O.E. (1990). Approximate interval probabilities. *J. Royal Statist. Soc. B* **52**, 485–96.

Barndorff-Nielsen, O.E. (1991). Modified signed log likelihod ratio. *Biometrika* **78**, 557–63.

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

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Cox, D.R. (1988). Some aspects of conditional and asymptotic inference : a review. *Sankhyā* A **50**, 314–37.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Royal Statist. Soc. B* **49**, 1–39.

Cox, D.R. and Snell, E.J. (1981). *Applied Statistics : Principles and Examples.* Chapman and Hall, London.

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–50.

Fraser, D.A.S. (1987). Fibre analysis and tangent models. *Statistical Papers* **28**, 163–81.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods. *Biometrika* **77**, 65-76.

Fraser, D.A.S. (1991). Statistical inference: likelihood to significance. *J. Amer. Stat. Assoc.* **86**, 258–65.

Fraser, D.A.S., Guttman, I. and Srivastava, M.S. (1991). Conditional inference for treatment and error in multivariate analysis. *Biometrika* **78**, 565–572.

Fraser, D.A.S. and Massam, H. (1985). Conical tests: observed levels of significance and confidence regions. *Statistical Papers* **26**, 1–17.

Fraser, D.A.S. and Reid, N. (1993). Simple asymptotic connection between density and cumulant generating functions leading to accurate approximations for distribution functions. *Statist. Sinica* **3**, 67-82.

Fraser, D.A.S., Reid, N. and Wong, A. (1991). Exponential linear models: a two pass procedure for saddlepoint approximation. *J. Royal Statist. Soc. B.* **52**, 483–492.

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

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

Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *J. R. Stat. Soc. B* **50**, 269–280.

Wood, A.T.A., Booth, J.G. and Butler, R.W. (1993). Saddlepoint approximations to the cdf of some statistics with non-normal limit distributions. *J. Amer. Statist. Assoc.* **88**, 680–686.