ADJUSTMENTS TO LIKELIHOOD AND DENSITIES: CALCULATING SIGNIFICANCE

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

Tangent exponential models at a data point lead to third order significance for component parameters (Fraser & Reid, 1989, 1993a,b,1995). With nuisance parameters, adjustment factors to a tangent exponential model can arise; for example, with conditioning for exponential models and with marginalization for transformation and general asymptotic models. Methods of calculating these adjustments are discussed and different versions of the resulting significance formulas are derived. Examples and numerical comparisons are given.

1. INTRODUCTION

Saddlepoint methods (Daniels, 1954; Barndorff-Nielsen & Cox, 1979; Lugannani & Rice, 1980) provide density and distribution function approximations for the minimal sufficient statistic in a full exponential model. For the model $f(x;\theta) = \exp\{\theta y(x) - c(\theta)\}h(x)$ with θ as a row vector, the density approximation for y of dimension p,

$$\hat{f}(y;\theta) = \frac{c}{(2\pi)^{p/2}} \exp\Big\{-\frac{r^2}{2}\Big\}\hat{j}^{-1/2} \ , \tag{1.1}$$

is accurate to order $O(n^{-3/2})$, where $c = 1+O(n^{-1})$ is a constant, $-r^2/2 = \ell(\theta;y)-\ell(\hat{\theta};y)$ is the log likelihood ratio, and $\hat{j}$ is the observed information for θ. For the case $p = 1$, the distribution function approximation (Lugannani & Rice, 1980)

$$\hat{F}(y;\theta) = \Phi(r,q) = \Phi(r) + \varphi(r)\Big\{\frac{1}{r} - \frac{1}{q}\Big\} \tag{1.2}$$

is accurate to order $O(n^{-3/2})$ where $q = (\hat{\theta}-\theta)\hat{j}^{1/2}$ is the standardized maximum likelihood departure and $r = \mathrm{sgn}(q)(r^2)^{1/2}$ is the signed log likelihood ratio statistic. This assumes that the initial model is asymptotic as a sample size parameter $n \to \infty$.

An alternative formula with argument r^* is available from Barndorff-Nielsen (1991)

$$\hat{F}(y;\theta) = \Phi\{r - r^{-1}\log(r/q)\} + O(n^{-3/2}) \ . \tag{1.3}$$

Earlier formulas similar to this may be found, for example Barndorff-Nielsen(1986), but the components were different, involving mean and variance corrections. In applications the preceding formulas are found to have remarkable practical accuracy, seemingly due to the use of likelihood invariants rather than the mean and variance type corrections as for example in Bartlett (1955), McCullagh (1984), Barndorff-Nielsen (1986).

Pierce & Peters (1992) give a detailed survey of such approximations for discrete exponential models, and find it valuable to distinguish between two influences on correction terms such as the second term in (1.2): nonnormality for the likelihood ratio; and nuisance

parameter effects. They also discuss different versions of the approximation in (1.2). We examine some general aspects of the correction terms and of the approximations related to (1.2) and (1.3).

In Section 2 we briefly survey some general model density and distribution function approximations. In Section 3, for an exponential model, we examine the norming of likelihood or density adjustment factors and then determine in Section 4 the corresponding distribution function. Some alternative formulas are developed in Section 5 and numerical comparisons presented in Section 6.

2. SOME BACKGROUND

For a general asymptotic model under quite general conditions, Fraser & Reid (1993b, 1995) describe a method of reducing the dimension n of the basic variable to the dimension p of the parameter and determining the minimum additional information for third order inference; all that is needed are p tangent vectors $v_1, \ldots, v_p$ to the conditioning surface at the observed data point y^0. Then, relative to the data point y^0, we obtain an approximating exponential model

$$g(y;\theta)dy = \frac{c}{(2\pi)^{p/2}} \exp\{\ell(\theta;y^0) - \ell(\hat{\theta};y^0) + (\varphi - \hat{\varphi}^0)s\}\tilde{j}_{\varphi}^{-1/2}ds \tag{2.1}$$

where

$$\varphi = \varphi(\theta) = \ell_{;V}(\theta;y^0) = \frac{d}{dV}\ell(\theta;y)|_{y^0} \tag{2.2}$$

is the gradient of the likelihood in the tangent directions $V = (v_1, \ldots, v_p)$ at the data point y^0,

$$s = s(y) = \ell_{\varphi}(\hat{\theta}^0;y) = \frac{\partial}{\partial\varphi'}\ell(\theta;y)|_{\hat{\theta}^0} \tag{2.3}$$

is the corresponding score variable, and $\tilde{j}_{\varphi}$ is the observed information for φ from the tilted likelihood in (2.1); the approximation is accurate to $O(n^{-3/2})$ in a first derivative neighbourhood of y^0 and to $O(n^{-1})$ on a compact set for the standardized variable; the determination of the vectors V is discussed in Fraser & Reid (1993b, 1995).

Also, for example, if $p = 1$, the Lugannani & Rice (1990) approximation (1.2) applied to (2.1) at the data point y^0 agrees to order $O(n^{-3/2})$ with the distribution function of the conditioned model described above.

More generally, with $\theta = (\lambda, \psi)$ and a scalar interest parameter ψ, a further reduction can be obtained to a scalar variable that measures departure from a particular value for ψ. The approximating distribution is an adjusted exponential model

$$g(y;\chi)A(y;\chi)dy \tag{2.4}$$

where $g(y;\chi)$ is exponential and $A(y;\chi) = 1 + O(n^{-1/2})$ is an adjustment factor; the parameter $\chi = \sum a_i\varphi_i$ is a linear function of φ in (2.2) such that $d\chi = 0$ means $d\psi = 0$ at $\hat\theta^0_\psi$; at the data point y^0

$$A(y;\chi) = \frac{|j_{(\lambda)}(\hat\theta^0)|^{-1/2}}{|j_{(\lambda)}(\hat\theta^0_\psi)|^{-1/2}} \tag{2.5}$$

is a quotient of nuisance parameter informations, recalibrated on the φ scaling; for details see Fraser & Reid (1993b, 1995). This adjusted exponential model gives $O(n^{-3/2})$ significance for ψ at the data point y^0; we examine this generally in the next section.

An alternative formula with a different expression replacing q is given by Barndorff-Nielsen (1991); the formula however needs the variable and parameter to be of the same dimension; thus as a separate issue an appropriate ancillary must be found as for example by the methods using the vectors V discussed above.

3. ADJUSTMENTS TO DENSITIES AND LIKELIHOOD

Consider an adjusted exponential model $g(y;\theta)A(y;\theta)dy$ as defined at (2.4) but now assume that y and θ have dimension d.

To derive properties of this model it is convenient to use standardized canonical coordinates for the exponential model $g(y;\theta)$ relative to some initial data point y^0 of interest. Let a new θ measure departure of some initial θ from $\hat\theta^0$ in units of observed information

for that parameter, and let a new y measure departure of the initial y from y^0 scaled so that it is the score variable for the new θ; for details see Fraser & Reid (1993a, 1995). Also we rewrite the adjustment as $A(\hat{\theta};\theta)$ where $\hat{\theta}$ is the maximum likelihood estimate as calculated from the exponential component model $g(y;\theta)$. The model then has the form

$$\frac{c}{(2\pi)^{d/2}} e^{-r^2/2} |\hat{j}|^{-1/2} A(\hat{\theta};\theta) dy \tag{3.1}$$

where $-r^2/2 = \ell^0(\theta) - \ell^0(\hat{\theta}^0) + \theta y$, $A(\hat{\theta};\theta) = 1 + O(n^{-1/2})$; the accuracy is as described after (2.3). For the analysis, we have

$$E(\hat{\theta}) = \theta + bn^{-1/2} + O(n^{-1}) \ , \ E(\hat{\theta}'\hat{\theta}) = I + \theta'\theta + O(n^{-1/2}) \tag{3.2}$$

based on the approximate d variate standard normal for y and for $\hat{\theta}$.

Sometimes, the adjustment factor may be only partially available $\bar{A}(\hat{\theta};\theta)$ as a likelihood; then $A(\hat{\theta};\theta) = \bar{A}(\hat{\theta};\theta)\ h(\hat{\theta})$ with h unknown. Other times, the adjustment factor may be only partially available $\bar{A}(\hat{\theta};\theta)$ as a relative density: then $A(\hat{\theta};\theta) = \bar{A}(\hat{\theta};\theta)h(\theta)$ with the norming constant $h(\theta)$ unknown. We examine these cases in an $O(n^{-3/2})$ context.

We expand to the second order in $\hat{\theta}$ and θ:

$$A(\hat{\theta};\theta) = a_{00} + \frac{a_{10}}{n^{1/2}}\hat{\theta} + \frac{a_{01}}{n^{1/2}}\theta + \frac{a_{20}}{2n}\hat{\theta}^2 + \frac{2a_{11}}{2n}\hat{\theta}\theta + \frac{a_{02}}{2n}\theta^2 + O(n^{-3/2}) \ , \tag{3.3}$$

where for simplicity of presentation we examine just the $d = 1$ case. The adjustment factor must have expectation unity with respect to the exponential model; thus by substitution from (3.2) we obtain $E_g\{A(\hat{\theta};\theta)\} = a_{00} + \frac{a_{10}+a_{01}}{n^{1/2}}\theta + \frac{a_{20}+2a_{11}+a_{02}}{2n}\theta^2 + \frac{B}{n} + O(n^{-3/2})$ and thus

$$a_{00} = 1 - \frac{B}{n} + O(n^{-3/2}) \ , \ a_{10} + a_{01} = 0 + O(n^{-1}) \ , \ a_{20} + 2a_{11} + a_{02} = 0 + O(n^{-1/2}) \ . \tag{3.4}$$

The same holds for general d using vector and matrix coefficients.

Now suppose the adjustment factor $\bar{A}(\hat{\theta};\theta)$ is known only as a likelihood; then

$$A(\hat{\theta};\theta) = \bar{A}(\hat{\theta};\theta)h(\hat{\theta}) \ ,$$

$$A(\hat{\theta};\hat{\theta}) = \bar{A}(\hat{\theta};\hat{\theta})h(\hat{\theta}) \ .$$

It follows that

$$A(\hat{\theta};\theta) = \frac{\bar{A}(\hat{\theta};\theta)}{\bar{A}(\hat{\theta};\hat{\theta})} \cdot \big(1 - \frac{B}{n}\big) \tag{3.5}$$

and the model can be written

$$\frac{c'}{(2\pi)^{d/2}} \exp\{-r^2/2\} |\hat{j}|^{-1/2} \frac{\bar{A}(\hat{\theta};\theta)}{\bar{A}(\hat{\theta};\hat{\theta})} dy \tag{3.6}$$

where the extra constant is incorporated into the c'; note that we have normed by replacing θ by $\hat{\theta}$.

Alternatively, suppose the adjustment is known only as a density function with norming constant; then

$$A(\hat{\theta};\theta) = \bar{A}(\hat{\theta};\theta) h(\theta) \ ,$$

$$A(\theta;\theta) = \bar{A}(\theta;\theta) h(\theta) \ .$$

It follows that

$$A(\hat{\theta};\theta) = \frac{\bar{A}(\hat{\theta};\theta)}{\bar{A}(\theta;\theta)} \big(1 - \frac{B}{n}\big) \tag{3.7}$$

and the model can be written

$$\frac{c''}{(2\pi)^{d/2}} \exp\big\{ - \frac{r^2}{2} \big\} |\hat{j}|^{-1/2} \frac{\bar{A}(\hat{\theta};\theta)}{\bar{A}(\theta;\theta)} dy \tag{3.8}$$

where the extra constant is incorporated into the c''; here we have normed by replacing $\hat{\theta}$ by θ.

Example 1. Consider an exponential model $\exp\{\lambda y_1 + \psi y_2 - c(\lambda,\psi)\} h(y_1, y_2)$ where λ, ψ have dimensions $p-d$, d. The distribution for inference concerning ψ is $y_2 | y_1 = y_1^0$ which is free of λ.

The full model likelihood at a point (y_1^0, y_2) is $\ell(\lambda, \psi; y_1^0, y_2)$ in the usual notation.

For a fixed value of ψ, y_1 is a sufficient statistic for λ: thus $\hat{\lambda}_\psi = \hat{\lambda}_\psi^0$ for points (y_1^0, y_2) and the density of y_1 at y_1^0 can be approximated by (1.1) as

$$\frac{c}{(2\pi)^{(p-d)/2}} \ \exp\{\ell(\lambda,\psi;y_1^0,y_2) - \ell(\hat{\lambda}_\psi^0,\psi;y_1^0,y_2)\} |j_\lambda(\hat{\lambda}_\psi^0, \psi : y_1^0)|^{-1/2} \tag{3.9}$$

The conditional log likelihood is then available by subtraction

$$\ell(\hat{\lambda}^0_\psi, \psi; y^0_1, y_2) + \frac{1}{2} \ln |j_\lambda(\hat{\lambda}^0_\psi, \psi : y^0_1)| \ .$$

The profile likelihood is of exponential type

$$\ell(\hat{\lambda}^0, \psi; y^0_1, y_2) = \ell^0(\hat{\lambda}^0, \psi) + \psi y_2 \ .$$

Thus we can use (3.5) and obtain the adjustment

$$\frac{|j_\lambda(\hat{\lambda}_\psi, \psi)|^{1/2}}{|j_\lambda(\hat{\lambda}, \hat{\psi})|^{1/2}} \ .$$

This corresponds to the approximation

$$\frac{c}{(2\pi)^{d/2}} \ \exp\{-\frac{r_p^2}{2}\} \frac{|\hat{j}|^{-1/2}}{|j_\lambda(\hat{\lambda}_\psi, \psi)|^{-1/2}} dy_2$$

for the conditional model for inference concerning ψ, where $-r_p^2/2 = \ell(\hat{\lambda}_\psi, \psi) - \ell(\hat{\lambda}, \hat{\psi})$ is the profile likelihood for ψ; see Barndorff-Nielsen(1979), Cox & Reid (1987), Fraser & Reid (1993a).

4. SIGNIFICANCE AT OBSERVED DATA

Consider an adjusted exponential model $g(y;\theta)A(\hat{\theta};\theta)dy$ for the scalar $d = 1$; we analyse the model in the equivalent form (3.1):

$$\frac{c}{\sqrt{2\pi}} e^{-r^2/2} A(\hat{\theta};\theta) |\hat{j}|^{-1/2} dy \ . \tag{4.1}$$

Without an adjustment, the corresponding distribution function to order $O(n^{-3/2})$ is given by (1.2). We now examine the effect of the adjustment factor.

First we change the variable (Fraser & Reid, 1993a) and have

$$\hat{j}^{-1/2} dy = \frac{r}{q} dr \tag{4.2}$$

where r, q are defined for the exponential model after (1.2). Then from Fraser & Reid (1993a) or Abebe et al (1993) we have that $r/q = 1$ to order $O(n^{-1/2})$. For the integration we will need the asymptotic form of rA/q. For given θ we expand in terms of r:

$$\frac{rA}{q} = 1 + a_1 r/n^{1/2} + a_2 r^2/2n + O(n^{-3/2}) \ . \tag{4.3}$$

Now consider the distribution function for the model (4.1)

$$\begin{aligned} F(y;\theta) &= \int_{-\infty}^{r} \frac{c}{\sqrt{2\pi}} e^{-r^2/2} A \frac{r}{q} dr \\ &= c\Phi(r) + c\int_{-\infty}^{r} \varphi(r)\big(A\frac{r}{q} - 1\big)dr \\ &= c\Phi(r) + c\varphi(r)\big(\frac{1}{r} - \frac{A}{q}\big) + c\int_{-\infty}^{r} \varphi(r) d\big(\frac{A}{q} - \frac{1}{r}\big) \\ &= c\Phi(r) + c\varphi(r)\big(\frac{1}{r} - \frac{A}{q}\big) + c(a_2/2n)\Phi(r) \\ &= \Phi(r) + \varphi(r)\big(\frac{1}{r} - \frac{1}{Q}\big) \ , \end{aligned} \tag{4.4}$$

with $Q = q/A$; in these calculations, we use $r\varphi(r)dr = -d\varphi(r)$ before integrating by parts, $d(A/q - 1/r) = (a_2/2n)dr$, and then (4.3) and the norming property of the distribution functions to consolidate the constants. It follows that the distribution function is given by (1.2) or (1.3) with q replaced by $Q = q/A$.

5. ALTERNATIVE FORMULAS FOR SIGNIFICANCE

Consider an adjusted exponential model in the form (4.1),

$$\frac{c}{(2\pi)^{1/2}} e^{-r^2/2} \ A\hat{j}^{-1/2} dy; \tag{5.1}$$

we assume that a likelihood $-r^2/2 = \ell^0(\varphi) - \ell^0(\hat{\varphi}^0) + \varphi y$ is expressed in terms of a canonical parameter φ and has corresponding score variable y and information $\hat{j}$; also that the adjustment factor is asymptotic as at (2.4) and that interest lies in significance

$p(\varphi) = P(\hat{\varphi} \leq \hat{\varphi}^0; \varphi)$ for the data value y^0 underlying the exponential model expression. From Section 4 we have

$$p(\varphi) = \Phi(r, q/A) = \Phi(r) + \varphi(r)\big(\frac{1}{r} - \frac{A}{q}\big) \tag{5.2}$$

to third order; r and q are the signed likelihood ratio and the standardized maximum likelihood departure in the φ scaling; see the definitions following (1.2). We refer to this as the initial method M_0 with adjusted maximum likelihood estimate.

A numerical procedure for calculating the significance function $p(\varphi)$ with $A = 1$ is described in Fraser, Reid, & Wong (1991). The procedure needs the input of a grid of values for the likelihood ℓ^0, which may be in terms of some initial parameterization θ, and corresponding values for the canonical parameter φ. The computation produces the significance function expressed in terms of φ or in terms of the initial θ. If an adjustment factor A is present, it can also be input to the program, giving the more general significance function $p(\varphi)$ in (5.2).

In summary, if we have a likelihood ℓ, a parameterization φ, and an adjustment A, then by the program or directly we can calculate the significance function. For this we note that an affine change in φ to say $a + c\varphi$ has no effect on the calculations; it just alters the canonical parameterization with compensating adjustments otherwise.

We now discuss some alternative formulas that arise from shifting overall likelihood contained in A into the likelihood of the exponential model. For this we assume the initial φ arises as a likelihood gradient (2.2).

Consider a factorization $A = A_1 A_2$ such that each component has the asymptotic properties at (2.1). We suppose that the component adjustment A_1 is moved to the exponent to give a new exponential likelihood,

$$-\frac{r_1^2}{2} = -\frac{r^2}{2} + \ln A_1 + a \tag{5.3}$$

where a constant a may be needed to give maximum likelihood equal to zero.

Does this alter the canonical parameterization associated with the exponential model? Using the standardized coordinates as at (3.1) and expanding $\log A_1$, we obtain

$$\log A_1 = a_{00} + \frac{a_{10}}{n^{1/2}}\theta + \frac{a_{01}}{n^{1/2}}\hat{\theta} + \frac{a_{20}}{2n}\theta^2 + \frac{2a_{11}}{2n}\theta\hat{\theta} + \frac{a_{02}}{2n}\hat{\theta}^2 \tag{5.4}$$

with gradient relative to say $\hat{\theta}$

$$\frac{a_{01}}{n^{1/2}} + \frac{a_{11}}{n}\theta \tag{5.5}$$

at the data y^0 with $\hat{\theta}^0 = 0$; this makes only an affine adjustment to the initial φ and can be ignored as noted earlier. We can obtain significance numerically (Fraser, Reid, Wong, 1991) or directly with

$$p(\varphi) = \Phi(r_1) + \varphi(r_1)\{\frac{1}{r_1} - \frac{A_2}{q_1}\} \tag{5.6}$$

where r_1 is the signed likelihood ratio from the modified exponent, and q_1 is the maximum likelihood departure using the original φ but standardized with respect to the modified likelihood.

Consider a first modified formula M_1 obtained by placing the full adjustment factor into the exponent

$$-\frac{r_1^2}{2} = -\frac{r^2}{2} + \ln A + a \ . \tag{5.7}$$

This gives a new signed likelihood r_1, the same canonical parameter φ with departure q_1 standardized with respect to r_1, and no adjustment factor; we thus have a first formula M_1 with adjusted likelihood:

$$p_1(\varphi) = \Phi(r_1, q_1) \ . \tag{5.8}$$

For a second modified formula M_2 we move the combined Ar/q to the exponent

$$-\frac{r_2^2}{2} = -\frac{r^2}{2} + \ln\big(A\frac{r}{q}\big) + a \ . \tag{5.9}$$

In the pattern of (5.3) this makes a quadratic contribution to the new $-r_2^2/2$ and correspondingly makes r_2 an affine function of r with scale factor of order $O(n^{-1})$. This gives a second formula M_2 with a normalized likelihood:

$$p_2(\varphi) = \Phi(r_2) = \Phi(r_2, r_2) \tag{5.10}$$

A third formula is obtained by a minor modification of the preceding. We complete the square for the expression on the right side of (5.9),

$$-\frac{r_2^2}{2} = -\frac{1}{2}\Big\{r - r^{-1}\ln\big(A\frac{r}{q}\big)\Big\}^2, \tag{5.11}$$

and note that $\{r^{-1}\ln(Ar/q)\}^2$ is a constant K/n using calculations as in (5.4). This, with the preceding, then gives a third formula M_3 based on a normalized variable

$$p_3(\varphi) = \Phi\{r - r^{-1}\ln(Ar/q)\}\ , \tag{5.12}$$

which gives an extension of Barndorff-Nielsen's (1991) r^* formula to cover the adjusted density case.

We now have four formulas M_0, M_1, M_2, M_3 for $p(\varphi)$ given by (5.2), (5.8), (5.10), (5.12).

6. EXAMPLES

For the general case of a real variable and a real parameter with no adjustment factor A, the formulas M_0 and M_1 are identical. The third order accuracy of the r^* formula M_3 is established in Barndorff-Nielsen (1990,1991); the third order accuracy of an earlier version of r^*, operationally using mean and variance connections, was given in Barndorff-Nielsen (1986). The third order accuracy of $M_0 = M_1$ was established in Fraser & Reid (1989; see also 1993); Barndorff-Nielsen (1991) indicates that the accuracy can be established using a formula in Barndorff-Nielsen (1986). We compare these asymptotic formulas for the extreme case with $n = 1$.

For the special case of an exponential model the third order accuracy of $M_0 = M_1$ was established by Lugannani & Rice (1980) and we could expect high practical accuracy by viewing the formula as an approximate Fourier inversion. In a range of examples we have found that M_0, M_1, M_2, M_3 all give excellent accuracy, two figure accuracy verging

on three figure accuracy even in the extreme tails. Some minor preferences are indicated as follows,

Logistic	$M_2 > M_0 \geq M_3$
Gamma (3)	$M_2 > M_0 \geq M_3$
Beta $(3,3)$	$M_0 \geq M_3 \geq M_2$

using selected points on the range of the variable; overall $M_0 = M_1$ seems preferable.

For nonexponential models we record results for a number of location models. A more general model obtained (Fraser & Reid, 1993) as a blend of a location and exponential model shows no new features. Consider the location examples in Fraser (1990) and Barndorff-Nielsen (1991),

Log gamma(3)	$(1/2)\exp\{3(y-\theta) - e^{y-\theta}\}$
Gamma(3)	$(1/2)(y-\theta)^2 \exp\{-(y-\theta)\}$
Logistic	$\exp\{y-\theta\}(1+e^{y-\theta}\}^{-2}$
Cauchy	$\pi^{-1}\{1+(y-\theta)^2\}^{-1}$,

for selected values of the variable. In each case the tail probability is recorded as a percent and for the left tail or right tail (marked with R) as appropriate; see Tables 1 - 4.

Table 1. Log gamma (3)

$y-\theta$	-1.0	0.0	1.0^R	1.5^R	2.0^R	2.5^R
M_0	0.63	8.04	48.92	17.58	2.21	0.045
M_2	0.63	8.05	48.89	17.56	2.21	0.045
M_3	0.63	8.02	48.96	17.60	2.21	0.045
Exact	0.63	8.03	48.92	17.56	2.21	0.045

Table 2. Gamma (3)

$y-\theta$	1.0	3.0^R	5.0^R	7.0^R	10.0^R	12.0^R
M_0	7.30	43.28	12.83	3.06	0.29	0.054
M_2	8.29	41.78	12.22	2.89	0.27	0.051
M_3	7.78	42.75	12.65	3.02	0.28	0.053
Exact	8.03	42.32	12.47	2.96	0.28	0.052

Table 3. Logistic

$y-\theta$	-8.0	-6.0	-4.0	-2.0	-1.0	-0.5
M_0	0.037	0.27	1.91	12.22	27.15	37.90
M_2	0.038	0.28	1.96	12.42	27.35	38.02
M_3	0.037	0.27	1.90	12.21	27.15	37.90
Exact	0.034	0.25	1.80	11.92	26.89	37.75

Table 4. Cauchy

$y-\theta$	-100.0	-30.0	-10.0	-5.0	-2.0	-1.0
M_0	0.28	0.94	2.81	5.58	13.30	23.22
M_2	0.41	1.35	3.97	7.67	17.00	27.25
M_3	0.15	0.61	2.14	4.69	12.49	22.83
Exact	0.32	1.06	3.17	6.28	14.76	25.00

The Cauchy values (Fraser, 1990) for the r^* type formula differ from improved values in Barndorff-Nielsen (1991): the first used r^* based on a mean and variance correction as in Barndorff-Nielsen (1986), the second used a new version of r^* with certain likelihood invariance properties. The change to likelihood invariance produces the improvement and is consistent with comparisons made in (Fraser, 1990, Section 6). Method $M_0 = M_1$ seems mildly preferable for two of the examples, method M_3 seems preferable for the second example, and M_0 and M_3 are roughly comparable for the third example.

Adjustment factors arise with nuisance parameters and with curvature of canonical parameters. The latter is typically not accessible to exact calculations; accordingly we choose the nuisance parameter extension and examine location models for which exact calculations are reasonably accessible.

For a first example consider the Cauchy distribution in R^3, a special case of the spherical Student distribution:

$$\pi^{-2}\{1+(y_1-\lambda_1)^2+(y_2-\lambda_2)^2+(y-\psi)^2\}^{-2};$$

see Table 5.

Table 5. Spherical Cauchy

$y-\psi$	-100.0	-30.0	-10.0	-5.0	$-\ 1.0$
M_0	0.20	0.66	1.99	3.99	18.75
M_1	0.28	0.94	2.81	5.58	23.22
M_2	0.41	1.35	3.97	7.67	27.25
M_3	0.02	0.10	0.61	1.84	17.29
Exact	0.32	1.06	3.17	6.28	25.00

As a second example consider a spherical logistic

$$c\exp(r)\{1+\exp(r)\}^{-2}$$

where $r^2=(y_1-\lambda_1)^2+(y_2-\lambda_2)^2+(y-\psi)^2$; see Table 6.

Table 6. Spherical Logistic

$y-\psi$	-8.0	-6.0	-4.0	$-\ 2.0$	$-\ 1.0$
M_0	0.15	0.83	4.13	17.09	31.09
M_1	0.14	0.78	4.02	17.24	31.36
M_2	0.17	0.96	4.73	18.74	32.53
M_3	0.14	0.77	3.94	16.91	31.05
Exact	0.10	0.60	3.32	15.68	30.11

For the third example consider a spherical beta (3,3) distribution with r on $(0,1)$,

$$c(1+r)^2(1-r)^2\ ,$$

where $r^2=(y_1-\lambda_1)^2+(y_2-\lambda_2)^2+(y-\psi)^2$; see Table 7.

Table 7. Spherical Beta (3,3)

$y-\psi$	-0.9	-0.7	-0.5	$-\ 0.3$	$-\ 0.1$
M_0	-0.032	0.64	5.87	18.76	38.65
M_1	0.017	1.16	6.94	19.86	39.12
M_2	0.020	1.22	7.08	20.01	39.18
M_3	0.010	0.90	6.13	18.86	38.66
Exact	0.019	1.21	7.06	19.98	39.17

7. DISCUSSION

In Sections 4 and 5 we discussed four methods M_0, M_1, M_2, M_3 of deriving a significance function for a scalar parameter in a general asymptotic model. All four methods require the tangent directions $V = (v_1, \ldots, v_p)$ for an approximate second order ancillary surface at the data point (Section 2).

Methods M_0 and M_3 are more direct using the likelihood ratio r_ψ and a standardized maximum likelihood departure Q; the first method then uses a Lugannani & Rice type formula while the last uses a Barndorff-Nielsen type formula based on a third order standard normal variable. Having a third order normal variable is attractive but such is equally available as $\Phi^{-1}\Phi(r_\psi, Q)$ from the initial method.

The middle two methods M_1 and M_2 use a modified profile likelihood which incorporates the logarithm of an adjustment factor; for the M_1 method only the nuisance adjustment factor is incorporated while for the M_2 method a nonnormality adjustment is also incorporated. For some related discussion see Pierce & Peters (1992).

For exponential models all four methods give excellent accuracy with perhaps some mild preference for the initial $M_0 = M_1$ over the remaining methods.

For the nonexponential models $M_0 = M_1$ and M_3 seem preferable with M_0 better for two examples and M_3 better for one example.

For the multiparameter examples with two nuisance parameters the adjusted likelihood methods M_1 and M_2 seem preferable for two of the examples, while the normalized variable method M_3 seems mildly preferable for the spherical logistic (Table 6). The logistic example has long tails among distributions that have a moment generating function and tail probabilities tend to be overestimated (Tables 3, 6).

The very short tailed distribution with spherical beta form (Table 7) produces a negative value with the initial method M_0. While this is a mathematical possibility with the approximation formula in M_0 this is the first case that we have noted. It is avoided with

methods M_2 and M_3. The possibility seems to correspond to heavy nuisance parameter effect with change in interest parameter.

With large numbers of nuisance parameters the approximations tend to degrade (Butler, Huzubazar & Booth, 1992); we do not examine this issue here.

In summary we find mild preference for the direct methods M_0 and M_3 in the absence of nuisance parameters, and some preference for the numerically more extended adjusted likelihood methods M_1, M_2 in the nuisance parameter context.

REFERENCES

Abebe, A., Cakmak, S., Cheah, P.K., Fraser, D.A.S., Kuhn, J., Reid, N. and Tapia, A. (1993). Third order asymptotic model: exponential and location type approximations. *Parisankhyan Samikkha*, to appear.

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

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

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. *J. R. Statist. Soc. B* **41**, 279-312.

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

Butler, R.W., Huzubazar, S. and Booth, J.G. (1992). Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika* **79**, 705-716.

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

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

Fraser, D.A.S. and Reid, N. (1989). From multiparameter likelihood to tail probabilities for a scalar parameter. Technical Report, Department of Statistics, University of Toronto.

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

Fraser, D.A.S. and Reid, N. (1993b). Ancillaries and third order significance. Technical Report, University of Toronto.

Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica*, to appear.

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.

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.

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71**, 233-244.

Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families, *J. R. Statist. Soc. B* **54**, 701-737.