COMPUTATION OF DISTRIBUTION FUNCTIONS FROM LIKELIHOOD INFORMATION NEAR OBSERVED DATA

This annotated copy includes typographical corrections relative to the printed JSPI version

by

D.F. Andrews and D.A.S. Fraser

Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3

and

A.C.M. Wong

SASIT, York University, Toronto, Canada M3J 1P3

SUMMARY

Likelihood is widely used in statistical applications, both for the full parameter by obvious direct calculation and for component interest parameters by recent asymptotic theory. Often, however, we want more detailed information concerning an inference procedure, information such as say the distribution function of a measure of departure which would then permit power calculations or a detailed display of p-values for a range of parameter values. We investigate how such distribution function approximations can be obtained from minimal information concerning the likelihood function, a minimum that is often available in many applications. The resulting expressions clearly indicate the source of the various ingredients from likelihood, and they also provide a basis for understanding how nonnormality of the likelihood function affects related p-values. Moreover they provide the basis for removing a computational singularity that arises near the maximum likelihood value when recently developed significance function formulas are used.

MSC: 62F25

Keywords: Saddlepoint approximation; Tangent exponential model; Tangent location model; Taylor series

1. INTRODUCTION

Central limit theorem procedures are useful in both theoretical and applied statistical inference; such procedures are derived from sample space averages, the corresponding mean and variance. In this paper we examine asymptotic inference procedures that are based on simple characteristics of the density function at or near a data point of interest. Example are then used to compare these procedures with some familiar procedures that make use of more comprehensive information such as the saddlepoint approximation and the more recent approximations from likelihood theory, for this the Edgeworth would be considered as a special case of the saddlepoint approximation.

The proposed procedures use only model derivatives at a data point and its corresponding maximum likelihood value, while the more familiar procedures would require a full cumulant generating function or a full likelihood function at other data points of interest. With recent statistical inference methods that use approximate conditioning, the likelihood functions at related data points with the same conditioning are typically not accessible; thus the large opportunity for the proposed methods. The methods also require just a general assumption of asymptotic properties and thus are applicable with models of more general form than the usual exponential type.

In Section 2 we record a brief overview of background results concerning asymptotic expansions of the logarithm of a statistical model in terms of both the variable and the parameter. In Section 3 we develop explicit formulas (3.3), (3.4) and (3.12), (3.13) for the distribution function of a scalar variable with scalar parameter; these are based on the likelihood function and the sample space first derivative of the likelihood function at some relevant data point. They have third order properties in the moderate deviations range. These formulas provide a basis for understanding how nonnormality of the likelihood function affects related p-values. They also provide the basis for removing a computational singularity that arises with recent significance function expressions near the maximum likelihood value, initially noted in Daniels (1987). In Section 4 we assess the performance of these distribution function formulas for two extreme examples, extreme in the sense of small sample size and nonnormality of the model, and we find that high accuracy is available for these extreme cases even with the very minimal but specialized information. Some brief concluding discussion is recorded in Section 5. In the remainder of this section we illustrate the methodology by summarizing simple results for the expansion and integration of a scalar log density without a parameter.

Consider a density function for a scalar y and suppose that it has asymptotic properties as some parameter n goes to infinity; for example n might be the size of some antecedent sample and the model of interest might be a one dimensional continuous model obtained be conditioning. Let $z = (y - \hat{y})\hat{\jmath}^{1/2}$ be a standardized variable calculated relative to the maximum density point $\hat{y}$ in units derived from the second derivative $\hat{\jmath}$ of the log density at the maximum. The resulting standardized density $g(z)$ then has the form

$$g(z) = \exp\left\{-\frac{5a_3^2 + 3a_4}{24n}\right\}\phi(z)\exp\left\{a_3\frac{z^3}{6n^{1/2}} + a_4\frac{z^4}{24n}\right\} \qquad (1.1)$$

to order $O(n^{-3/2})$. For this the asymptotic assumption gives that the log-density grows like $O(n)$ and has a unique maximum; the centering and scaling then gives the standard normal density component $\phi(z)$ with third and fourth order terms that drop off in powers of $n^{1/2}$ giving standardized third and fourth derivatives at the mode, and the order of the coefficients $a_3/n^{1/2}$ and a_4/n is indicated explicitly by the powers of n. For some details and background see DiCiccio *et al* (1990) and the references therein. We note that the cubic and quartic terms drop off as $n^{-1/2}$ and n^{-1} as a consequence of the asymptotic assumption, a phenomenon found for moment generating functions in the proof of the Central Limit Theorem and for log-density functions in Edgeworth and saddlepoint derivations. The asymptotic properties also show that higher order terms continue to drop off as powers of $n^{-1/2}$ and we work from the assumption that the first term not recorded gives the magnitude for the full missing terms; we do not here pursue formal proofs for this error analysis.

The distribution function $G(z)$ corresponding to the density (1.1) and calculated ro order $O(n^{-3/2})$ is

$$\begin{aligned} G(z) &= \Phi(z) - \phi(z)\left\{\frac{a_3}{6n^{1/2}}(2+z^2) + \frac{a_4}{24n}(3z+z^3) + \frac{a_3^2}{72n}(15z+5z^3+z^5)\right\} \\ &= \Phi\left(z - \frac{a_3}{6n^{1/2}}(2+z^2) - \frac{a_4}{24n}(3z+z^3) - \frac{a_3^2}{72n}(11z+z^3)\right) \end{aligned} \tag{1.2}$$

where Φ is the standard normal distribution function; this can be obtained by integrating (1.1) by parts or by using the relations

$$\begin{aligned} \int_{-\infty}^{z} x\phi(x-\theta)dx &= -\phi(z-\theta) + \theta\Phi(z-\theta) \\ \int_{-\infty}^{z} x^2\phi(x-\theta)dx &= -(z+\theta)\phi(z-\theta) + (\theta^2+1)\Phi(z-\theta) \\ \int_{-\infty}^{z} x^3\phi(x-\theta)dx &= -(2+z^2+z\theta+\theta^2)\phi(z-\theta) + (3\theta+\theta^3)\Phi(z-\theta) \\ \int_{-\infty}^{z} x^4\phi(x-\theta)dx &= -(3z+z^3+5\theta+z^2\theta+z\theta^2+\theta^3)\phi(z-\theta) + (3+6\theta^2+\theta^4)\Phi(z-\theta)\ . \end{aligned} \tag{1.3}$$

Note that the two versions of (1.2) are interconnected using $\Phi'(z) = \phi(z)$, $\Phi''(z) = -z\phi(z)$. The details are easily checked by hand or computer algebra.

2. BACKGROUND: ASYMPTOTIC STATISTICAL MODELS

Consider a statistical model $f(y;\theta)$ with scalar y and θ and with asymptotic properties as some background parameter n becomes large so that y for given θ is $O_p(n^{-1/2})$ about a maximum density point and $\ell(\theta;y) = \log f(y;\theta)$ is $O(n)$ with a unique maximum when either argument is fixed. As in the preceding section we are assuming that the asymptotic properties come from some antecedent sample size n and that the model otherwise is fully general; exponential models and location models can have these asymptotic properties but the general third order model form is a blend taking the model form between and beyond the exponential and location versions using a structuring parameter γ to be defined below. Our aim is to obtain the distribution function for this statistical model using likelihood information in the neighbourhood of an observed data point y^0.

In this section, we summarize needed asymptotic results for this model; for more details see Cakmak *et al* (1998). The results can be used widely, typically for models obtained by conditioning.

We expand the log density as a Taylor series to fourth degree about some convenient point. As our interest lies typically in probabilities calculated at or near an observed data point, we take the expansion point to be (y^0, θ^0) where $\theta^0 = \hat{\theta}^0$ is the maximum likelihood value corresponding to y^0, and obtain

$$\log f(y;\theta) = \ell(\theta;y) = \sum_{i,j\geq 0} a_{ij}(\theta-\theta^0)^i(y-y^0)^j/i!j! \tag{2.1}$$

where $a_{ij} = (\partial/\partial\theta)^i(\partial/\partial y)^j \ell(\theta;y)\big|_{(y^0,\hat{\theta}^0)}$ and $a_{10} = 0$ from the maximum likelihood property at $\theta^0 = \hat{\theta}^0$; the coefficients inherit the $O(n)$ property of the log-density itself.

To examine this in the moderate deviations range we first standardize the parameter θ relative to $\hat{\theta}^0$ in units obtained from the observed information $\hat{\jmath} = -a_{20}$,

$$\bar{\theta} = (\theta - \hat{\theta}^0)\hat{\jmath}^{1/2} \ , \tag{2.2}$$

and standardize the variable with respect to the observed y^0 in units that give a unit cross

Hessian between y change and θ change

$$\bar{y} = (y - y^0) a_{11} \hat{\jmath}^{-1/2} \ . \tag{2.3}$$

The log density of the transformed variable $\bar{y}$ with the new parameter $\bar{\theta}$ has the form

$$\log f(\bar{y}; \bar{\theta}) = \sum \bar{a}_{ij} \bar{\theta}^i \bar{y}^j / i! j! \tag{2.4}$$

where the new coefficients $\bar{a}_{ij}$ are $O(n^{-(i+j)/2+1})$ for $i + j \geq 2$. Note that the calculations are to order $O(n^{-3/2})$ and thus coefficients with $i + j > 4$ can be ignored. Also for display purposes we record just the matrix array $(\bar{a}_{ij})$ of coefficients,

$$\bar{A} = \begin{pmatrix} \bar{a}_{00} & \bar{a}_{01} & \bar{a}_{02} & \bar{a}_{03} & \bar{a}_{04} \\ 0 & 1 & \bar{a}_{12} & \bar{a}_{13} & - \\ -1 & \bar{a}_{21} & \bar{a}_{22} & - & - \\ \bar{a}_{30} & \bar{a}_{31} & - & - & - \\ \bar{a}_{40} & - & - & - & - \end{pmatrix} , \tag{2.5}$$

where omitted entries are $O(n^{-3/2})$.

It is possible to monotonely transform the variable and monotonely transform the parameter so that the model appears in some convenient form; the exponential and the location forms have useful and understandable structure. For an exponential model in canonical form we have log $f(x; \theta) = \varphi x - k(\theta) + h(y)$. Accordingly if we transform the variable and transform the parameter to target on the exponential model pattern we would would require that $a_{1j} = 0, a_{i1} = 0$ for $i, j > 1$; this and the norming property of the density then gives

Correct & different from published version

$$\bar{A} = \begin{pmatrix} a + \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} & -\frac{\alpha_3}{2n^{1/2}} & -\{1 + \frac{\alpha_4 - 2\alpha_3^2 - 5\gamma}{2n}\} & \frac{\alpha_3}{n^{1/2}} & \frac{\alpha_4 - 3\alpha_3^2 - 6\gamma}{n} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \gamma/n & - & - \\ \frac{-\alpha_3}{n^{1/2}} & 0 & - & - & - \\ \frac{-\alpha_4}{n} & - & - & - & - \end{pmatrix} \tag{2.6}$$

where $a = -\frac{1}{2}\log(2\pi)$, and $\alpha_3/n^{1/2}$, α_4/n, γ/n are the standardized third and fourth cumulants and a measure of nonexponentiality respectively; the asymptotic assumptions

Collected terms in γ are → $+ \frac{\gamma}{4n}\{y^2\theta^2 - y^4 + 5y^2 - 2\}$

imply that various terms drop off as powers of n and this is indicated explicitly. The α parameters are available from the likelihood and its gradient but the γ parameter requires a second derivative on the sample space, and thus is not available from such. We do note that the α_3, α_4, and γ are different from the versions obtained in the exponential type expansion given by (2.6).

If γ is set equal to zero we obtain an exponential model, called the tangent exponential model (Fraser & Reid, 1993). It can be presented generally as

$$f(x;\theta) = e^{k/n} \exp\{\ell(\theta) - \ell(\hat{\theta}) + \varphi(\theta)x\} \hat{\jmath}_{\varphi\varphi}^{-1/2} \tag{2.7}$$

where $\varphi(\theta)$ is taken as a new parameterization and is given as the centered likelihood gradient at y^0 taken with respect to $\bar{y}$

$$\varphi = \frac{\partial}{\partial \bar{y}} \{\ell(\theta;\bar{y}) - \ell(\hat{\theta};\bar{y})\}\big|_{\bar{y}=0}$$

and $x(y)$ is the corresponding score variable at $\hat{\theta}^0$ with respect to $\bar{\theta}$,

$$x = \frac{\partial}{\partial \bar{\theta}} \ell(\bar{\theta};\bar{y})\big|_{\bar{\theta}=0};$$

hence in the moderate deviations range that φ and θ are one-one equivalent and x and y are one-one equivalent. The tangent exponential model is a tangent generalization of the p^* formula of Barndorff-Nielsen (1983).

By contrast for a location model in canonical form we have $\log f(x;\beta) = h(x-\beta)$. Accordingly if we transform the variable and transform the parameter to target on location model form we would then require that $a_{30} = -a_{21} = a_{12}$ and $a_{40} = -a_{31} = -a_{13}$; this and the norming property then give

Corrections re JSPI version

$$\bar{A} = \begin{pmatrix} a + \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} & 0 & -1 + \frac{5\gamma}{2n} & \frac{\alpha_3}{n^{1/2}} & \frac{-\alpha_4 - 6\gamma}{n} \\ 0 & 1 & \frac{-\alpha_3}{n^{1/2}} & \frac{\alpha_4}{n} & - \\ -1 & \frac{\alpha_3}{n^{1/2}} & \frac{-\alpha_4+\gamma}{n} & - & - \\ \frac{-\alpha_3}{n^{1/2}} & \frac{\alpha_4}{n} & - & - & - \\ \frac{-\alpha_4}{n} & - & - & - & - \end{pmatrix}, \tag{2.8}$$

where $\alpha_3/n^{1/2}$, α_4/n, and γ/n are the third, fourth standardized cumulants and a measure of nonlocationality. If the present γ is set equal to zero we obtain a location model $f\{x-\beta(\theta)\}$ called the tangent location model with

$$\beta(\theta) = \int_0^\theta -\frac{\ell_\theta(\theta)}{\varphi(\theta)} d\theta \ . \tag{2.9}$$

Thus the exponential parameterization $\varphi(\theta)$, the score parameterization $\ell_\theta(\theta)$ and the location parameterization $\beta(\theta)$ are intimately connected. Cakmak *et al* (1998) examine in detail the connection between the tangent exponential model and the tangent location model.

3. INTEGRATING AN ASYMPTOTIC DENSITY

We work with an asymptotic statistical model $f(y;\theta)$ as described in the preceding section. And we suppose that the only information available is given by the local shape of the likelihood and of its sample space gradient at the data value y^0. We develop expressions for the corresponding distribution function $F(y;\theta)$ for y values near the data value y^0, and of course for $y=y^0$ which gives the observed p-value function $p^0(\theta)$.

3.1 Distribution function for the exponential type reexpression

The exponential type reexpression has Taylor coefficients recorded in the array (2.6): $\alpha_3/n^{1/2}$ and α_4/n are available from the first four derivatives of observed likelihood and the first three derivatives of the observed likelihood gradient; the γ/n however requires a second derivative with respect to the variable. The other coefficients in the array are then determined by the norming of the original model. First we record the expression for $f(x;\varphi)$ based on the Taylor series for $\log f(x;\varphi)$ with coefficients from (2.6):

$$\begin{aligned} f(x,\varphi) = \phi(x-\varphi)\exp\Big\{ & \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} - \frac{\alpha_3\varphi^3}{6n^{1/2}} - \frac{\alpha_4\varphi^4}{24n} \\ & - \frac{\alpha_3 x}{2n^{1/2}} - \frac{(\alpha_4 - 2\alpha_3^2 - 5\gamma)x^2}{4n} + \frac{\gamma\varphi^2x^2}{4n} + \frac{\alpha_3x^3}{6n^{1/2}} + \frac{(\alpha_4 - 3\alpha_3^2 - 6\gamma)x^4}{24n}\Big\} \end{aligned} \tag{3.1}$$

to order $O(n^{-3/2})$. The exponential factor in (3.1) can be expanded by computer algebra giving

$$f(x;\varphi) = \phi(x-\varphi)\Big\{1 + \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} - \frac{\alpha_3\varphi^3}{6n^{1/2}} - \frac{\alpha_4\varphi^4}{24n} + \frac{\alpha_3^2\varphi^6}{72n}\Big\} \Big\{1 - \frac{\alpha_3 x}{2n^{1/2}} - \frac{(2\alpha_4 - 5\alpha_3^2 - 10\gamma - 2\gamma\varphi^2)x^2}{8n} + \frac{\alpha_3 x^3}{6n^{1/2}} + \frac{(\alpha_4 - 5\alpha_3^2 - 6\gamma)x^4}{24n} + \frac{\alpha_3^2 x^6}{72n}\Big\}. \tag{3.2}$$

To obtain the distribution function we now integrate in a large but bounded region and then let n go to ∞; for background theory concerning the general methodology see Fraser & McDunnough (1984). We integrate term-by-term using the equalities given in (1.3) and obtain the distribution function

$$F(x;\varphi) = \Phi(x-\varphi) + \phi(x-\varphi)\left(\begin{array}{l} \frac{\alpha_3}{6n^{1/2}}\{1-\varphi^2-\varphi x - x^2\} \\ +\frac{\alpha_4}{24n}\{(\varphi-\varphi^3)+(3-\varphi^2)x - \varphi x^2 - x^3\} \\ +\frac{\alpha_3^2}{72n}\{(-3\varphi-\varphi^3+\varphi^5)+(-15+3\varphi^2+\varphi^4)x+(6\varphi+\varphi^3)x^2 \\ +(10-\varphi^2)x^3-\varphi x^4 - x^5\} \\ +\frac{\gamma}{4n}\{-2x+\varphi x^2+x^3\} \end{array}\right). \tag{3.3}$$

Note that the terms that appear with $\phi(x-\varphi)$ can be moved inside $\Phi(\cdot)$ by using the identities $\Phi'(x) = \phi(x)$ and $\Phi''(x) = -x\varphi(x)$. Then solving the equation $\Phi(z) = F(x;\phi)$ for z, we obtain

$$F(x;\varphi) = \Phi\left(\begin{array}{ll} (x-\varphi) & +\frac{\alpha_3}{6n^{1/2}}\{1-\varphi^2-\varphi x-x^2\} \\ & +\frac{\alpha_4}{24n}\{(\varphi-\varphi^3)+(3-\varphi^2)x-\varphi x^2-x^3\} \\ & +\frac{\alpha_3^2}{72n}\{(-4\varphi+\varphi^3)+(-14+3\varphi^2)x+6\varphi x^2+8x^3\} \\ & +\frac{\gamma}{4n}\{-2x+\varphi x^2+x^3\} \end{array}\right). \tag{3.4}$$

Equations (3.3) and (3.4) are asymptotically equivalent up to order $O(n^{-3/2})$.

If we are interested in just the p-value at the data point y^0, given as $x = 0$ in the standardized notation, we have $p^0(\varphi) = F(y^0;\theta) = F(0,\varphi)$,

$$p^0(\varphi) = \Phi(-\varphi) + \phi(-\varphi)\left\{\frac{\alpha_3}{6n^{1/2}}(1-\varphi^2) + \frac{\alpha_4}{24n}(\varphi-\varphi^3) + \frac{\alpha_3^2}{72n}(-3\varphi-\varphi^3+\varphi^5)\right\} \tag{3.5}$$

$$p^0(\varphi) = \Phi\left\{-\varphi + \frac{\alpha_3}{6n^{1/2}}(1-\varphi^2) + \frac{\alpha_4}{24n}(\varphi-\varphi^3) + \frac{\alpha_3^2}{72n}(-4\varphi+\varphi^3)\right\}, \tag{3.6}$$

which are obtained from (3.3) and (3.4) respectively. We note the important fact that the p-value is independent of γ and thus depends only on α_3 and α_4; and as we recall these can be obtained from just the observed likelihood and the reparameterization φ obtained as the gradient of the likelihood at the observed data point y^0. In other words, the nonexponentiality as represented by the γ coefficient has no effect on the p-value to order $O(n^{-3/2})$ and we can thus quite generally obtain the p-values from the tangent exponential model (2.7); this has remarkable implications. Also note that (3.3) and (3.4) go to 0 and 1 at $-\infty$ and $+\infty$ respectively, thus verifying the integration property for (2.6).

3.2 Distribution function for the location type reexpression

The location type reexpression has Taylor coefficients recorded in the array (2.8). As in the exponential type reexpression, all the coefficients in the array can be obtained from the observed likelihood function and norming properties of the model. The density function from (2.8) has the form

$$f(x;\beta) = h(z)\exp\left\{\frac{\gamma}{4n}(z^2\beta^2 - z^4 + 5z^2 - 2)\right\} \tag{3.7}$$

where with $\gamma = 0$, $z = x - \beta$ has density

$$h(z) = \exp\left\{\frac{3\alpha_4 - 5\alpha_3^2}{24n}\right\}\varphi(z)\exp\left\{\alpha_3\frac{z^3}{6n^{1/2}} - \alpha_4\frac{z^4}{24n}\right\}\ . \tag{3.8}$$

The error type component (3.8) can be expanded to the third order giving

$$h(z) = \exp\left\{\frac{3\alpha_4 - 5\alpha_3^2}{24n}\right\}\varphi(z)\left\{1 + \alpha_3\frac{z^3}{6n^{1/2}} - \alpha_4\frac{z^4}{24n} + \alpha_3^2\frac{z^6}{72n}\right\}\ . \tag{3.9}$$

Computer algebra is then used as at (3.3) to give the corresponding distribution function

$$H(z) = \Phi(z) + \varphi(z)\left\{\begin{array}{l}\frac{\alpha_3}{6n^{1/2}}(-2 - z^2) \\ +\frac{\alpha_4}{24n}(3z + z^3) \\ +\frac{\alpha_3^2}{72n}(-15z - 5z^3 - z^5)\end{array}\right\} \tag{3.10}$$

which to order $O(n^{-3/2})$ can be reexpressed as

$$H(z) = \Phi \left\{ \begin{array}{ll} z & +\frac{\alpha_3}{6n^{1/2}}(-2 - z^2) \\ & +\frac{\alpha_4}{24n}(3z + z^3) \\ & +\frac{\alpha_3^2}{72n}(-11z - z^3) \end{array} \right\} . \tag{3.11}$$

Thus we have the following two versions for the distribution functions $F(u;\beta)$:

$$F(x;\beta) = \Phi(x-\beta) + \phi(x-\beta) \left[\begin{array}{l} \frac{\alpha_3}{6n^{1/2}}\{-2 - (x-\beta)^2\} \\ +\frac{\alpha_4}{24n}\{3(x-\beta) + (x-\beta)^3\} \\ +\frac{\alpha_3^2}{72n}\{-15(x-\beta) - 5(x-\beta)^3 - (x-\beta)^5\} \\ +\frac{\gamma}{4n}\{-2x + \beta x^2 + x^3\} \end{array} \right] \tag{3.12}$$

$$F(x;\beta) = \Phi \left[\begin{array}{ll} (x-\beta) & +\frac{\alpha_3}{6n^{1/2}}\{-2 - (x-\beta)^2\} \\ & +\frac{\alpha_4}{24n}\{3(x-\beta) + (x-\beta)^3\} \\ & +\frac{\alpha_3^2}{72n}\{-11(x-\beta) - (x-\beta)^3\} \\ & +\frac{\gamma}{4n}\{-2x + \beta x^2 + x^3\} \end{array} \right] . \tag{3.13}$$

If we are interested in just the p-value at the data y^0 which is $x = 0$ in the standardized notation, we have $p^0(\beta) = F(y^0;\theta) = F(0;\beta)$ which gives

$$p^0(\beta) = \Phi(-\beta) + \phi(-\beta) \left\{ \frac{\alpha_3}{6n^{1/2}}(-2-\beta^2) + \frac{\alpha_4}{24n}(-3\beta - \beta^3) + \frac{\alpha_3^2}{72n}(15\beta + 5\beta^3 + \beta^5) \right\} \tag{3.14}$$

$$p^0(\beta) = \Phi \left\{ -\beta + \frac{\alpha_3}{6n^{1/2}}(-2-\beta^2) + \frac{\alpha_4}{24n}(-3\beta - \beta^3) + \frac{\alpha_3^2}{72n}(11\beta + \beta^3) \right\} . \tag{3.15}$$

As in the exponential type reexpression, the p-value is independent of γ which means that the nonlocationality γ has no effect on the p-value to order $O(n^{-3/2})$ and we can then obtain p-values also from the tangent location model.

Notice that (3.14), (3.15) have a different form from that in (3.5), (3.6) but this is due to the different parameterization and the correspondingly different Taylor series coefficients. An important result of these equations, (3.3) to (3.6) and (3.12) to (3.15) is that they are continuous functions. That is, there is no singularity, whereas the usual Lugannani and Rice and Barndorff-Nielsen formulas have a computational singularities at the maximum likelihood estimate (Daniels, 1987). Of course there are ways to avoid the singularity: see for example, Robinson (1982) and Daniels (1987) which in effect make use of derivatives of the likelihood as we have here; there are also computational and plotting methods that bridge the singularity (for example, Fraser, Reid, Li, & Wong, 2003). We thus have an important tool to eliminate the singularity problem that exists in the standard likelihood based higher order asymptotic methods.

4. EXAMPLES

We choose to assess the integration methods by examining the smallest possible sample size and two extreme types of statistical model: the exponential model with rate parameter θ; and the location Cauchy model with location parameter θ. We will see that remarkable accuracy is obtained even for unnaturally small sample sizes.

Example 4.1. Consider a sample of $n = 1$ from the extremely nonnormal scale exponential model $f(y;\theta) = \theta \exp\{-\theta y\}$ with $\theta > 0$, $y > 0$. The exact distribution function is of course available:

$$F(y;\theta) = 1 - \exp\{-\theta y\} \ . \tag{4.1}$$

Without loss of generality we take the parameter values $\theta = \theta^0 = 1$ and seek distribution function values for various y values around $y^0 = 1$ which is the point with maximum likelihood value $\hat{\theta}(y^0) = \theta^0 = 1$. The standardization equations (2.2) and (2.3) give

$$\bar{y} = (y-1), \quad \bar{\theta} = (1-\theta).$$

These produce a new observed information equal to 1 and make the observed gradient of the score parameter also equal to 1. Following Cakmak *et al* (1998) for the exponential reexpressions, we have

$$x = \bar{y} \ , \quad \varphi = \bar{\theta}$$

together with the revised coefficient array

$$\bar{A} = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & 0 & - & - \\ -2 & 0 & - & - & - \\ -6 & - & - & - & - \end{pmatrix} \; ;$$

whereas for the location reexpressions, we have

$$x = \bar{y} - \bar{y}^2/2 + 2\bar{y}^3/6, \qquad \beta = \bar{\theta} + \bar{\theta}^2/2 + 2\bar{\theta}^3/6$$

together with the revised coefficient array

$$\bar{A} = \begin{pmatrix} -1 & 0 & -1 & -1 & -1 \\ 0 & 1 & 1 & 1 & - \\ -1 & -1 & -1 & - & - \\ 1 & 1 & - & - & - \\ -1 & - & - & - & - \end{pmatrix} \; .$$

The above two expressions for $\bar{A}$ together with the formulas (3.4) and (3.13) respectively then give Table 4.1 which records approximation values for the distribution function for $\theta = 1$ and selected values of y near 1. The table also records the approximation values when full likelihood information is available at each point and Barndorff-Nielsen formula is used directly. Note that the Barndorff-Nielsen formula cannot be applied directly to give values at $y = 1$ because at this point it has a computational singularity, but various limiting expression are available (see for example Daniels, 1987; Fraser *et al*, 2003, Robinson, 1982; Yand and Kolassa, 2002). Of particular importance is that the special likelihood information at the observed data point seems to provide broadly almost the same accuracy as full likelihood information at the various points of interest.

Table 4.1. Approximate values for $F(y;1)$ using the third order exponential expansion (3.4) about $(y^0, \hat{\theta}^0) = (1,1)$, the third order location expansion (3.13) about $(y^0, \hat{\theta}^0) = (1,1)$, the multi-point likelihood approximation about the y-value being examined, and the exact value (4.1).

y	0.8	0.9	1.0	1.1	1.2
Exponential (3.4)	0.5494	0.5920	0.6306	0.6655	0.6971
Location (3.13)	0.5497	0.5920	0.6306	0.6655	0.6973
Barndorff-Nielsen	0.5494	0.5920	0.6306	0.6655	0.6971
Exact (4.1)	0.5507	0.5934	0.6321	0.6672	0.6988

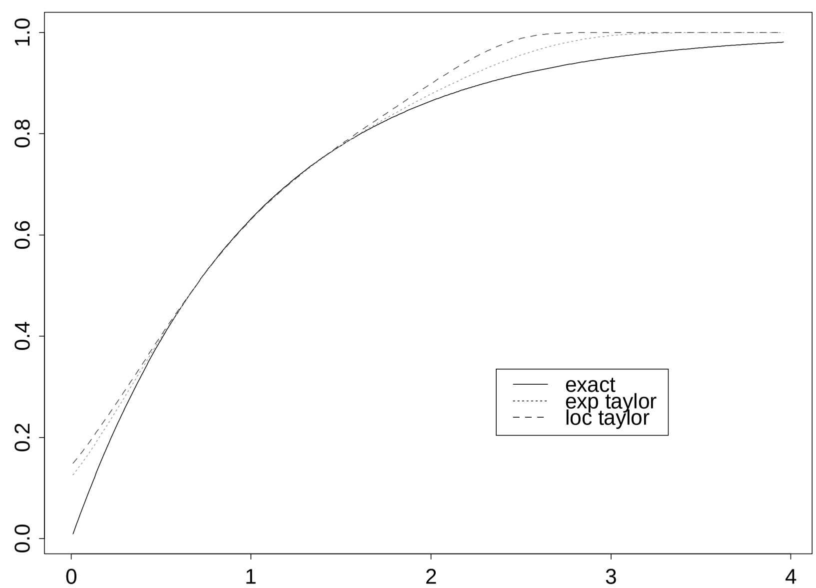

Figure 1. Exponential Example 4.1: The exact distribution function, the exponential type approximation (3.4), and the location type approximation (3.13).

Figure 1 plots the results for the distribution function $F(y;1)$ as calculated from the third order exponential expansion (3.4), from the third order location expansion (3.13), and from the exact (4.1). In a substantial range around $y^0 = 1$, both (3.4) and (3.13) give uniformly excellent approximations with (3.4) slightly more accurate than (3.13). However, beyond this range the Taylor series approximations suddenly degrade quite rapidly.

Example 4.2. Consider the location Cauchy $f(y;\theta) = \pi^{-1}\{1+(y-\theta)^2\}^{-1}$ on $(-\infty,\infty)$. The exact distribution function is

$$F(y;\theta) = .5 + \pi^{-1}\tan^{-1}(y-\theta) \ . \qquad (4.2)$$

Again we take the parameter value $\theta^0 = 1$ and seek distribution function values for various y values around $y^0 = 1$ which is the point with maximum likelihood value $\hat{\theta}(y^0) = 1$. The standardization equations give

$$\bar{\theta} = \sqrt{2}(\theta - 1), \qquad \bar{y} = \sqrt{2}(y - 1) \ .$$

The exponential reexpressions in Section 2 then give

$$x = \bar{y} - 3\bar{y}^3/6, \qquad \varphi = \bar{\theta} - 3\bar{\theta}^3/6$$

together with the revised coefficient array

$$\bar{A} = \begin{pmatrix} -0.3466 & 0 & 2 & 0 & -9 \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & 3 & - & - \\ 0 & 0 & - & - & - \\ -9 & - & - & - & - \end{pmatrix} .$$

In a related way the location reexpressions give

$$\beta = \bar{\theta} - 2\bar{\theta}^3/6, \qquad x = \bar{y}$$

with the location type coefficient array

$$\bar{A} = \begin{pmatrix} -0.3466 & 0 & -1 & 0 & 3 \\ 0 & 1 & 0 & -3 & - \\ -1 & 0 & 3 & - & - \\ 0 & -3 & - & - & - \\ 3 & - & - & - & - \end{pmatrix} .$$

Table 4.2 records approximation values for the distribution function for $\theta = 1$ and selected y value near 1. Again of particular importance we have that the restricted (observed data) likelihood information seems to perform as well as the full likelihood information at the particular points of interest.

Table 4.2. Approximate values for $F(y; 1)$ using the Laplace (4.1), the third order exponential expansion (3.4) about (1,1), the third order location expansion (3.13) about (1,1),

the multipoint likelihood approximation about the y value being examined, and the exact value (4.2).

y	0.8	0.9	1.0	1.1	1.2
Exponential (3.4)	0.4297	0.4647	0.5000	0.5353	0.5703
Location (3.15)	0.4310	0.4649	0.5000	0.5351	0.5690
Barndorff-Nielsen	0.4305	0.4649	0.5000	0.5351	0.5695
Exact (4.2)	0.4372	0.4683	0.5000	0.5317	0.5628

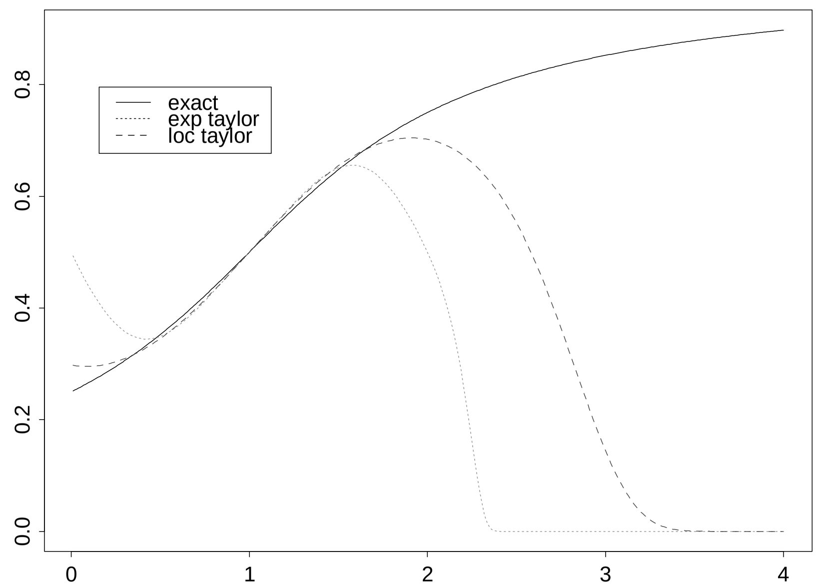

Figure 2. Cauchy Example 4.2: The exact distribution function, the exponential type approximation (3.4) and the location type approximations (3.13).

Figure 2 plots the results for the distribution function $F(y;1)$ as calculated from the third order exponential expansion (3.4), from the third order location expansion (3.13), and from the exact (4.2). When y is within a broad range about $y^0 = 1$, both (3.4) and (3.13) give excellent approximations with (3.13) slightly more accurate than (3.4). However, when y is beyond this range the approximations degrade very quickly. What seems quite remarkable however is how accurately the two Taylor series based formulas track the exact in a rather broad range about the observed data point y_0.

5. DISCUSSION AND ACKNOWLEDGEMENTS

We use computer algebra to obtain explicit third order expressions for the distribution function of an asymptotic scalar model. The coefficients for the distribution function expansion come from Taylor series coefficients obtained from likelihood at some initial data point. Also the resulting distribution function expressions provide explanations for some high accuracy results in recent likelihood theory. It is seen that the distribution function at the expansion data point y_0 is fully third order accurate using only likelihood and likelihood gradient at the data point and thus does not need any sample space second or higher derivatives; in other words at the observed data point the likelihood function provides all the information normally found in a cumulant generating function.

The expansion and the resulting formulas also provide elementary proofs for key formulas of likelihood asymptotics: the p^* formula and the tangent exponential model. They also provide simple numerical integration results: we obtain accurate distribution function values using only information at and to first derivative at the expansion point.

An interactive program written in C with Maple is available for the integrations and can be obtained from the third author.

For work in progress towards third order methods for combining independent data, a key ingredient is the distribution function for components in the neighbourhood of a data point. The present paper provides key background towards this. The Natural Sciences and Engineering Research Council provided support for this research.

REFERENCES

Abebe, F., Cakmak, S., Cheah, P.K., Fraser, D.A.S., Kuhn, J., Reid, N.(1995) Third order asymptotic model: Exponential and location type approximations, *Parisankhyan*

Samikkha. **2**, 25-34.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimation. *Biometrika* **70**, 343-365.

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. (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. *Jour. Roy. Statist. Soc.* **B41**, 279-312.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* London: Chapman and Hall.

Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N. and Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference* **66**, 211-222. Some minor typographical corrections may be accessed through http://www.utstat.toronto.edu/dfraser/research.html

Daniels, H.E. (1954). Saddlepoint approximation in statistics. *Annals Math. Statist.* **25**, 631-650.

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

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

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

Fraser, D.A.S. and McDunnough, P. (1984). Further remarks on asymptotic normality and conditional analysis. *Canadian Jour. Statist.* **12**, 183-190.

Fraser, D.A.S. and Reid, N. (1993). Third order asymptotic models: likelihood functions leading to accurate approximations for distribution functions. *Statist. Sinica* **3**, 67-82.

Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third-order significance. *Utilitas Mathematica* **77**, 33-53.

Fraser, D.A.S. and Reid, N. (2001). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference,* New York: Springer-Verlag, 185-207.

Fraser, D.A.S., Reid, N., Li, R., and Wong, A. (2003). On bridging the singularities of p-value formulas from likelihood analysis. *J. Statist. Research* **37**, 1-15.

Fraser, D.A.S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249-264.

Fraser, D.A.S., Wong, A., and Wu, J. (1999). Regression analysis, nonlinear or nonnormal: simple and accurate p-values from likelihood analysis. *J. Amer. Statist. Assoc.* **94**, 1286-1295.

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

Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44**, 91-101.

Yang, B. and Kolassa, J.E. (2002). Saddlepoint approximation for the distribution function near the mean. *Ann. Inst. Statist. Math.* **54**, 743-747.