Statistica Sinica 3(1993), 67-82

THIRD ORDER ASYMPTOTIC MODELS: LIKELIHOOD FUNCTIONS LEADING TO ACCURATE APPROXIMATIONS FOR DISTRIBUTION FUNCTIONS

D. A. S. Fraser and N. Reid

York University and University of Toronto

Abstract: A very accurate saddlepoint approximation formula for the distribution function of the sample average was obtained by Lugannani and Rice (1980), and reformulated for exponential models in terms of likelihood by Daniels (1958, 1987). In this paper we obtain a simple third order asymptotic correspondence between cumulant generating functions and corresponding log density functions, leading to a correspondence between likelihood functions and distribution functions; a multivariate analog is also obtained. We use this correspondence to establish the third order accuracy of the invariant tail probability formula proposed by Fraser (1990) and of the conversion of conditional likelihood for an exponential model and marginal likelihood for a location model to tail probabilities for testing scalar parameters.

Key words and phrases: Asymptotic expansion, conditional inference, conditional likelihood, cumulant generating function, exponential families, location families, marginal likelihood, saddlepoint approximation, tail probability formula.

1. Introduction

The saddlepoint approximation for the density and distribution function of the average $\bar{y}$ or sum $S = \sum_1^n y_i$ based on a sample $y_1, \cdots, y_n$ from a density $f(y)$, is based on the corresponding cumulant generating function $c(t)$. For statistical purposes the approximation is more convenient if based on the likelihood function for an embedding exponential model

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

where y and θ are $p \times 1$ vectors and 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

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

where $l(\theta) = l(\theta; y) = \theta' S - nc(\theta)$ is the likelihood function from the sample, $\hat{\theta} = \hat{\theta}(y)$ is the maximum likelihood estimate, and $j(\hat{\theta}) = nc''(\hat{\theta})$ is the observed information matrix $-\partial^2 l(\theta)/\partial\theta\partial\theta'$ at $\theta = \hat{\theta}$. The approximate density 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 and in McCullagh (1984) for the vector case. As $\hat{\theta}$ is a monotone function of $S = n\bar{y}$ with $dS = |j(\hat{\theta})|d\hat{\theta}$, the density of the maximum likelihood estimate is approximated by

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

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

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

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

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

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

are respectively the signed likelihood ratio quantity and the data-standardized maximum likelihood departure. The computation is easily implemented with an observed likelihood function tabulated on a fine grid (see Fraser, Reid and Wong (1991)).

A density function $f(y)$ can also be embedded in a location model $f(y - \theta)$, and an appropriate distribution for $\hat{\theta}$ is conditional on the configuration statistic $a = (y_1 - \hat{\theta}, \ldots, y_n - \hat{\theta})$ (Fisher (1934)). The likelihood function $l(\theta; y)$ for the original or the conditional model when substituted into (1.3) gives the exact density when renormalized (Barndorff-Nielsen (1986)). A modification of (1.4) approximates $F(\hat{\theta}|a;\theta)$ using 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 (1.6). This is accurate to $O(n^{-3/2})$ (DiCiccio, Field and Fraser (1990)).

The two tail area approximations, (1.4) with r in (1.5) and q given by (1.6) or (1.7), are special cases of different invariant versions of the Lugannani and Rice formula (1980). The invariant versions are due to Barndorff-Nielsen (1988, 1990b) and Fraser (1990). 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) with r in (1.5) and q specified as

$$q = q_3(\theta;\hat{\theta}) = \left\{\dot{l}(\hat{\theta}) - \dot{l}(\theta)\right\} k^{-1}(y) j^{1/2}(\hat{\theta}) \tag{1.8}$$

where $\dot{l}(\theta;y) = \partial l(\theta;y)/\partial y$ and $k(y) = \partial \dot{l}(\theta;y)/\partial\theta|_{\hat{\theta}}$. The derivation was based on an approximating exponential model (see Fraser (1988)) and the $O(n^{-3/2})$ accuracy established 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.

In Section 2, for a third order asymptotic model, it is shown that the coefficients in the expansions of log density functions and cumulant generating functions can be put into a simple one-to-one correspondence. This then gives a simple procedure for assessing approximations based on appropriate likelihood functions. All the results come from Taylor series expansions, which provide an easy method 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 derived in Section 4. Tail probability formulas based on likelihood in the presence of nuisance parameters are derived in Section 5. Some numerical examples are given in Section 6.

2. Asymptotic Connection: Log Densities and Cumulant Generating Functions

A large part of applied statistical inference is based on the log density approaching the quadratic form of a normal distribution. By including cubic and quartic terms the accuracy of the approximation can be substantially increased. We approach this by examining higher order terms for the log density and for the cumulant generating function and obtaining a simple correspondence.

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 (see Fraser and McDunnough (1984)). A location-scale standardization then gives a variable y, with location 0 and scale 1 which is $O_p(1)$ as $n \to \infty$ and has log density, except for the normalizing constant, of the form

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

where a_3 and a_4 are $O(1)$ and will be referred to as pseudo-cumulants for a nominal variable corresponding to $n = 1$. Note that the leading term in the approximation to $\exp\{l(y)\}$ is the standard normal density function.

The expression $\exp\{l(y)\}$ can be integrated (see Hinkley (1978)) to determine the norming constant giving

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

where

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

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 gives a cumulant generating function of the form

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

where α_3 and α_4 are the standardized third and fourth cumulants for a nominal variable corresponding to $n = 1$.

Simple computation outlined below then shows that the log density in (2.1), after a mean and standard deviation adjustment

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

has cumulant generating function (2.4) with

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

Similarly, a cumulant generating function for a variable z, say, has a log density after a location and scale adjustment,

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

given by (2.1), with

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

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

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

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

and can be simplified by 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 gives

$$\begin{aligned}\tilde{c}(t) &= \frac{a_3}{2n^{1/2}}t + \left[1 + \frac{a_4 + 2a_3^2}{2n}\right]\frac{t^2}{2} + \frac{a_3}{6n^{1/2}}t^3 + \frac{(a_4 + 3a_3^2)t^4}{24n} \\ &= c(t) + \frac{a_3}{2n^{1/2}}t + \frac{a_4 + 2a_3^2}{2n}\frac{t^2}{2},\end{aligned} \tag{2.9}$$

which verifies the correspondence between the log density and the cumulant generating function.

As mentioned in Section 1 the 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

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

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

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

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 of the remainder R_n (see Daniels (1954)). The components needed for (2.11) have asymptotic expansions

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

$$\jmath(\hat{\theta}) = \exp\left[\frac{a_3}{n^{1/2}}y + \frac{a_4 + a_3^2}{2n} + \frac{a_4 + a_3^2}{2n}y^2\right],$$

$$l(0;y) - l(\hat{\theta};y) = -\frac{1}{2}r^2 \tag{2.12}$$

$$= -\frac{1}{2}y^2 + \frac{a_3}{2n^{1/2}}y + \frac{a_3}{6n^{1/2}}y^3 + \frac{1}{n}\Big[-\frac{a_3^2}{8} + \frac{a_4 + a_3^2}{4}y^2 + \frac{a_4}{24}y^4\Big].$$

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

Example 2.2. The Lugannani and Rice (1980) formula (1.4) with (1.5) and (1.6) and $\theta = 0$, say, can be obtained by simple integration of the asymptotic model (2.10), reexpressed as (2.11). From $-r^2/2 = \tilde{c}(\theta) - y\theta$ we obtain $rdr = \hat{\theta}dy = \hat{j}^{1/2}qdy$ and thus

$$\begin{aligned}\int \phi(r)\frac{r}{q}dr &= \Phi(r) + \int (r^{-1} - q^{-1})d\phi(r)\\ &= \Phi(r) + (r^{-1} - q^{-1})\phi(r) - \int \phi(r)d(r^{-1} - q^{-1}).\end{aligned} \tag{2.13}$$

From (2.12), we obtain with z as the mean and variance standardized version of y

$$\begin{aligned}q &= r + \frac{\alpha_3}{n^{1/2}}\frac{r^2}{6} + \frac{9\alpha_4 - 13\alpha_3^2}{n}\frac{r^3}{72},\\ r &= z - \frac{a_3}{6n^{1/2}}z^2 - \frac{3a_4 + a_3^2}{72n}z^3,\end{aligned} \tag{2.14}$$

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

Thus $r^{-1} - q^{-1}$ is $O(n^{-1/2})$ and $d(r^{-1} - q^{-1}) = -\frac{\delta}{2}dr$, which yields formula (1.4) by rearranging terms. Also the density $f(y)$ can be transformed to the density of the exponential likelihood root r using (2.12), showing that r is normally distributed to $O(n^{-3/2})$ with mean $-a_3/(6n^{1/2})$ and variance $1-(2a_4-13a_3^2)/(36n)$; these are the Bartlett corrections for the exponential model (see, e.g., DiCiccio, Field and Fraser (1990)).

Example 2.3. The location model tail probability formula (1.4) and (1.5) with q given by (1.7) can also be obtained by the same integration pattern. In the model $f(y-\theta)$ based on the density $f(y)$ in (2.2), we have $\hat{\theta} = y$, $j(\hat{\theta}) = \hat{j} = 1$, and $d(r^{-1} - q^{-1}) = -(b/2)dr$. Also, the density $f(y)$ in (2.2) can be transformed to the density for the location likelihood root r, showing that r is asymptotically normal with mean and variance given by $a_3/(3n^{1/2})$ and $1 + (9a_4 + 11a_3^2)/(36n)$.

3. Third Order Asymptotic Model and the Invariant Tail Probability Formula

In this section we derive a canonical third order asymptotic model, and then establish the $O(n^{-3/2})$ accuracy of the invariant tail probability formula proposed by Fraser (1990) (see also Barndorff-Nielsen (1988)). For a model $f(y;\theta)$, where both y and θ are scalars, the approximation uses

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

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

$$\begin{aligned} q^0 &= \left\{\dot{l}(\hat{\theta}^0) - \dot{l}(\theta)\right\}k^{-1}(y^0)j^{1/2}(\hat{\theta}^0) \\ &= (\hat{\phi}^0 - \phi)j^{1/2}(\hat{\phi}^0) \end{aligned} \tag{3.2}$$

is the standardized maximum likelihood estimate for the constructed parameter

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

$k(y^0) = \partial\dot{l}(\theta;y^0)/\partial\theta|_{\hat{\theta}^0}$, $\hat{\theta}^0 = \hat{\theta}(y^0)$, and y^0 designates the observed data value. The observed data value is emphasized in (3.1) and (3.2) because the constructed parameter ϕ depends on it, and it is the value for which significance and confidence intervals are usually required; the formulas, however, apply for arbitrary y^0.

Consider some relative density $f(y;\theta)$ with scalar variable y and parameter θ and 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 now derive various expansions for this model in the neighbourhood of a parameter value θ_0, and for a Taylor's series

$$l(\theta;y) = \sum_{i,j=0}^{4} A_{ij}(\theta-\theta_0)^i(y-y_0)^j \big/ i!j!, \tag{3.4}$$

we record just the 5×5 matrix of derivatives,

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

These are initially of order $O(n)$, but various standardizations modify this. As an initial expansion point for y, let y_0 be the point at which the density $f(y;\theta_0)$ is maximized. We describe successive transformations of variable and parameter but do not record the lengthy details here.

First, we standardize the variable with respect to its second derivative at the maximum, and standardize the parameter with respect to the cross Hessian between θ and y at the reference value θ_0. The new coefficients A_{ij} are of order $O(1), O(n^{-1/2})$, $O(n^{-1})$, respectively for $i+j = 2,3,4$. Second, we transform the variable so that the new mixed derivatives A_{12} and A_{13} are zero, as in a canonical exponential model; the new variable is $y + A_{12}y^2/2 + A_{13}y^3/6$, and the Jacobian introduces linear and quadratic terms. Third, we recenter the variable so that the null density has maximum at zero; the new variable is $y - A_{01}$. Fourth, we define a new parameter so that the new mixed derivatives A_{21} and A_{31} are $O(n^{-3/2})$; the new parameter is $\theta + A_{21}\theta^2/2 + A_{31}\theta^3/6$. Finally, we rescale the variable and parameter so that the null density curvature is -1. The new variable and parameter are, respectively, $(-A_{02})^{1/2}y$ and $(-A_{02})^{-1/2}\theta$. The resulting coefficient array is

$$\begin{bmatrix} a_{00} & 0 & -1 & a_{03}n^{-1/2} & a_{04}n^{-1} \\ a_{10} & 1 & 0 & 0 & - \\ a_{20} & 0 & cn^{-1} & - & - \\ a_{30}n^{-1/2} & 0 & - & - & - \\ a_{40}n^{-1} & - & - & - & - \end{bmatrix}, \tag{3.6}$$

where we now explicitly show the asymptotic order.

If $c = 0$ the expansion is that for the canonical exponential model (2.10) and the coefficients in the first column can be expressed in terms of $a_3 = a_{03}$ and $a_4 = a_{04}$. The integral of the non-exponential term $\exp(cy^2\theta^2/4n) = 1 + cy^2\theta^2/(4n)$ is obtained by noting that to $O(n^{-1/2})$, y is $N(0,1)$ and $E(y^2) = 1 + \theta^2$ so that the integral is $1 + c(1+\theta^2)\theta^2/(4n)$. Thus, the asymptotic expansion for the density is given by the array

$$\begin{bmatrix} -(1/2)\log(2\pi) - (3a_4 + 5a_3^2)/(24n) & 0 & -1 & a_3/n^{1/2} & a_4/n \\ -a_3/(2n^{1/2}) & 1 & 0 & 0 & - \\ -\{1 + (a_4 + 2a_3^2 + c)/(2n)\} & 0 & c/n & - & - \\ -a_3/n^{1/2} & 0 & - & - & - \\ -(a_4 + 3a_3^2 + 6c)/n & - & - & - & - \end{bmatrix}. \tag{3.7}$$

We call (3.7) the *canonical third order asymptotic model.*

We can now establish the third order accuracy of the invariant tail probability formula (3.1) with (3.2) and (3.3) using the canonical model (3.7). We do this for $y = 0$ and general θ_0 without loss of generality. For the value θ_0 the density given by (3.7) coincides with the following exponential model

$$\begin{bmatrix} a-(3a_4+5a_3^2+6c\theta_0^2)/(24n) & 0 & -(1-c\theta_0^2)/(2n) & a_3/n^{1/2} & a_4/n \\ -a_3/(2n^{1/2}) & (1-c\theta_0^2)/(4n) & 0 & 0 & - \\ -\{1+(a_4+2a_3^2)/n\} & 0 & 0 & - & - \\ -a_3/n^{1/2} & 0 & - & - & - \\ -(a_4+3a_2^2)/n & - & - & - & - \end{bmatrix}, \tag{3.8}$$

with parameter ϕ, say, set equal to $\theta_0(1+c\theta_0^2)/(4n)$. For this, let $F_1(0,\theta_0)$ designate the Lugannani and Rice (Example 2.2) approximation with associated r_1 and q_1, and let $F_2(0,\theta_0)$ be the parameterization invariant expression (3.1) with associated r_2 and q_2.

Simple calculation shows that $r_2 - r_1 = -c\theta_0/n$ and $q_2 - q_1 = \dfrac{c\theta_0^3}{n} - \dfrac{c\theta_0}{n}$, from which it follows that $(F_2 - F_1)/\phi(r_1) = O(n^{-3/2})$, thus verifying the parameterization invariant formula.

The parameterization invariant version is closely related to an approximation given by Barndorff-Nielsen (1988, 1990b) which involves differentiation with respect to the maximum likelihood estimate conditional on an ancillary, and uses

$$u = \left\{ \frac{\partial}{\partial\hat\theta} l(\theta;\hat\theta,a)\Big|_{\theta=\hat\theta} - \frac{\partial}{\partial\hat\theta} l(\theta;\hat\theta,a) \right\} \left|j(\hat\theta)\right|^{-1/2}, \tag{3.9}$$

in place of the q given by (1.8). Using quite different techniques, Barndorff-Nielsen (1991) indicates that results from Barndorff-Nielsen (1986) give the $O(n^{-3/2})$ accuracy to this formula, extending the $O(n^{-1})$ accuracy in Barndorff-Nielsen (1990b). In (1986), (1990a), (1991) he also discusses the adjusted likelihood root $r^* = r + r^{-1}\log(r/u)$, which attains the corresponding accuracy as a standard normal variable.

4. Multivariate Connections: Log Densities and Cumulant Generating Functions

The univariate connection between the log density and the cumulant generating function in Section 2 extends to the multivariate case and provides a basis for various inference analyses which we do not examine here.

Consider a relative density function in p variables, and assume that it is $O(n^{-1/2})$ and that the log density is $O(n)$ at each point. We assume a location and scale standardization has been made, so that y has log density without norming constant in the form

$$l(y) = -\frac{1}{2}\Gamma^{ij}y_iy_j + \frac{1}{6}A^{ijk}y_iy_jy_k + \frac{1}{24}A^{ijkl}y_iy_jy_ky_l + O(n^{-3/2}) \tag{4.1}$$

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 $I^{ij} = O(1)$; thus $A^{ijk} = O(n^{-1/2})$, and $A^{ijkl} = O(n^{-1})$, as would be the case if y were a standardized sum of independent random vectors, or the conditioned variable in a transformation model. The normalizing constant e^a for (4.1) has

$$\begin{aligned} a &= (-p/2)\log(2\pi) + (1/2)\log|I^{ij}| + O(n^{-1}), \\ &= (-p/2)\log(2\pi) + (1/2)\log|I^{ij}| - b/2 + O(n^{-3/2}) \end{aligned} \tag{4.2}$$

where

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

and consists of scalar contractions of the arrays $A^{ijk} \cdot A^{lmn}$ and A^{ijkl} contracted by I_{ij}, the inverse of I^{ij}. These scalars are 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

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

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

$$a + P_0(t) + x_i\xi^i - \frac{1}{2}(I^{ij} + D^{ij})x_ix_j + \frac{1}{6}\tilde{A}^{ijk}x_ix_jx_k + \frac{1}{24}\tilde{A}^{ijkl}x_ix_jx_kx_l, \tag{4.5}$$

where

$$\begin{aligned} &P_0(t) = \frac{1}{2}I_{ab}t^at^b + \frac{1}{6}A^{ijk}I_{ia}I_{jb}I_{kc}t^at^bt^c + \frac{1}{24}A^{ijkl}I_{ia}I_{jb}I_{kc}I_{ld}t^at^bt^ct^d, \\ &\xi^i = \frac{1}{2}A^i_{ab}t^at^b + \frac{1}{6}A^i_{abc}t^at^bt^c, \quad D^{ij} = -A^{ij}_a t^a - \frac{1}{2}A^{ij}_{ab}t^at^b, \\ &\tilde{A}^{ijk} = A^{ijk} + A^{ijk}_a t^a, \quad \tilde{A}^{ijkl} = A^{ijkl}. \end{aligned} \tag{4.6}$$

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

$$a + P(t) - \frac{1}{2}(I^{ij} + \tilde{D}^{ij})z_iz_j + \frac{1}{6}\tilde{A}^{ijk}z_iz_jz_k + \frac{1}{24}\tilde{A}^{ijkl}z_iz_jz_kz_l, \tag{4.7}$$

where

$$
\begin{aligned}
P(t) &= P_0(t) + \frac{1}{2} D_{ij} \xi^i \xi^j, \\
\tilde{D}^{ij} &= D^{ij} - A^{ijk} D_{kl} \xi^l = D^{ij} - \frac{1}{2} A^{ijk} A^{lmn} I_{kl} I_{ma} I_{nb} t^a t^b.
\end{aligned} \tag{4.8}
$$

The moment-generating function is then obtained as

$$
\exp\{\tilde{c}(t)\} = \exp\{a + P(t) - A(t)\}
$$

where $\exp\{-A(t)\}$ is given by

$$
\begin{aligned}
& \int \exp\Big\{ -\frac{1}{2}(I^{ij} + \tilde{D}^{ij}) z_i z_j + \frac{1}{6} \tilde{A}^{ijk} z_i z_j z_k + \frac{1}{24} \tilde{A}^{ijkl} z_i z_j z_k z_l \Big\} dz \\
= \;& \exp\Big\{ - \Big[(-p/2) \log(2\pi) + (1/2) \log |\tilde{D}_{ij}| + B(t)/2 \Big] \Big\},
\end{aligned}
$$

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

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

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

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

where

$$
\mu_a = \frac{1}{2} A_a^{ij} I_{ij}, \qquad \sigma_{ab} = I_{ab} + \frac{1}{2} \Delta_{ab} \tag{4.10}
$$

$$
\alpha_{abc} = A^{ijk} I_{ia} I_{jb} I_{kc}, \qquad \alpha_{abcd} = A^{ijkl} I_{ia} I_{jb} I_{kc} I_{ld} + 3 A_{ab}^i A_{cd}^l I_{il} \tag{4.11}
$$

which gives the multivariate version of (2.5) and (2.6) using

$$
\Delta_{ab} = A_{ab}^{ij} I_{ij} + A_a^{ij} A_b^{kl} I_{il} I_{jk} + A^{ijk} A_{ab}^l I_{ij} I_{kl}.
$$

For the reverse connection, the standardized cumulant generating function

$$
c(t) = \frac{1}{2} \sigma_{ab} t^a t^b + \frac{1}{6} \alpha_{abc} t^a t^b t^c + \frac{1}{24} \alpha_{abcd} t^a t^b t^c t^d \tag{4.12}
$$

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

$$
m_i = -\frac{1}{2} \alpha_{iab} \sigma^{ab}, \qquad I^{ij} = \sigma^{ij} + \frac{1}{2} \Delta^{ij} \tag{4.13}
$$

$$
A^{ijk} = \alpha_{abc} \sigma^{ia} \sigma^{jb} \sigma^{kc}, \qquad A^{ijkl} = \alpha_{abcd} \sigma^{ia} \sigma^{jb} \sigma^{kc} \sigma^{ld} - 3 \alpha_a^{ij} \alpha_b^{kl} \sigma^{ab} \tag{4.14}
$$

using σ^{ij} as the inverse of σ_{ij} and

$$\begin{aligned}\Delta^{ij} &= \Delta_{ab}\sigma^{ia}\sigma^{jb} \\ &= \left(\alpha_{cdab}I^{cd} - 3\alpha^{klm}\alpha_{abm}I_{kl} + \alpha^{k}_{ca}\alpha^{c}_{kb} + \alpha_{cdk}\alpha^{k}_{ab}I^{cd}\right)I^{ia}I^{jb}. \quad (4.15)\end{aligned}$$

In these formulas, σ^{ij} can be taken equal to I^{ij}, since Δ^{ij} is $O(n^{-1})$.

5. Accurate Tail Probability Approximation with Nuisance Parameters

5.1. Exponential models

For the general exponential model with scalars ψ and y,

$$f(x,y;\theta) = \exp\left\{\psi y + \lambda' x - k(\psi,\lambda) + h(x,y)\right\}, \quad (5.1)$$

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

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

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),

$$l_m(\psi,\lambda;x) = l(\psi,\lambda;x,y) - l(\psi;\hat\lambda_\psi;x,y) - \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi,\hat\lambda_\psi)\right| - \frac{1}{2}\delta(x;\psi,\lambda) \quad (5.2)$$

and is accurate to order $O(n^{-3/2})$. The term δ is of order $O(n^{-1})$ but is constant to order $O(n^{-3/2})$; thus the conditional likelihood from y given x is

$$l_c(\psi) = l(\psi,\hat\lambda_\psi) + \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi,\hat\lambda_\psi)\right| + O(n^{-3/2}). \quad (5.3)$$

From the connection established in Section 2 it then follows that the conversion of l_c using (1.4) and (1.5) with q given by (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 proposed by Fraser, Reid and Wong (1991) as part of the development of a computer program to numerically convert a likelihood function to a significance function.

In contrast, Skovgaard (1987) obtained a Lugannani and Rice (1980) type tail probability approximation using a double saddlepoint approach (see Barndorff-Nielsen and Cox (1979)). Further discussion of the two approaches with numerical examples can be found in Butler, Huzurbazar and Booth (1992a,b) and Pierce and Peters (1992).

5.2. Transformation models

Consider the location model

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

where y and ψ have dimension 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 and Fraser (1990): let $\hat{u}(t)$ be the value of u that maximizes $l(t,u) = \log f(t,u)$ for given t, let $\hat{\jmath}_{uu}(t)$ 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 correction term (2.3) or (4.3) for the conditional density of u given t. The marginal log density for t is then

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

and it follows easily that $b(t)$ is constant to $O(n^{-3/2})$. This gives the marginal likelihood for ψ,

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

and then, using (1.4) and (1.5) with q given in (1.7), gives a marginal significance function for ψ accurate to $O(n^{-3/2})$. This corresponds to DiCiccio, Field and Fraser (1990) but is expressed in terms of likelihood, thus permitting the direct use of the numerical conversion of Fraser, Reid and Wong (1991).

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 $\log|\hat{\jmath}|$ term would reverse, thus bringing it into line with (5.6). Of course, the two likelihoods are converted to significance functions by using different numerical inversions.

For more general models, versions of the canonical parameter are obtained by differentiating the likelihood in various directions in the sample space. One proposed technique is to try to find an appropriate direction of differentiation directly (Fraser and Reid (1989), or Fraser (1991)). A related technique is to find approximate ancillary statistics on which to condition (see Barndorff-Nielsen (1986, 1991)).

6. Numerical Results

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

$$f(x;\theta) = \frac{e^{x-\theta}}{\{1+e^{(x-\theta)}\}^2} e^{\theta(x-\theta)-c(\theta)}, \tag{6.1}$$

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 q given in (1.8); the constructed parameter ϕ is given by $1 + \theta - e^{(x-\theta)}/\{1 + e^{(x-\theta)}\}$.

Table 1. Exact and approximate values of $\mathrm{pr}(X > x)$ for the logistic

x	2	3	4	5	6	
exact	.2689	.1192	.0474	.0180	.0067	.0025
param. invariant	.2786	.1200	.0462	.0173	.0065	.0024
Lugannani-Rice	.2669	.1177	.0466	.0179	.0067	.0025

The logistic distribution function can also be approximated by the Lugannani and Rice (1980) formula with q given in (1.6); and with the saddlepoint connection this should be more accurate. The final row of Table 1 shows the tail areas by this approximation.

We find the accuracy of the general model version to be remarkably good. The surprising accuracy of the more specialized exponential version has been discussed in the literature. 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 derived by Fu, Leu and Peng (1990) is slightly less accurate, except in the far tails ($P < .0004$) of that distribution. Fu et al's approach has the advantage of providing uniform asymptotic error bounds, which is not available using our techniques.

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 (1992), Butler, Huzurbazar and Booth (1992a,b) and Jensen (1992) for the exponential model version, DiCiccio, Field and Fraser (1990), and Fraser, Lee and Reid (1990) for the transformation model version.

In all the published examples, the tail area approximation is surprisingly accurate. Butler et al. (1992a) and Pierce and Peters (1992) provide a good discussion of different methods of implementing the approximation for the exponential case, and Pierce and Peters (1992) provide a discussion of the role of continuity corrections for discrete models.

References

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Barndorff-Nielsen, O. E. (1988). Contribution to discussion of Reid (1988).

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

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

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

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

Butler, R. W., Huzurbazar, S. and Booth, J. G. (1992a). Saddlepoint approximations to the generalized variance and Wilks' statistic. *Biometrika* **79**, 157–170.

Butler, R. W., Huzurbazar, S. and Booth, J. G. (1992b). Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika*, to appear.

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

Daniels, H. E. (1958). Discussion of "The regression analysis of binary sequences" by Dr. Cox. *J. Roy. Statist. Soc. Ser.B* **20**, 236–238.

Daniels, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55**, 37–48.

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

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

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

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

Fraser, D. A. S. (1991). 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**, 333–341.

Fraser, D. A. S. and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analyses. *Canad. J. Statist.* **12**, 183–190.

Fraser, D. A. S. and Reid, N. (1989). Adjustments to profile likelihood. *Biometrika* **76**, 477–488.

Fraser, D. A. S. and Reid, N. (1990). From multiparameter likelihood to tail probabilities for a scalar parameter. Technical Report, University of Toronto, Department of Statistics.

Fraser, D. A. S., Reid, N. and Wong, A. (1991). Exponential linear models: a two-pass procedure for saddlepoint approximation. *J. Roy. Statist. Soc. Ser.B* **53**, 483–492.

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

Hinkley, D. V. (1978). Likelihood inference about location and scale parameters. *Biometrika* **65**, 253–261.

Jensen, J. L. (1992). The modified signed likelihood statistic and saddle point approximation. *Biometrika* **79**, to appear.

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

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

McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, London.

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

Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3**, 213–238.

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

Department of Mathematics and Statistics, York University, North York, Canada M3J 1P3.
Department of Statistics, University of Toronto, Toronto, Canada M5S 1A1.

(Received July 1991; accepted September 1992)