Received: by utstat; Mon Dec 23 08:51 EST 1991 From: ida@math.toronto.edu Date: Mon, 23 Dec 91 08:52:26 EST To: dfraser@utstat

Third order asymptotic models: likelihood functions leading to accurate approximations for distribution functions

D. A. S. Fraser
Department of Mathematics and Statistics
York University
North York, Canada M3J 1P3
and
N. Reid
Department of Statistics
University of Toronto
Toronto, Canada M5S 1A1

Summary

An approximation (Lugannani and Rice, 1980) to the distribution function of the sample average obtained by saddlepoint methods is based on the moment or cumulant generating function; it has been found to be remarkably accurate. For exponential models this has been reformulated (Daniels, 1958) in terms of likelihood and provides approximations for the distribution function of the score, maximum likelihood estimate or likelihood ratio statistic. In this paper we derive a simple asymptotic correspondence between the standardized expansion of the cumulant generating function and the corresponding expansion of the standardized log density with the objective of a general correspondences between likelihood and distribution function. The multivariate version although not simple is also obtained. This is then used to establish third order $O(n^{-3/2})$ accuracy for various distribution and tail probability approximations: for a general one parameter model (Fraser, 1990); for a scalar canonical parameter in a multiparameter exponential model (Fraser, Reid, Wong, 1991); for a scalar parameter in a transformation model (alternative calculation to DiCiccio, Field, Fraser, 1990). In particular, it provides an explanation for the numerical accuracy found for the parmeterization invariant tail probability formula (Fraser, 1990). The emphasis throughout is on the accurate conversion of likelihood to significance. The correspondence also provides a basis for more general approximations and inference methods which will be presented subsequently. Numerical comparisons are given.

Some key words: Approximate distributions; Asymptotic expansion; Conditional inference; Conditional likelihood; Cumulant generating function; Exponential families; Likelihood function; Location families; Marginal likelihood; Saddlepoint approximation; Tail probability formula.

1. Introduction

Various saddlepoint approximations for density and distribution functions of the sample average $\bar{y}$ or sample sum $S = \Sigma_1^n y_i$ based on a density $f(y)$ have been obtained from the corresponding cumulant generating function $c(t)$. For statistical purposes it is usually more convenient to express these in terms of a likelihood function for an embedding exponential model

$$f(y;\theta) = f(y)\exp\{\theta' y - C(\theta)\} , \tag{1.1}$$

where interest lies in $\theta = 0$; the connection was suggested by Daniels (1958).

The saddlepoint approximation (Daniels, 1954) for the density of the sample sum S from the model (1.1) is

$$\hat{f}(S;\theta) = n\hat{f}(\bar{y};\theta) = (2\pi)^{-k/2}|j(\hat{\theta})|^{-1/2}\exp\{l(\theta) - l(\hat{\theta})\} \tag{1.2}$$

where k is the dimension of the variable, $l(\theta) = l(\theta; y)$ is the likelihood function from the sample y which is $\theta' S - nC(\theta)$, $\hat{\theta}=\hat{\theta}(y)$ is the maximum likelihood estimate, and $j(\hat{\theta})$ is the observed information given by $-\partial^2 l(\theta)/\partial\theta\partial\theta'$ at $\theta = \hat{\theta}$ which is $nc(\hat{\theta})$. It is related to the exact density by $f(S;\theta) = \hat{f}(S;\theta)(1 + R_n)$ where R_n is $O(n^{-1})$ and has its leading term recorded in Daniels (1954) for the scalar case (see (2.11) below) and in McCullagh (1984) for the vector case. As $\hat{\theta}$ is a one-one function of $S=n\bar{y}$ with $dS = |j(\hat{\theta})|d\hat{\theta}$, the density of the maximum likelihood estimate is

$$\hat{f}(\hat{\theta};\theta) = (2\pi)^{-k/2}|j(\hat{\theta})|^{1/2}\exp\{l(\theta) - l(\hat{\theta})\} . \tag{1.3}$$

For the scalar parameter case the Lugannani and Rice (1980) saddlepoint approximation for the distribution function $F(S; \theta)$ of $S = n\bar{y}$ or of $\hat{\theta}$ is given by

$$F(\hat{\theta}; \theta) = \Phi(r) + \phi(r)\{ \frac{1}{r} - \frac{1}{q} \} + O(n^{-3/2}) \tag{1.4}$$

where ϕ and Φ are the standard normal density and distribution functions and

$$r = \operatorname{sgn}(\hat{\theta} - \theta)[2\{l(\hat{\theta}) - l(\theta)\}]^{1/2} \tag{1.5}$$

$$q = q_1 = (\hat{\theta} - \theta) j^{1/2}(\hat{\theta}) \tag{1.6}$$

are the signed likelihood ratio quantity $r = r(\theta; \hat{\theta})$ and the data-standardized maximum likelihood departure $q_1 = q_1\{\theta; (\hat{\theta})\}$. The computation is easily implemented (Fraser, Reid, Wong, 1991) or an observed likelihood function tabulated on a fine grid and gives the observed significance function $p(\theta) = F(\hat{\theta}^0; \theta)$ or the distribution function $F(\hat{\theta}; \theta)$ for arbitrary θ or values for any of the usual quantities.

A density function $f(y)$ can also be embedded in a location model $f(y - \theta)$ with interest in $\theta = 0$. An appropriate distribution for $\hat{\theta}$ is obtained (Fisher, 1934) from the conditional distribution given the configuration statistic $a = (y_1 - \hat{\theta}, \ldots, y_n - \hat{\theta})$. The likelihood function $l(\theta; y)$ for the original (or the conditional) model when substituted in (1.3) gives the exact density when renormalized (Barndorff-Nielsen, ????). A modification of (1.4) with (1.5) suggested by Fraser (1990) uses the data standardized score

$$q = q_2\{\theta, (\hat{\theta})\} = \frac{\partial l(\theta; y)}{\partial \theta} \hat{j}^{-1/2} \tag{1.7}$$

in place of q_1 in (1.6); the right side of (1.4) then approximates $F(\hat{\theta}|a; \theta)$ and is accurate to $O(n^{-3/2})$ (DiCiccio, Field, Fraser, 1990).

The two tail area approximations, (1.4) and (1.5) with (1.6) or (1.7), are special cases of different invariant versions of the Lugannani and Rice formula due to Fraser (1990), Fraser and Reid (1990) and Barndorff-Nielsen (1988, 1990b). Fraser's (1990) invariant version, discussed in Section 3, uses a data dependent parameter, ϕ, that is obtained as the sample space derivative of the observed likelihood and is given by (1.4) and (1.5) with

$$q \doteq q_3\{\theta, (\hat{\theta})\}\ mark = \{\dot{l}(\hat{\theta}) - \dot{l}(\theta)\} k^{-1}(y) j^{1/2}(\hat{\theta}) \qquad (1.8)$$

$$\doteq (\hat{\phi} - \phi) j^{1/2}(\hat{\phi})$$

where $\phi \doteq \dot{l}(\theta) \doteq \partial l(\theta; y)/\partial y$, $k(y) \doteq \partial \dot{l}(\theta; y)/\partial\theta|_{\hat{\theta}}$, and $j(\hat{\phi}) \doteq j(0)$ is the observed information for the new parameter ϕ. The derivation was based on using an approximating or tangent exponential model (Fraser, 1988) and the $O(n^{-3/2})$ accuracy estblished in the technical report, Fraser and Reid (1990); the proof is recorded in Section 3. The resulting approximation seems to be more accurate than asymptotically equivalent approximations based on adjusting the mean and variance of the likelihood ratio statistic. Barndorff-Nielsen's invariant version replaces q with u defined in Section 3. The relationship between the two invariant expressions is also discussed in Section 3.

In Section 2 it is shown for a third order asymptotic context that the coefficients of the expansions for log density functions and for cumulant generating functions can be put into a simple one-to-one correspondence. This then gives a simple means for comparing approximations based on the two model types just discussed, and examining their relationship to the the appropriate likelihood functions. All the results come from Taylor series expansions, which provide an easy for handling the asymptotic accuracy of the density and distribution function approximations. Section 3 uses approximating exponential models to show that the parameterization invariant tail formula is accurate to $O(n^{-3/2})$. The multivariate version of the log density and cumulant generating function connection is examined in Section 4. Tail probability formulas in the presence of nuisance parameters are discussed in Section 5. Some numerical examples are given in Section 6

We use asymptotic methods informally, as for example in DiCiccio, Field, Fraser (1990). Thus means, variances, and moment and cumulant generating functions corresponding to a density expansion are to be interpreted asymptotically.

2. Asymptotic connection: log densities and cumulant generating functions

A large part of applied statistical inference is based on the normal distribution as an approximation to the null distribution of the maximum likelihood estimate, likelihood ratio statistic, or score statistic. The discussion in the preceding section suggests that the approach to normality can be finely calibrated using simple likelihood properties.

Asymptotic normality holds for the sample mean in both exponential and location family models and of course, others, but the usual methods of proof are different. In the exponential model case where the sample mean has a cumulant generating function, the familiar proof of the central limit theorem shows that the cumulant generating function for the standardized sample mean goes to the quadratic form $t^2/2$. In the location model case, the conditional density of the sample average $\bar{y}$, conditional on the residuals $y_1 - \bar{y}, \ldots, y_n - \bar{y}$, is also asymptotically normal but the proof uses likelihood function results to show that the log of the conditional density for the standardized sample average $\bar{y}$ goes to the quadratic form $a - \bar{y}^2/2$; this is developed in Brenner, Fraser and McDunnough (1982) and Fraser and McDunnough (1984).

Both these asymptotic expansion limits can be refined to include terms to fourth order, and we now show that there is a simple one to one correspondence between the coefficients of these 4th order expansions, neglecting terms of order $O(n^{-3/2})$. Simple proofs of the tail probability formulas (1.4) and (1.5) with (1.6) or (1.7) are then given as examples.

Consider some $O_p(n^{-1/2})$ variable whose log relative density is assumed to be $O(n)$ at each point and has a unique maximum; this can arise in conditional analysis of location or location-scale models. Then a location-scale standardization gives a variable y, with location 0 and scale 1 which is $O_p(1)$ as $n \to \infty$. The log density for y less a norming constant then has the form

$$l(y) \doteq -\frac{1}{2} y^2 + \frac{a_3}{\sqrt{n}} \frac{y^3}{6} + \frac{a_4}{n} \frac{y^4}{24} + O(n^{-3/2}) \tag{2.1}$$

where a_3 and a_4 are $O(1)$ and will be referred to as pseudo-cumulants for a nominal $n = 1$ variable. Note that the leading term in the approximation to $\exp\{l(y)\}$ is the normal density function. For several examples, see DiCiccio, Field and Fraser (1990).

The density $\exp\{l(y)\}$ can be integrated to determine the norming constant; see, for example, Hinkley (1978), DiCiccio, Field and Fraser (1990), or Example 2.3 below. The density then is

$$f(y) \doteq (2\pi)^{-1/2} \exp\left(-\frac{b}{2} - \frac{y^2}{2} + \frac{a_3}{n^{1/2}} \frac{y^3}{6} + \frac{a_4}{n} \frac{y^4}{24}\right), \tag{2.2}$$

where

$$b \doteq \frac{3a_4 + 5a_3^2}{12n} \tag{2.3}$$

and terms of order $O(n^{-3/2})$ are omitted, a pattern to be followed below.

Similarly, consider some variable whose cumulant generating function is $O(n)$ at each point other than zero; this can happen with simple convolution of independent variables as in the Central Limit Theorem context. Then a mean and variance standardization give a cumulant generating function

$$c(t) \doteq \frac{1}{2} t^2 + \frac{\alpha_3}{\sqrt{n}} \frac{t^3}{6} + \frac{\alpha_4}{n} \frac{t^4}{24} + O(n^{-3/2}) \tag{2.4}$$

where α_3 and α_4 are the standardized third and fourth cumulants for a nominal $n \doteq 1$

variable.

Simple computation described below then show that the density $f(y)$ in (2.2) or log density in (2.1) has a cumulant generating function which, after a mean and standard deviation adjustment

$$\begin{aligned} \mu &= \frac{a_3}{2n^{1/2}} , \\ \sigma &= 1 + \frac{a_4 + 2a_3^2}{4n} , \end{aligned} \tag{2.5}$$

takes the form (2.4) where

$$\alpha_3 = a_3 , \quad \alpha_4 = a_4 + 3a_3^2 ; \tag{2.6}$$

in other words, the cumulant generating function of $z = (y - \mu)/\sigma,$ is $c(t)$; the unadjusted cumulant generating function is recorded as $\tau(t)$ in (2.9) below. Conversely, if a variable z has cumulant generating function $c(t)$ given by (2.4), then the density after a location and scale adjustment

$$\begin{aligned} m &= -\frac{\alpha_3}{2n^{1/2}} , \\ s &= 1 - \frac{\alpha_4 - \alpha_3^2}{4n} \end{aligned} \tag{2.7}$$

takes the form (2.2) where

$$a_3 = \alpha_3 , \quad a_4 = \alpha_4 - 3\alpha_3^2 ; \tag{2.8}$$

in other words the density of the variable $y = (z - m)/s$ is given by (2.2).

The cumulant generating function say $\tilde{c}(t)$ of the variable y in (2.2) is given by

$$\exp\{\tilde{c}(t)\} = \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\{ -\frac{b}{2} + yt - \frac{y^2}{2} + \frac{a_3}{n^{1/2}} \frac{y^3}{6} + \frac{a_4}{n} \frac{y}{24} \} \, dy .$$

A substitution to complete the quadratic produces a new linear term from the higher power. Repetition of this then leads to the change of variable

$$y = x + t + \left(\frac{a_3}{2n^{1/2}}\, t^2 + \frac{a_4 + 3a_3^2}{6n}\, t^3 \right) + x \left(\frac{a_3}{2n^{1/2}}\, t + \frac{2a_4 + 5a_3^2}{8n}\, t^2 \right)$$

which fully recenters the exponent. The normalization implicit in (2.2) then leads to

$$\tilde{c}(t) = \frac{a_3}{2n^{1/2}}\, t + \left(1 + \frac{a_4 + 2a_3^2}{2n} \right) \frac{t^2}{2} + \frac{a_3}{6n^{1/2}}\, t^3 + \frac{(a_4 + 3a_3^2)t^4}{24n}\,, \qquad (2.9)$$

$$= c(t) + \frac{a_3}{2n^{1/2}}\, t + \frac{a_4 + 2a_3^2}{2n}\, \frac{t^2}{2}$$

justifying the correspondence from density to cumulant generating function described above. The reverse transformation can be derived in various ways or deduced from the uniqueness theorem.

As described in the Introduction many approximations can be more easily described in terms of a corresponding exponential model. From the standardized density (2.2) with cumulant generating function (2.9) we obtain the exponential model

$$f(y;\theta) = \frac{1}{(2\pi)^{1/2}} \exp\left\{ -\frac{b}{2} + l(y) + y\theta - \tilde{c}(\theta) \right\}. \qquad (2.10)$$

Example 2.1. The asymptotic exponential model (2.10) leads to a simple proof of the saddlepoint density approximation (1.2) with first correction term, which for the case $\theta = 0$ is

$$\hat{f}(y) = (2\pi)^{-1/2} \cdot \hat{j}^{-1/2} \cdot \exp\{l(0;y) - l(\hat{\theta};y)\}(1 - \delta/2) \qquad (2.11)$$

where $\delta = (5\alpha_3^2 - 3\alpha_4)/12n = -(3a_4 + 4a_3^2)/12n$, $l(0;y) = l(y)$ is the likelihood function from (2.10) and $-\delta/2$ is the leading term (Daniels, 1954) of the remainder R_n. To verify (2.11) we need expansions for the various components; by straightforward calculation we obtain

$$\hat{\theta} \doteq -\frac{a_3}{2n^{1/2}} + y\left[1 - \frac{a_4 + a_3^2}{2n}\right] - \frac{a_3}{2n^{1/2}}\, y^2 - \frac{a_4}{6n}\, y^3 \,,$$

$$j(\hat{\theta}) \doteq \exp\left\{\frac{a_3}{n^{1/2}}\, y + \frac{a_4 + a_3^2}{2n} + \frac{a_4 + a_3^2}{2n}\, y^2\right\} \tag{2.12}$$

$$l(0; y) - l(\hat{\theta}; y) \doteq -\frac{r}{2}\, r^2$$

$$\doteq -\frac{1}{2}\, y^2 + \frac{a_3}{2n^{1/2}}\, y + \frac{a_3}{6n^{1/2}}\, y^3 + \frac{1}{n}\left[-\frac{a_3^2}{8} + \frac{a_4 + a_3^2}{4}\, y^2 + \frac{a_4}{24}\, y^4\right].$$

Substitution of these in (2.11) togetehr with (2.6) then gives the density (2.2).

Example 2.2. The Lugannani and Rice formula (1.4) with (1.5) and (1.6) can be obtained by simple integration of the asymptotic model (2.10); it suffices to calculate for $\theta \doteq 0$ and we use the reexpressed density (2.11). As primary variable we use the signed likelihood ratio r implicit in (2.12):

$$r \doteq z - \frac{a_3}{6n^{1/2}}\, z^2 - \frac{3a_4 + a_3^2}{72n}\, z^3 \tag{2.13}$$

where z is the mean and variance standardized y recorded after (2.6). As secondary variable we use the data standardized maximum likelihood departure (1.6),

$$q \doteq \hat{\theta}\, \hat{j}^{1/2} \doteq \hat{\theta} + \frac{\alpha_3}{n^{1/2}}\, \frac{\hat{\theta}^2}{2} + \frac{6\alpha_4 - 3\alpha_3^2}{n}\, \frac{\hat{\theta}^3}{24} \tag{2.14}$$

$$\doteq r + \frac{\alpha_3}{n^{1/2}}\, \frac{r^2}{6} + \frac{9\alpha_4 - 13\alpha_3^2}{n}\, \frac{r^3}{72}\,.$$

The two variables are closely related, for example

$$r^{-1} - q^{-1} \doteq \frac{\alpha_3}{n^{1/2}}\, \frac{1}{6} - \delta\, \frac{r}{2} \tag{2.15}$$

with δ defined after (2.11).

From (2.12) and (2.10) with $\theta = 0$ we obtain $r^2/2 = l(\hat\theta; y) = \hat\theta y - c(\hat\theta)$ giving $rdr = \hat\theta dy$ and then $\hat{j}^{-1/2} dy = (r/q)dr$ where we have written $\hat{j} = j(\hat\theta)$. We can then integrate (2.11) obtaining

$$\begin{aligned} e^{\delta/2} F(y;0) &= \int_{-\infty}^{r} \phi(r) \frac{r}{q}\, dr \\ &= \Phi(r) + \phi(r)(r^{-1} - q^{-1}) - \int_{-\infty}^{r} \phi(r) d(\Lambda^{-1} - q^{-1}) \\ &= (1 + \delta/2)\Phi(r) + \phi(r)(r^{-1} - q^{-1}) \end{aligned} \tag{2.16}$$

where the last step uses $d(r^{-1} - q^{-1}) = (-\delta/2)dr$. Formula (1.4) then follows by noting that $r^{-1} - q^{-1} = O(n^{-1/2})$ and $\delta = O(n)$. The density $f(y)$ in (2.2) reexpressed as (2.11) can be transformed to the density of r using (2.13) showing that it is normal $(-a_3/6n^{1/2}, 1 - 2a_4 + 13a_3^2/36n)$ and that δ is given by (2.11); these are the Bartlett corrections for the signed likelihood statistic with the exponential model.

Example 2.3. The location model tail probability formula (1.4) and (1.5) with (1.7) can also be obtained from the model $f(y - \theta)$ based on the density $f(y)$ in (2.2). With this different model we have $\hat\theta = y$ and $j(\hat\theta) = \hat{j} = 1$. The signed likelihood ratio statistic r for testing $\theta = 0$ is obtained from

$$\begin{aligned} \frac{r^2}{2} &= -l(y) = \frac{y^2}{2} - \frac{a_3}{n^{1/2}} \frac{y^3}{6} - \frac{a_4}{n} \frac{y^4}{24}; \\ r &= y - \frac{a_3}{6n^{1/2}} y^2 + \frac{3a_4 + a_3^2}{72n} y^3 . \end{aligned} \tag{2.17}$$

The score statistic q in (1.7) for testing $\theta = 0$ is

$$\begin{aligned} q &= \frac{\partial}{\partial\theta}\, l(y-\theta)\big|_{\theta=0} = -\frac{\partial}{\partial y}\, l(y) = -\, l'(y)\,. \\ &= y - \frac{a_3}{2n^{1/2}}\, y^2 - \frac{a_4}{6n}\, y^3\,. \end{aligned} \tag{2.18}$$

It follows that $r^{-1} - q^{-1} = a_3/3n^{1/2} - (3a_4 + 5a_3^2)r/24n$ and thus $d(r^{-1} - q^{-1}) = -(b/2)dr$. We can then integrate

$$e^{b/2} F(y;0) = \int_{-\infty}^{r} \phi(r)\,\frac{r}{q}\,dr$$

following the same steps as in (2.16) and obtain the location tail probability formula. The density $f(y)$ in (2.2) with say b unknown can be transformed to the density for r given by (2.17) showing that r is normal $(a_3/3n^{1/2};\, 1 + (9a_4 + 11a_3^2)/36n)$ and that b is as given in (2.3); these are the Bartlett corrections for the signed likelihood ratio statistic with the location model.

The distribution function approximation (1.4) for the null density $f(y)$ in (2.2) can be obtained from the exponential embedding (Example 2.2) using the exponential statistics r_1 and q_1 given by (2.13) and (2.14), or from the location embedding (Example 2.3) using the location statistics r_2 and q_2 given by (2.17) and (2.18); both approximations are $O(n^{-3/2})$. It is appropriate then to consider the difference between two such approximations, say F_1 and F_2.

Let $r_2 = r_1 + \delta$ and $q_2 = q_1 + \varepsilon$ and expand the second version of the formula to second derivatives using $\delta = O(n^{-1/2})$, $\varepsilon = O(n^{-1/2})$:

$$\begin{aligned} \frac{F_2 - F_1}{\phi(r_1)} &= \delta\left(\frac{r_1}{q_1} - \frac{1}{r_1^2}\right) + \varepsilon\,\frac{1}{q_1^2} + \delta^2\left(\frac{1}{2q_1} - \frac{r_1^2}{2q_1} + \frac{1}{2r_1} + \frac{1}{r_1^3}\right) \\ &\quad + \varepsilon^2\left(\frac{-1}{q_1^2}\right) + \delta\varepsilon\left(-\frac{r_1}{q_1^2}\right). \end{aligned} \tag{2.20}$$

We then substitute

$$\delta \doteq d_1 n^{-1/2} + d_2 n^{-1}\,, \quad \varepsilon \doteq e_1 n^{-1/2} + e_2 n^{-1}$$

where the d's and e's are functions of a_3, a_4, and y, and obtain

$$\frac{F_2 - F_1}{\phi(r_1)} \doteq n^{-1/2} H_1 + n^{-1} H_2\,,$$

where the H's depend on r_1, q_1, y. In the case of Examples 2.2 and 2.3 substitution produces 0 to order $O(n^{-3/2})$.

In the next section we will see that the accuracy is not restricted to the two special model types but is obtained for a general embedding satisfying asymptotic properties.

The asymptotic connections among densities, likelihoods, and cumulant generating functions as developed in this section forms a basis for a general asymptotic analysis of inference to order $O(n^{-3/2})$.

3. Third order asymptotic model and the invariant tail probability formula

In this section we derive a general third order asymptotic model in an exponential-type canonical form (3.11). This then provides a simple proof of the $O(n^{-3/2})$ accuracy of the invariant tail probability formula (1.4) and (1.5) with (1.8).

The exponential and location versions of the tail area approximation as mentioned in section 1, are both special cases of the invariant tail probability formula (Fraser, 1988, 1990) involving sample space differentiation of likelihood. For a model $f(y;\theta)$, where both y and θ are scalars, the invariant tail area approximation is

$$p(\theta) \doteq F(\hat{\theta}^0;\theta) \doteq \Phi(r^0) + \phi(r^0)\left(\frac{1}{r^0} - \frac{1}{q^0}\right) \tag{3.1}$$

where r^0 is as usual the signed square root of the log likelihood ratio statistic,

$$\begin{aligned} q^0 &\doteq \{\dot{l}(\hat{\theta}^0) - \dot{l}(\theta)\} k^{-1}(x^0) j^{\frac{1}{2}}(\hat{\theta}^0) \\ &\doteq (\hat{\phi} - \phi) j(\hat{\phi}^0) \end{aligned} \tag{3.2}$$

is the standardized maximum likelihood estimate for the constructed parameter

$$\phi \;=\; \dot{l}(\theta; x^0)\,, \tag{3.3}$$

and $k(y^0) = \partial \dot{l}(\theta; y^0)/\partial\theta|_{\hat{\theta}^0}$, $\hat{\theta}^0 = \hat{\theta}(y^0)$, and with y^0 designating the observed data value; The second version in (3.2) uses the observed information for ϕ which can be obtained using the scale factor $k^{-1}(y_0)$. The observed data value is emphasized in (3.1) and (3.2) because the constructed parameter ϕ depends on it; and it is of course the value for which significance and confidence intervals are usually required; the formulas however apply for arbitrary x^0. The tail probability function $p(\theta)$ given by (3.1) is called the significance function in Fraser (1991).

As context for the third order asymptotic model consider some relative density $f(y;\theta)$ with scalar variable and parameter. We assume that for each θ the variable y is $O_p(n^{-1/2})$ about the maximum density value $\hat{y}(\theta)$, and that $l(\theta; y) = \log f(y;\theta)$ with either argument fixed is $O(n)$ and has a unique maximum. We will derive various canonical representations for this model in the neighbourhood of a data point y^0 and the corresponding parameter range of maximum likelihood values. These will be of further use in statistical theory but one of them called the canonical exponential type approximation provides a simple verification of the accuracy $O(n^{-3/2})$ of the invariant tail area formula.

The various representations are based on fourth order by Taylor expansions; thus

$$l(\theta; x) \;=\; \sum_{ij=1}^{4} A_{ij}(\theta-\theta_0)^i (y-y_0)^j / i!\, j! \tag{3.4}$$

and we will record just the 5×5 matrix of derivatives

$$A_{ij} \;=\; \frac{\partial^i}{\partial\theta^i}\,\frac{\partial^j}{\partial y^j}\, l(\theta; y)|_{(\theta_0, y_0)} \tag{3.5}$$

with rows for parameter and columns for variable.

As an initial expansion point consider (θ_0, y_0) where $y_0 \doteq \hat{y}(x_0)$ is the maximum density value for θ_0. We will make successive changes of variable and parameter, targeting on an exponential model pattern, but falling short ones for a single fourth order derivative. The composite transformation is complicated and not of interest here, only the form of the final coefficient array. For simplicity then we record the successive arrays relative to a nominal θ and nominal y. The lengthy details will be presented elsewhere.

First we standardize the variable with respect to its second derivative at the maximum giving

$$(-A_{02})^{1/2}(y - y_0)$$

and standardize the parameter with respect to the cross Hessian between θ and the new y at the reference value giving

$$(-A_{02})^{-1/2} A_{11}(\theta - \theta_0)\,.$$

The coefficient array for the new affinely transformed arguments is

$$\begin{bmatrix} a_{00} & 0 & -1 & a_{03}n^{-1/2} & a_{04}n^{-1} \\ a_{10} & 1 & a_{12}n^{-1/2} & a_{13}n^{-1} & - \\ a_{20} & a_{21}n^{-1/2} & a_{22}n^{-1} & - & - \\ a_{30}n^{-1/2} & a_{31}n^{-1} & - & - & - \\ a_{40}n^{-1} & & & & \end{bmatrix} \qquad (3.6)$$

For the higher derivative we have shown the dependence on n obtained from the asymptotic assumptions; the missing terms are $O(n^{-3/2})$.

Second we transform the variable so the new a_{12} and a_{13} are zero as with a canonical exponential model. The new variable is

$$y + a_{12}y^2/2n^{1/2} + a_{13}y^3/6n\,,$$

which gives a linear and quadratic Jacobian effect. The new matrix array has the form

$$\begin{bmatrix} a_{00} & b_1 n^{-1/2} & -(1 + b_2/n) & a_{03}n^{-1/2} & a_{04}n^{-1} \\ a_{10} & 1 & 0 & 0 & - \\ a_{20} & a_{21}n^{-1/2} & a_{22}n^{-1} & - & - \\ a_{30}n^{-1/2} & a_{31}n^{-1} & - & - & - \\ a_{40}n^{-1} & - & - & - & - \end{bmatrix} \qquad (3.7)$$

where many of the entries have changed.

Third, we recenter the variable so that the null density has maximum at zero. The new variable is

$$y - b_1 n^{-1/2}$$

giving an array; of the following form

$$\begin{bmatrix} a_{00} & 0 & -(1+bn^{-1}) & a_{03}n^{-1/2} & a_{04}n^{-1} \\ a_{10} & 1 & 0 & 0 & - \\ a_{20} & a_{21}n^{-1/2} & a_{22}n^{-1} & - & - \\ a_{30}n^{-1/2} & a_{31}n^{-1} & - & - & - \\ a_{40}n^{-1} & - & - & - & - \end{bmatrix}. \tag{3.8}$$

Fourth, we define a new parameter so that the new a_{21} and a_{31} are zero as with a canonical exponential model. The new parameter is

$$\theta + a_{21}\theta^2/2n^{1/2} + a_{31}\theta^3/6n$$

and the resulting array has the form

$$\begin{bmatrix} a_{00} & 0 & -(1+bn^{-1}) & a_{03}n^{-1/2} & a_{04}n^{-1} \\ a_{10} & 1 & 0 & 0 & - \\ a_{20} & 0 & a_{22}n^{-1} & - & - \\ a_{30}n^{-1/2} & 0 & - & - & - \\ a_{40}n^{-1} & - & - & - & - \end{bmatrix}. \tag{3.9}$$

Fifth we rescale the variable and parameter so that the null density curvature is -1. The new variable and parameter are

$$(1+bn^{-1})^{1/2}y\,, \quad (1-bn^{-1})\theta\,.$$

There is a constant Jacobian effect giving the coefficient array

$$\begin{bmatrix} a_{00} & 0 & -1 & a_3n^{-1/2} & a_4n^{-1} \\ a_{10} & 1 & 0 & 0 & - \\ a_{20} & 0 & cn^{-1} & - & - \\ a_{30}n^{-1/2} & 0 & - & - & - \\ a_{40}n^{-1} & - & - & - & - \end{bmatrix}. \tag{3.10}$$

If $c = 0$ the model is the exponential model (2.10) and the coefficients in the first column can be expressed in terms of a_3 and a_4; we use $a = -(1/2)\log(2\pi)$ and obtain in $(c = 0)$

$$\begin{bmatrix} a-(3a_4+5a_3^2)/24n & 0 & -1 & a_3/n^{1/2} & a_4/n \\ -a_3/2n^{1/2} & 1 & 0 & 0 & - \\ -\{1+(a_4+2a_3^2+c)/2n\} & 0 & c/n & - & - \\ -a_3/n^{1/2} & 0 & - & - & - \\ -(a_4+3a_3^2+6c)/n & - & - & - & - \end{bmatrix}. \tag{3.11}$$

The integral of a factor

$$\exp\{cy^2\theta^2/4n\} = 1+cy^2\theta^2/4n$$

with the exponential density needs only the $O(n^{-1/2})$ version of the model which is normal $(\theta; 1)$; thus $E(y^2) = 1+\theta^2$ and the integral is

$$\exp\{c(1+\theta^2)\theta^2/4n\} = 1+c(1+\theta^2)\theta^2/4n\,.$$

This explains the c terms in the first column of (3.11) which we call the *canonical third order asymptotic model.*

Without loss of generality we can consider the approximation of $F(0;\theta)$, for the data point $y = 0$ together with some general θ, say θ_0; we use the canonical model centered on the original data point of interest. The model at $\theta = \theta_0$ coincides with the following exponential model with parameter value $\theta_0(1+c\theta_0^2/4n)$.

$$\begin{bmatrix} a-(3a_4+5a_3^2+6c\theta_0^2)/24n & 0 & -(1-c\theta_0^2/2n) & a_3/n^{1/2} & a_4/n \\ -a_3/2n^{1/2} & (1-c\theta_0^2/4n) & 0 & 0 & - \\ -\{1+(a_4+2a_3^2)/n\} & 0 & 0 & - & - \\ -a_3/n^{1/2} & 0 & - & - & - \\ -(a_4+3a_2^2)/n & - & - & - & - \end{bmatrix}, \tag{3.12}$$

which is the density for $(1+c\theta_2^2/4n)$ times the variable for (2.10). The equivalence is easily checked by substitution.

The tail area can be evaluated by Lugannani and Rice, say Example 2.2, applied to the exponential model above, giving say $F_1(0,\theta_0)$. It can also be evaluated tentatively using the parameterization invariant formula (3.1) with (3.2) giving say F_2. These formulas have the differences

$$\delta \doteq r_2 - r_1 \doteq -c\theta_0/n\,, \quad \varepsilon \doteq q_2 - q_1 \doteq \frac{c\theta_0^3}{n} - \frac{c\theta_0}{n}\,.$$

Substitution in (2.20),

$$\frac{F_2 - F_1}{\phi(r_1)} \doteq \delta\left(\frac{r_1}{q_1} - \frac{1}{r_3^2}\right) + \varepsilon\left(\frac{1}{q_3^2}\right)$$

with $r_1 \doteq q_1 \doteq -\theta_0 + O(n^{-3/2})$ gives zero to order $O(n^{-3/2})$, thus verifying the parameterization invariant formula.

If an exponential model $f(y;\theta)$ is not presented in canonical form, the canonical parameter and variable, $\phi \doteq \phi(\theta) \doteq k^{-1}(y^0)l(\theta;y^0)$, $x = x(y) = \partial l(\theta;y)/\partial\theta|_{\hat\theta^0}$ with $c(\phi) \doteq -l(\theta(\phi))$, can be obtained from the likelihood function at y^0 giving the version $g(x;\phi) \doteq \exp\{\phi x - c(\phi)\}g(x;0)$ Fraser (1988). If the original model $f(y;\theta)$ is not of exponential family form than an observed likelihood function $l(\theta;y^0)$ and its derivative can lead to g as an approximating exponential model for a neighbourhood of the observed data point. This leads to the location model version (1.4) and (1.5) with (1.7), and more generally to the parameterization invariant version (1.4) and (1.5) with (1.8).

The parameterization invariant version is closely related to an approximation given by Barndorff-Nielsen (1988, 1990b) which uses (1.4) and (1.5) with

$$u \doteq \{\frac{\partial}{\partial\hat\theta}\, l(\theta;\hat\theta,a)|_{\theta\doteq\hat\theta} - \frac{\partial}{\partial\hat\theta}\, l(\theta;\hat\theta,a)\}|j(\hat\theta)|^{-1/2}. \qquad (3.13)$$

In the definition of u, it is assumed that the minimal sufficient statistic has been transformed to the pair $(\hat{\theta}, a)$, where $\hat{\theta}$ is the maximum likelihood estimate, and a is exactly or approximately ancillary. The left hand side of (1.4) then approximates the conditional distribution of $\hat{\theta}$, given a.

The main difference in the definitions of u and q defined in (1.8) or (3.2) is the choice of variable for differentiation on the sample space. In (3.13) it is specified to be $\hat{\theta}$, whereas in (3.2) it can be any one-dimensional sample space variable. The factor $k^{-1}\hat{j}$ needed to convert one expression to the other provides an adjustment and makes q invariant to the choice of variable for sample space differentiation. Barndorff-Nielsen (1991) indicates that resutls from Barndorff-Nielsen (1986) would give $O(n^{-3/2})$ accuracy extending the $O(n^{-1})$ accuracy in Barndorff-Nielsen (1990b).

4. Multivariate connections: log densities and cumulant generating functions

The univariate connection between log density and cumulant generating function in Section 2 is extended now to the multivariate case. This provides a basis for various inference analyses which we do not examine directly in this paper.

Consider a relative density function in p variables, and assume that it is $O_p(n^{-1/2})$ and that the log density is $O(n)$ at each point; this can arise in both conditional and marginal analyses as some sample size n becomes larger. A location scale standardization then gives a variable y where log density less a norming constant has the form

$$l(y) \doteq -\frac{1}{2} I^{ij}\, y_i y_j + \frac{1}{6} A^{ijk} y_i y_j y_k + \frac{1}{24} A^{ijkl} y_i y_j y_k y_l + O(n^{-3/2}) \qquad (4.1)$$

where the indices run from 1 to p and summation over repeated indices is implied. The standardized log-density takes its maximum at 0, and has asymptotic information matrix such that $I^{ij} \doteq 0(1)$; thus $A^{ijk} \doteq O(n^{-1/2})$, and $A^{ijkl} \doteq O(n^{-1})$, as would be the case, for example, if y were a standardized sum of independent random vectors, or the conditioned variable in a transformation model.

When (4.1) is normed by a factor e^a to have integral equal to 1, we obtain an expression as given by eq.(8) of DiCiccio, Field and Fraser (1990):

$$a \doteq (-p/2)\log(2\pi) + (1/2)\log|\, I^{ij}\,| + O(n^{-1})\ , \tag{4.2}$$

$$\doteq (-p/2)\log(2\pi) + (1/2)\log|\, I^{ij}\,| - b/2 + O(n^{-3/2})$$

where with I_{ij} as the inverse of I^{ij}

$$b \doteq \frac{1}{12}\left(3A^{ijkl}I_{ij}I_{kl} + 3A^{ijk}A^{lmn}I_{ij}I_{kl}I_{mn} + 2A^{ijk}A^{lmn}I_{il}I_{jm}I_{kn}\right) \tag{4.3}$$

and consists of scalar contractions of the arrays $A^{ijk} \cdot A^{lmn}$ and A^{ijkl} (denoted $F_{4,3}, F_{6,7}, F_{6,8}$ in DiCiccio, Field and Fraser (1990)).

The moment generating function, $\exp\{\tilde{c}(t)\}$ say, is obtained by integrating an expression whose logarithm is

$$a + y_i t^i - \frac{1}{2} I^{ij} y_i y_j + \frac{1}{6} A^{ijk} y_i y_j y_k + \frac{1}{24} A^{ijkl} y_i y_j y_k y_l\ \ . \tag{4.4}$$

As a first step in centering this quartic let $y_i \doteq x_i + I_{ij}t^j$, giving

$$a + P_0(t) + x_i \xi^i - \frac{1}{2}(I^{ij} + D^{ij})x_i x_j + \frac{1}{6}\tilde{A}^{ijk} x_i x_j x_k + \frac{1}{24}\tilde{A}^{ijkl} x_i x_j x_k x_l\ \ . \tag{4.5}$$

where

$$P_0(t) \doteq \frac{1}{2} I_{ab}t^a t^b + \frac{1}{6} A^{ijk} I_{ia} I_{jb} I_{kc} t^a t^b t^c + \frac{1}{24} A^{ijkl} I_{ia} I_{jb} I_{kc} I_{ld} t^a t^b t^c t^d\ \ ,$$

$$\xi^i \doteq \frac{1}{2} A^i_{ab} t^a t^b + \frac{1}{6} A^i_{abc} t^a t^b t^c\ \ ,$$

$$D^{ij} \doteq -A_a^{ij} t^a - \frac{1}{2} A_{ab}^{ij} t^a t^b \quad , \tag{4.6}$$

$$\tilde{A}^{ijk} \doteq A^{ijk} + A_a^{ijk} t^a \quad , \qquad \tilde{A}^{ijkl} \doteq A^{ijkl} \quad ,$$

Indices for the coefficient arrays are lowered by multiplying by the matrix I_{ij}; for exampled, $A^i_{ab} \doteq A^{ijk} I_{ia} I_{jb}$, etc.. A linear term in x is still present, but can be removed by the further recentering $x_i \doteq z_i + D_{ij} \xi^i$, giving

$$a + P(t) - \frac{1}{2} I + \tilde{D}^{ij} z_i z_j + \frac{1}{6} \tilde{A}^{ijk} z_i z_j z_k + \frac{1}{24} \tilde{A}^{ijkl} z_i z_j z_k z_l \quad , \tag{4.7}$$

where

$$P(t) \doteq P_0(t) + \frac{1}{2} D_{ij} \xi^i \xi^j \quad ,$$

$$\tilde{D}^{ij} \doteq D^{ij} - A^{ijk} D_{kl} \xi^l \doteq D^{ij} - \frac{1}{2} A^{ijk} A^{lmn} I_{kl} I_{ma} I_{nb} t^a t^b \quad . \tag{4.8}$$

The moment-generating function is then obtained as

$$\exp\{\tilde{c}(t)\} \doteq \exp\{a + P(t)\} \int \exp\left(-\frac{1}{2} \tilde{D}^{ij} z_i z_j + \frac{1}{6} \tilde{A}^{ijk} z_i z_j z_k + \frac{1}{24} \tilde{A}^{ijkl} z_i z_j z_k z_l\right) dz \ .$$

$$\doteq \exp\{a + P(t) - A(t)\}$$

where

$$A(t) \doteq (-p/2) \log(2\pi) + (1/2) \log|\tilde{D}_{ij}| + B(t)/2,$$

as in (4.2). Note however that $B(t) = b$ to order $O(n^{-3/2})$ as a consequence of the fact that the t-terms in $\tilde{A}_{ijk}$ are of higher order than the leading term. We then obtain

$$\tilde{c}(t) \doteq P(t) + \frac{1}{2} \log |\, I^{ij} \,| - \frac{1}{2} \log |\, I^{ij} + \tilde{D}^{ij} \,|$$

Using the expansion given in McCullagh (1987, p.21) for $\log |\, I + X \,|$, we obtain

$$\tilde{c}(t) \doteq \mu_a t^a + \frac{1}{2}\,\sigma_{ab} t^a t^b + \frac{1}{6}\,\alpha_{abc} t^a t^b t^c + \frac{1}{24}\,\alpha_{abcd} t^a t^b t^c t^d \qquad (4.9)$$

where

$$\mu_a \doteq \frac{1}{2}\,A^{ij}_a I_{ij} \qquad (4.10)$$

$$\sigma_{ab} \doteq I_{ab} + \frac{1}{2}\,\Delta_{ab} \qquad (4.11)$$

$$\alpha_{abc} \doteq A^{ijk} I_{ia} I_{jb} I_{kc} \quad , \qquad (4.12)$$

$$\alpha_{abcd} \doteq A^{ijkl} I_{ia} I_{jb} I_{kc} I_{ld} + 3A^i_{ab} A^l_{cd} I_{il} \qquad (4.13)$$

$$\Delta_{ab} \doteq A^{ij}_{ab} I_{ij} + A^{ij}_a A^{kl}_b I_{il} I_{jk} + A^{ijk} A^l_{ab} I_{ij} I_{kl}$$

which is the multivariate version of (2.5).

For the reverse connection, the standardized cumulant generating function

$$c(t) \doteq \frac{1}{2}\,\sigma_{ab} t^a t^b + \frac{1}{6}\,\alpha_{abc} t^a t^b t^c + \frac{1}{24}\,\alpha_{abcd} t^a t^b t^c t^d \qquad (4.14)$$

corresponding to a variable z leads to expression (4.1) for $y_i \doteq z_i - m_i$, where

$$m_i \doteq -\frac{1}{2}\,\alpha_{iab}\sigma^{ab} \qquad (4.15)$$

$$I^{ij} \doteq \sigma^{ij} + \frac{1}{2}\,\Delta^{ij} \qquad (4.16)$$

$$A^{ijk} \doteq \alpha_{abc}\sigma^{ia}\sigma^{jb}\sigma^{kc} \qquad (4.17)$$

$$A^{ijkl} \doteq \alpha_{abcd}\sigma^{ia}\sigma^{jb}\sigma^{kc}\sigma^{ld} - 3\alpha^{ij}_a \alpha^{kl}_b \sigma^{ab} \qquad (4.18)$$

$$\Delta^{ij} \doteq \Delta_{ab}\sigma^{ia}\sigma^{jb} \qquad (4.19)$$

$$\doteq (\alpha_{cdab} I^{cd} - 3\alpha^{klm}\alpha_{abm} I_{kl} + \alpha^k_{ca}\alpha^c_{kb} + \alpha_{cdk}\alpha^k_{ab} I^{cd}) I^{ia} I^{jb}$$

where σ^{ij} is the inverse matrix of σ_{ij}. In formulas (4.15), (4.17-19), σ^{ij} can be taken to be equal to I^{ij}, since Δ^{ij} is $O(n^{-1})$.

The multivariate density and corresponding cumulant generating function produce a canonical version of a third order multiparameter exponential model:

$$f(y;\theta) \doteq \exp\{l(y) + y_i\theta^i - \tilde{c}(\theta)\}$$

using (4.1) and (4.9).

In the bivariate case, assuming I_{ab} is the identity matrix, and with all indices in (4.8) running from 1 to 2, and we have

$$\mu_1 \doteq \frac{1}{2}(a_{30} + a_{12}) \ , \quad \mu_2 \doteq \frac{1}{2}(a_{03} + a_{21})$$

$$\sigma_{11} \doteq 1 + \frac{1}{2}(a_{40} + a_{22} + 2a_{30}^2 + 3a_{21}^2 + a_{12}^2 + a_{30}a_{12} + 2a_{30}a_{21} + a_{12}a_{21} + a_{03}a_{30} + a_{03}a_{21})$$

$$\sigma_{12} \doteq \frac{1}{2}(a_{31} + a_{13} + 2a_{30}a_{21} + 2a_{21}a_{12} + a_{12}a_{03} + 2a_{12}a_{21} + a_{30}a_{12} + a_{21}^2 + a_{12}^2 + a_{03}a_{21})$$

$$\doteq \sigma_{21}$$

and σ_{22} is the same as σ_{11} with the subscripts reversed. The bivariate cumulant generating function is

$$\tilde{c}(t) \doteq \mu_1 t_1 + \mu_2 t_2 + \sigma_{11}t_1^2 + 2\sigma_{12}t_1t_2 + \sigma_{22}t_2^2 +$$

$$+ \frac{1}{6}\left(a_{30}t_1^3 + 3a_{21}t_1^2t_2 + 3a_{12}t_1t_2^2 + a_{03}t_2^3\right)$$

$$+ \frac{1}{24}\Big(\{3(a_{30}^2 + a_{21}^2) + a_{40}\}t_1^4 + \{12(a_{30}a_{21} + a_{21}a_{12}) + a_{31}\}t_1^3t_2$$

$$+ \{6(a_{30}a_{12} + a_{21}a_{03}) + 12(a_{21}^2 + a_{12}^2) + a_{22}\}t_1^2t_2^2$$

5. Accurate tail probability approximation with nuisance parameters

5.1. Exponential models

For an exponential model with a single scalar parameter, the Lugannani and Rice approximation, (1.4) and (1.5) with (1.6), gives $O(n^{-3/2})$ accurate probabilities $F(\theta^0; \theta)$; often the results are good even $n = 1$. This of course gives tests of significance and confidence intervals immediately. We now show that the general case with canonical interest parameter ψ and nuisance λ can produce a one-dimensional likelihood approximation that leads by the preceding conversion to $O(n^{-3/2})$ accurate probabilities $F(\hat{\psi}^0; \psi)$.

Consider the general exponential model

$$f(x, y; \theta) \doteq \exp\{\psi y + \lambda' x - k(\psi, \lambda) + h(x, y)\} ; \tag{5.1}$$

the standard inference for ψ is based on the conditional distribution of y given x. For this, the joint log likelihood for (x, y) is exact,

$$l(\psi, \lambda; x, y) \doteq a + \psi y + \lambda' x - k(\psi, \lambda),$$

and is of course available from whatever original variables preceded the sufficient statistic (x, y). The marginal log likelihood from x alone is available from (2.11)

$$l_m(\psi, \lambda; x) \doteq l(\psi, \lambda; x, y) - l(\psi; \hat{\lambda}_\psi; x, y) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| - \frac{1}{2}\delta(x, \psi, \lambda) \tag{5.2}$$

and is accurate to order $O(n^{-3/2})$; also from Section 5.4 we have that δ is constant to the same order. It follows that the conditional likelihood from y given x is

$$l_c(\psi) \doteq l(\psi, \hat{\lambda}_\psi) + \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + O(n^{-3/2}). \tag{5.3}$$

From the connection established in Section 2 it then follows that the conversion of l_c using (1.4) and (1.5) with (1.6) gives a conditional significance function for ψ that is accurate to $O(n^{-3/2})$. This verifies the accuracy $O(n^{-3/2})$ for the sequential saddlepoint in

Fraser, Reid, Wong (1991).

As noted above, a test of a component canonical parameter is based on the conditional distribution given the score for the nuisance parameter. For this, Skovgaard (1987) obtained a Lugannani and Rice type tail probability approximation using a double saddlepoint approach (Barndorff-Nielsen and Cox, 1979); both the joint density and the marginal density are approximated using the saddlepoint formula (2.11); after integrating the ratio of these two approximate densities, a tail probability formula accurate to $O(n^{-1})$ is obtained. Fraser, Reid, and Wong (1991), as part of the development of a computer program to numerically convert a likelihood function to a significance function (Fraser, 1991) suggested the direct approximation of the likelihood for the conditional distribution. The method was called the sequential saddlepoint but the asymptotic accuracy (established above) was not directly addressed there. Further discussion of the two approaches with numerical examples can be found in Butler, Huzurbazar and Booth (1991a,b) and Pierce and Peters (1991).

5.2 Transformation models

For a location model with a single scalar parameter, the DiCiccio, Field, Fraser (1990) approximation, (1.4) and (1.5) with (1.7) gives $O(n^{-3/2})$ accurate probabilities $F(\hat{\theta}^0; \theta)$; the results are often good with $n = 1$ (Fraser, 1990). We now show that the general location case with canonical interest parameter ψ and nuisance λ can produce a one-dimensional likelihood approximation that gives by the preceding conversion $O(n^{-3/2})$ accurate probabilities $F(\hat{\psi}^0; \psi)$.

Consider the location model

$$f(y-\psi, x-\lambda) \;=\; f(t, u) \tag{5.4}$$

where x, λ and u have dimension $p-1$; this reduced form could have come from a general location model by the usual conditioning on the configuration statistic. The standard inference procedure would use the marginal density for y or for $t = y - \psi$.

An approximation to the marginal density of t is available from DiCiccio, Field, Fraser (1990) or from the results in Section 5.4. Let $\hat{u}(t)$ be the value of u that maximizes $l(t, u) \doteq \log f(t, u)$ for given t, let $\hat{j}_{uu}(t)$ be the negative Hessian of $l(t, u)$ with respect to u for given t at the maximizing point $\hat{u}(t)$, and let $b = b(t)$ be the connection term (2.3) or (4.3) for the conditional density of u given t. The marginal log density for t is then

$$l_m(t) \doteq a + l(t, \hat{u}(t)) - \frac{1}{2} \log |\hat{j}_{uu}(t)| \frac{b}{2} + O(n^{-3/2}) ; \tag{5.5}$$

also from Section 5.4 we have that $b(t)$ is constant to the same order. We then allow the usual arbitrary additive constant for log likelihood, and obtain

$$l_m(\psi) \doteq l(\psi, \hat{\lambda}_\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + O(n^{-3/2}) . \tag{5.6}$$

It then follows from Section 2 that converting the approximate likelihood l_m using (1.4) and (1.5) with (1.7) gives a marginal significance function for ψ accurate to $O(n^{-3/2})$. This agrees with DiCiccio, Field, Fraser (ibid) but is expressed in terms of likelihood, thus permitting the direct use of the numerical conversion (Fraser, Reid, Wong, 1991) mentioned in Section 1.

Many transformation models such as the regression model and not pure location models and do not immediately lead to an appropriate marginal likelihood. An example however indicates the method.

Example 5.1. The regression model $y \doteq X\beta + \sigma e$ where e has a known density $f(y)$ in R^n has $r + 1$ parameters where r is the column rank of X; the highest order regression coefficient might be of interest. The standard conditional reduction with $r \doteq 2$ say would give

$$
\begin{aligned}
y &= \psi + \sigma z_1 \\
x &= \lambda + \sigma z_2 \qquad (5.7) \\
s &= \sigma \chi
\end{aligned}
$$

where the density $f(z_1, z_2, x)$ is known; a particular case would have z_1, z_2 standard normal and χ with a chi distribution on $n-2$ degrees of freedom. The original full data likelihood function

$$
l(\psi, lambda, \sigma) = \sigma^{-3} f(\sigma^{-1}(y-\psi), \sigma^{-1}(x-\lambda), \sigma^{-1}s)
$$

coincides with the conditional likelihood. Checking for the normal case, we have profile likelihood

$$
l_p(\psi) = a - \frac{n}{2} \log \{s^2 + (x-\psi)^2\} \; ,
$$

and information matrix

$$
\begin{bmatrix} \sigma^{-2} & 0 \\ 0 & 2 \end{bmatrix}
$$

relative to λ and the *location* version $\tau = \log \sigma$ of σ. The marginal log likelihood is then

$$
\begin{aligned}
l_m(\psi) &= a - \frac{n}{2} \log \{s^2 + (x-\mu)^2\} - \frac{1}{2} \log(\hat{\sigma}_{\psi}^{-2}) \\
&= a - \frac{n-1}{2} \log \{s^2 + (x-\mu)^2\} \; . \qquad (5.8)
\end{aligned}
$$

This inverts by the numerical procedure (location version) to the Student density on the appropriate degrees of freedom, $n-2$; the general case conforms to the asymptotics of Section 2 and the preceding theory.

The general transformation model with a canonical interest parameter ψ has an initial conditioning (Fisher, 1934; Fraser, 1957, 1968) in which likelihood does not change. The test for ψ is then based on the marginal distribution of the corresponding pivotal quantity. DiCiccio, Field, Fraser (1990) used the probability density of the pivotal quantity for the full parameter to obtain a tail probability formula accurate to $O(n^{-3/2})$ for a scalar parameter of interest . The method can be viewed as a Laplace integration to eliminate the nuisance parameter effect followed by an application of the parameterization invariant tail probability formula (1.4) and (1.5) with (1.8). Fraser, Lee and Reid (1990) obtained the same formula by a method of adjusted conditional distributions; the asymptotic accuracy was not examined but a Monte Carlo procedure was developed for directly assessing accuracy. In this section we developed a likelihood procedure for the pure location case; the more general will be examined elsewhere.

5.3. Discussion

The difference between (5.3) for the exponential model and (5.6) for the translation model is partly explained by the type of nuisance parameterization. If (5.3) is reexpressed in terms of the expectation parameterization which would be approximately location, then the sign on the $\log|\hat{j}|$ term would reverse bringing it into line with (5.6). Of course the two likelihoods are converted to significance functions by using different numerical inversions and different q formulas, (1.6) for exponential and (1.7) for location.

For the exponential and location models above a key ingredient is the use of the corresponding canonical parameter. This parameter can be determined by differentiating the likelihood on the sample space, in an arbitrary direction for the exponential model and along the ancillary orbit for the location model. For more general models different versions of the canonical parameter are obtained by differentiating in different directions with different implications for statistical inference. One approach is to try to find the appropriate direction of differentiation directly, as in Fraser and Reid (1989), or Fraser (1991b). A related approach is that of finding approximate ancillary statistics on which to condition, as in Barndorff-Nielsen (1986). Also for general models Barndorff-Nielsen

(1986,1990a,b,1991) has developed a tail probability approximation based on an adjusted likelihood root r^* which is asymptotically standard normal to $O(n^{-3/2})$. This general formula can be applied to transformation models, but will give a different approximation than the DiCiccio, Field and Fraser (1990) approach. Some numerical work is discussed in Fraser (1990), Barndorff-Nielsen (1991), and DiCiccio and Martin (1991). It can also be applied to component parameters in an exponential model, where it gives an approximation different from Skovgaard's or the sequential saddlepoint, although it is asymptotically equivalent to Skovgaard's version, as shown in Jensen (1991). A choice among these approximations needs input from their numerical performance, ease of implementation, and other factors.

Our emphasis in this section has focussed on approximate tail probabilities for testing real parameters, that are obtained by the numerical inversion of likelihood, the conditional likelihood (5.3) for the exponential models and the marginal likelihood (5.6) based on a location-type parameterization for the transformation models. Results for more general models depend strongly on the results in Sections 3 and 4; these will be discussed elsewhere.

5.4 Addendum

In this section we provide an integration result needed in Sections 5.1 and 5.2.

Consider a density model with the asymptotic properties discussed in Sections 3 and 4 so that the log density $\log f(x, y; \theta)$ in standardized coordinates has second derivatives $O(1)$, third derivatives $O(n^{-1/2})$, and fourth derivatives $O(n^{-1})$ as for example in (3.6). We want an approximation for

$$g(x; \theta) = \int f(x, y; \theta) dy$$

where x, y, θ can be real or vector and y has dimension d, say.

We can integrate directly using (2.2) or the generalization (4.2)

$$g(x;\theta) = (2\pi)^{d/2} f(x, \hat{y}(x,\theta);\theta))|\hat{j}(x, \hat{y}(x,\theta);\theta|^{-1/2}(1 + b/2)(1 + O(n^{-3/2})) \qquad (5.9)$$

where $b = b(x,\theta)$ contains fourth derivatives or products of third derivatives and is $O(n^{-1})$. For example when $d = 1$, $b = (3\hat{a}_4 + 5\hat{a}_3^2)/12n$ where $a_3 n^{-1/2}$, $a_4 n^{-1}$ are the standardized third and fourth derivatives with respect to y evaluated at $\hat{y}(x,\theta)$. Then to obtain dependence on x or θ we can further expand to first derivatives in each. These next derivatives have an additional factor $n^{-1/2}$ and thus are $O(n^{-3/2})$. It follows that $b(x,\theta)$ is constant to order $O(n^{-3/2})$ and thus

$$g(x;\theta) = cf(x, \hat{y}(x,\theta);\theta)|\hat{j}(x, \hat{y}(x;\theta);\theta)|^{-1/2} + O(n^{-3/2}) \qquad (5.10)$$

which is the result needed in Section 5.2.

In a similar manner we can use the exponential integration pattern (2.11) and note that δ contains fourth derivatives or products of third derivatives and is $O(n^{-1})$. Expansion then in other variables or parameters produces only $O(n^{-3/2})$ terms: Thus δ is constant to the order $O(n^{-3/2})$, which is the result needed in (5.1).

6. Numerical Results

To illustrate the parameterization invariant version of the tail probability approximation, we constructed a single parameter model by shifting and tilting a logistic density:

$$f(x;\theta) = \frac{e^{x-\theta}}{\{1 + e^{(x-\theta)}\}^2} e^{\theta(x-\theta)-c(\theta)} \qquad (A)$$

where $c(\theta) = \log\{\pi\theta/\sin\pi\theta\}$ is the cumulant generating function for the basic logistic density. Table 1 shows the exact and approximate tail areas for $\theta = 0$ and selected values of x. The approximate values were computed using (1.5) with (?); the constructed parameter ϕ is given by $1 + \theta - e^{(x-\theta)}/\{1 + e^{(x-\theta)}\}$.

The logistic cdf can also be approximated by simply embedding it in an exponential model, and then the tail area approximation is given by the Lugannani-Rice formula (1.5) with (1.6). Since the approximation was derived for exponential models, it is expected to be more accurate in their setting. The third row of Table 1 shows the approximate tail areas obtained this way.

Table 1. Exact and approximate logistics

$p_r(X > x)\backslash$ x	1	2	2.5	3	3.5	4	5	6
exact	.2689	.1192	.0759	.0474	.0293	.0180	.0067	.0025
param. invariant	.2786	.1200	.0750	.0462	.0283	.0173	.0065	.0024
Lugannani-Rice	.2669	.1177				.0179	.0067	.0025

Although the example above is somewhat artificial, in that the exact distribution for a logistic is known, it does provide an illustration of the accuracy of the tail area approximation in the entrance case of $n = 1$.

There are several illustrations in the literature of the sometimes surprising accuracy of the exponential version of the tail area approximation. We found it to be accurate to six decimal places for a sample of size 11 from the standard exponential distribution, for extreme tail areas, and somewhat less accurate in the center of the distribution. (A large deviation approximation discussed in Lu, Leu and Peng (1990) is slightly less accurate, except in the four tails ($p < .0004$) of the distribution.) Daniels (1987) provides comparisons of the exact and approximate tail area formulas for the exponential, Poisson, binormal, ... for sample sites ???, ??? and ???. Other examples are presented in Barndorff-

Nielsen and Cox (1989, Ch. ?).

Examples illustrating the accuracy of the formulas in the multiparameter setting discussed in Section 5 also appear in several papers, including Fraser, Reid and Wong (1991), Pierce and Peters (1991), Butler, Huzurbazar and Booth (1991a,b) and Jensen (1991) for the exponential model version, DiCiccio, Field and Fraser (1990), Fraser, Lee and Reid (1990) and ? for the transformation model version.

In all the pubished examples, the tail area approximation is surprisingly accurate. Butler et al (1991a) and Pierce and Peters (1991) provide a good discussion of different methods of implementing the approximation for the exponential case, and Pierce and Peters (1991) provide a discussion of the role of continuityu corrections for discrete models.

Table 1. Exact and approximate logistics

$P_r(Y>y)\backslash^{y}$	1.650	1.660	1.680	1.690	1.800	1.920	1.960	2.050	2.2960	2.5238	2.6189
exact	.0282	.0267	.0239	.0226	.0120	.0058	.0046	.0026	.00050	.00010	.000050
Lugannani-Rice	.0282	.0267	.0239	.0226	.0120	.0058	.0046	.0026	.00050	.00010	.000050
large-dev.	.0228	.0218	.0199	.0190	.0107	.0054	.0043	.0024	.00049	.00098	.000049

References

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. (1988). contribution to discussion of Reid (1988).

Barndorff-Nielsen, O.E. (1990b). Approximate interval probabilities. *J. Royal Statist. Soc. B* **52,** 485-496.

Barndorff-Nielsen, O.E. (1990a). A note on the standardized signed log likelihood ratio. *Scand. J. Statist.* **17,** 157-160.

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. Royal Statist. Soc. B* **41,** 279-312.

Brenner, D., Fraser, D.A.S. and McDunnough, P. (1982). On asymptotic normality of likelihood and conditional analysis. *Canad. J. Statist.* **10,** 163-72.

Butler, R.W., Huzurbazar, S. and Booth, J.G. (1990). Saddlepoint approximations to the generalized variance and Wilks' statistic. *Biometrika* to appear.

Butler, R.W., Huzurbazar, S. and Booth, J.G. (1991). Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Technical Report*, Department of Statistics, Colorado State University. to appear.

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

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

Daniels, H.E. (1958). Discussion of "The regression analysis of binary sequences" by D.R. Cox. *J. Royal Statist. Soc. B* **20,** 236-238.

Daniels, H.E. (1987). Tail probability approximation. *Int. Statist. Rev.,* **55,** 37-48.

DiCiccio, T.J., Field, C.A. and Fraser, D.A.S. (1990). Approximation of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77,** 77-95.

DiCiccio, T.J. and Martin, M.A. (1990). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Technical Report No. 346*, Department of Statistics, Stanford University.

Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65,** 457-87.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Royal Soc. A* **144,** 285-307.

Fraser, D.A.S. (1957). A regression analysis using the invariance method. *Annals Math. Statist.* **28,** 517-520.

Fraser, D.A.S. (1968). *The Structure of Inference*. New York: John Wiley.

Fraser, D.A.S. (1988). Normed likelihood as saddlepoint approximation. *J. Mult. Anal.* **27,** 181-93.

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

Fraser, D.A.S. (1991a). Statistical inference: from likelihood to significance. *J. Amer. Statist. Assoc.* **86,** 258-265.

Fraser, D.A.S. (1991b). Directional analysis and statistical frames. *Statistical Papers*, to appear.

Fraser, D.A.S., Lee, H.S. and Reid, N. (1990). Nonnormal linear regression; an example of significance levels in high dimensions. *Biometrika,* **77,** 333-41.

Fraser, D.A.S. and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analysis. *Canad. J. Statist.* **12,** 183-90.

Fraser, D.A.S., and Reid, N. (1989). Adjustments to profile likelihood. *Biometrika,* **76,** 477-88.

Fraser, D.A.S., Reid, N. and Wong, A. (1991). From observed likelihood to tail area: a two pass approximation. *J. Royal Statist. Soc. B*, **52,** 483-492.

Fu, J.C., Leu, C.M. and Peng. C.Y. (1990). A numerical comparison of normal and large duration approximation for tail probabilities. *J. Japan Statist. Soc.* **20,** 61-67.

Hinkley, D.V. (1978). Likelihood inference about location and scale parameters. *Biometrika* **65,** 253-61.

Jensen, J.L. (1991). The relation of r^* to saddlepoint approximations. *Technical report #223* , Department of Theoretical Statistics, Aarhus.

Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12,** 475-90.

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71,** 233-44.

McCullagh, P. (1987). *Tensor methods in statistics*. Chapman and Hall, London.

Pierce, D.A. and Peters, D. (1991). Practical use of higher order asymptotics for multiparameter exponential families. *Technical Report*, Department of Statistics, Oregon State University.

Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3,** 213-37.

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

Wong, A. (1991). Inference on the shape parameter of a gamma distribution: a conditional approach. *Technometrics*, to appear.