

NONLINEAR REGRESSION: THIRD ORDER SIGNIFICANCE

F. Abebe

Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

D.A.S. Fraser

Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

A. Wong

Department of Mathematics and Statistics
York University
North York, Canada, M3J 1P3

SUMMARY

Some recent methods of third order asymptotics produce remarkably accurate significance probabilities but implicitly require the variable to have the same dimension as the parameter. A method for constructing an approximate ancillary given in Fraser & Reid (1995) extends these methods to a general models with asymptotic properties. These extensions are used to give third order significance for a real parameter in a non-linear regression model. Emphasis is placed on the normal error model but extends to the non-normal additive error case and to generalized linear and nonlinear models; discussion here is restricted to the known variance case. Examples are given illustrating accuracy and generality, and comparisons are made with some first order methods.

1. INTRODUCTION

Recent methods of third order asymptotics can produce significance probabilities that have remarkable accuracy, but for implementation require the sample space dimension to be equal to that for the full parameter. From a practical viewpoint this effectively restricts the approximations to location, transformation and exponential models where the reduction methods of conditioning or sufficiency are available to give a variable of the same dimension as the parameter.

The recent third order significance approximations have evolved from the saddlepoint method (Daniels, 1954; Barndorff-Nielsen & Cox, 1979) or from the direct analysis and Taylor series expansion of log density functions (Barndorff-Nielsen, 1986; DiCiccio, Field & Fraser, 1990; Fraser, 1990; Fraser & Reid, 1993). Some early formulas used mean and variance corrected quantities but the exceptional accuracy of some recent formulas seems to arise from a natural invariance obtained from the use of simple likelihood characteristics. Some general remarks on significance and third order approximation are given in Fraser (1991).

For an exponential model with a real canonical parameter, see Lugannani & Rice (1980). For a real component parameter in a general model, a mean and variance corrected formula is given in Barndorff-Nielsen (1986), but the implementation was rather complicated as subsequently noted by Barndorff-Nielsen (1991). For a canonical component of a transformation model, see DiCiccio, Field & Fraser (1990). For a real parameter general model or a canonical component of an exponential model, see Fraser & Reid (1993). For a real component parameter in a general model, see Barndorff-Nielsen (1991) and Fraser & Reid (1995); the latter has an ease of implementation obtained from the use of tangent location and tangent exponential models.

The preceding approximations are essentially restricted to the case where the dimen-

sion of the variable is the same as that of the parameter. For the more common situation where the dimension of the variable is larger, Barndorff-Nielsen (1983) gave the existence of a third order ancillary but not an implementable construction procedure. Fraser & Reid (1993; 1995) show that only the observed likelihood and the likelihood gradient tangent to the ancillary at the data point are needed for third order inference and that it suffices to have the gradient with respect to a second order ancillary. Fraser & Reid (1995) used the tangent location method (Fraser, 1964) to generate the tangent directions for the second order ancillary.

The preceding results give third order significance for a real parameter in a general model context. For some examples see Fraser & Reid (1995) and Fraser, Monette, Ng & Wong (1992). The examples include generalized linear models but the method applies more generally needing only the determination of the effects of a parameter change on the individual coordinates of the model.

In Section 2 we summarize relevant formulas for third order significance. In Section 3 we discuss the minimum information needed concerning a third order ancillary. In Section 4 and 5 we discuss third order significance for the normal and non-normal regression models. Examples are given in Section 6.

The statistical analysis of nonlinear regression models was examined by Beale (1960) using differential geometry and extended in work by Guttman & Meeter (1965), Bates & Watts (1980,1981), and Hamilton, Watts & Bates (1982). In particular, Bates & Watts (1980) contributed significantly by identifying the two orthogonal components of the second derivative array $\ell_{\theta\theta}(\theta)$, the intrinsic curvature that relates to the geometry of the model and which is invariant to reparameterization, and parameter effect curvature that depends on the parametrization of the model. For normally distributed errors, Bates & Watts (1980) proposed two measures of nonlinearity directly related to the maximum relative intrinsic curvature and maximum relative parameter effects curvature, respectively. The

measures were extended to a general nonlinear model context in Amari (1982) and Kass (1984) (see also Cook and Witmer, 1985).

Second order inference for nonlinear regression models was treated in Hamilton, Watts & Bates (1982), Hamilton (1986), and Fraser & Massam (1987). The analyses by Hamilton, Watts & Bates (1982) and Hamilton (1986) are based on approximating the response surface by a tangent surface to the second order. Fraser & Massam's (1987) analysis is related to a conditional testing procedure developed in Fraser & Massam (1985) in which two levels of conditioning were introduced: one for the data direction and the other for the distance from data point to the response surface. For normal errors, they use the Fisher-von-Mises distribution in place of χ^2 to assess observed departure.

2. BACKGROUND

Consider a statistical model $f(y; \theta)$ and an observed data point y^0 and suppose the model has asymptotic properties (for example, DiCiccio, Field & Fraser, 1990; Fraser & Reid, 1993a) as some parameter n , say sample size, becomes large; let $\ell(\theta; y)$ be the log-likelihood function.

If y and θ are scalar, a tangent exponential model (Fraser & Reid, 1995)

$$\frac{c}{(2\pi)^{1/2}} \exp\{\ell(\theta; y^0) - \ell(\hat{\theta}^0; y^0) + (\varphi - \hat{\varphi}^0)s\} |\tilde{j}|^{-1/2} ds \quad (2.1)$$

gives an $O(n^{-3/2})$ approximation to first derivative at y^0 and an $O(n^{-1})$ approximation generally (within a compact set for the standardized variable); for this, φ is a locally defined canonical parameter, s is the corresponding score variable,

$$\begin{aligned} \varphi = \varphi(\theta) &= \frac{d}{dy} \ell(\theta; y)|_{y^0} = \ell_{;y}(\theta; y^0), \\ s = s(y) &= \frac{\partial}{\partial \varphi} \ell(\theta; y)|_{\hat{\theta}^0} = \ell_{\varphi}(\hat{\theta}^0; y), \end{aligned} \quad (2.2)$$

and \tilde{j} is the observed information for φ from the likelihood in the exponent of (2.1); for convenience we assume that φ is locally increasing in θ , otherwise we reverse the sign

of y . The nominal parameter φ appears in this as a computational device to obtain the significance approximation.

Fraser & Reid (1993) show that the distribution function for the exponential model (2.1) coincides with that for the given model to order $O(n^{-3/2})$ at the data point y^0 ; let $\ell^0(\theta) = \ell(\theta; y^0)$ be the observed log-likelihood function. The significance function $p^0(\theta) = P(\hat{\theta} \leq \hat{\theta}^0; \theta)$ can then be obtained from the exponential model formula of Lugannani & Rice (1980),

$$p^0(\theta) = \Phi(r) + \phi(r) \left(\frac{1}{r} - \frac{1}{q} \right) + O(n^{-3/2}) \quad (2.3)$$

where r is the signed root of the log likelihood ratio

$$r = \text{sgn}(\hat{\theta}^0 - \theta) \cdot \left[2\{\ell^0(\hat{\theta}^0) - \ell^0(\theta)\} \right]^{1/2}, \quad (2.4)$$

q is the standardized maximum likelihood departure on the φ scale

$$q = (\hat{\varphi}^0 - \varphi) \hat{j}_{\varphi\varphi}^{1/2}, \quad (2.5)$$

$\hat{j}_{\varphi\varphi}$ is the observed information for φ at the data point, and $\phi(z)$, $\Phi(z)$ are the standard normal density and distribution functions. The r^* formula of Barndorff-Nielsen (1986,1991) provides an alternative approximation to the significance function

$$p^0(\theta) = \Phi(r^*) + O(n^{-3/2}) \quad (2.6)$$

where the corrected likelihood ratio r^* is defined using (2.4) and (2.5),

$$r^* = r - r^{-1} \log(r/q), \quad (2.7)$$

and has a standard normal distribution to order $O(n^{-3/2})$. The two approximations (2.3) and (2.6) use the same r and q ; we find it convenient to use the notation $p^0(\theta) = \Phi(r, q)$ for either. For contrast, two common first order approximations to $p^0(\theta)$ are given by $\Phi(r)$ and

$\Phi(u)$ where r is the signed root of the log likelihood ratio as above and $u = (\hat{\theta} - \theta) \hat{J}_{\theta\theta}^{1/2}$ is the standardized maximum likelihood departure based on the original parameter θ rather than on the pseudo parameter φ as in (2.5).

If y and θ are vectors of dimension p , then (2.1) still applies with $(2\pi)^{1/2}$ replaced by $(2\pi)^{p/2}$ and with φ and s treated as row and column vectors. For this we take s as defined in (2.2) where φ is the sample space gradient of the likelihood function,

$$\varphi = \varphi(\theta) = \frac{\partial}{\partial y'} \ell(\theta; y) \Big|_{y^0} = \frac{d}{dV} \ell(\theta; y) \Big|_{y^0} \quad (2.8)$$

at the data point y^0 . The final expression with differentiation with respect to p basis vectors $V = (v_1, \dots, v_p)$ will be used in the presence of an ancillary where the vectors V are tangent to the ancillary at the data point, that is, $\mathcal{L}(V)$ is the tangent plane to the ancillary surface at the data point. For computation, any affinely equivalent version of φ can be used; this amounts to a relocation and change of basis vectors which does not affect the value of the expression (2.1). Then, for testing a scalar interest parameter ψ where $\theta = (\psi, \lambda)$, a generalization of approximation (2.3) is available. Let r now be the signed root of the profile log likelihood ratio

$$r = \text{sgn}(\hat{\psi}^0 - \psi) \cdot \left[2 \{ \ell^0(\hat{\lambda}^0, \hat{\psi}^0) - \ell^0(\hat{\lambda}_\psi^0, \psi) \} \right]^{1/2}, \quad (2.9)$$

where $\hat{\lambda}_\psi^0$ is the constrained maximum likelihood estimate with ψ fixed. It is straightforward to find a scalar linear function $\chi(\theta) = \sum a_i \varphi_i(\theta)$ of φ with $\sum a_i^2 = 1$ such that $d\chi = 0$ is equivalent to $d\psi = 0$ at $(\hat{\lambda}_\psi^0, \psi)$ and increments positively with respect to ψ at that point; this can be interpreted as a sample space gradient of the profile likelihood. Then $p^0(\psi) = P(\hat{\psi} \leq \hat{\psi}^0; \psi)$ is given by (2.3) with r now defined by (2.9) and q replaced by

$$Q = \{ \chi(\hat{\theta}^0) - \chi(\hat{\theta}_\psi^0) \} \cdot \frac{J_p^{1/2}}{J_{p-1}^{1/2}} = M \frac{J_p^{1/2}}{J_{p-1}^{1/2}} \quad (2.10)$$

where the full and nuisance informations, J_p and J_{p-1} , are calibrated on the φ scale,

$$J_p = |j_{\varphi\varphi}(\hat{\theta}_0)| = |j_{\theta\theta}(\hat{\theta}^0)| |\varphi_{\theta'}(\hat{\theta}^0)|^{-2}, \quad J_{p-1} = |j_{\lambda\lambda}(\hat{\theta}_\psi^0)| |\varphi_{\lambda'}(\hat{\theta}_\psi^0)|^{-2} \quad (2.11)$$

and $\varphi_{\theta'}$ is the Jacobian

$$\varphi_{\theta'}(\theta) = \frac{\partial\varphi(\theta)}{\partial\theta'} \quad (2.12)$$

for the parameter transformation to φ from θ . The determinant of the $(p-1) \times p$ matrix $\varphi_{\lambda'} = A$ is defined by $|A| = |AA'|^{1/2}$. The quotient calculation of information in (2.10) takes account of marginalization over the nuisance parameter distribution. For details see Fraser & Reid (1995).

3. ANCILLARY DIRECTIONS AT THE DATA POINT

Consider the nonlinear regression model

$$y_i = \eta_i(\theta) + e_i \quad (3.1)$$

where $\eta_i(\theta)$ is a smooth real-valued function of $\theta = (\lambda_1, \dots, \lambda_{p-1}, \psi)$ and the e_i are independent with error density $f_i(e_i)$. We assume here that the error scaling is known; the more general case with unknown error scaling will be discussed elsewhere. We also assume that the coordinates of the parameter have been defined and reordered so that the interest parameter ψ is the last coordinate. Each coordinate is already in location form with respect to its main parameter η_i .

Now consider the effect of a θ change on the full variable y at y^0 ; the vector directions for data change at y^0 that correspond to parameter change from $\hat{\theta}^0$ to $\hat{\theta}^0 + d\theta$ are given by the $n \times p$ matrix $X_p = X_p(\hat{\theta}^0) = (x_1, \dots, x_p)$, obtained from

$$X_p(\theta) = \eta_\theta(\theta) = \left\{ \frac{\partial\eta(\theta)}{\partial\lambda_1}, \dots, \frac{\partial\eta(\theta)}{\partial\lambda_{p-1}}, \frac{\partial\eta(\theta)}{\partial\psi} \right\}, \quad (3.2)$$

which records p tangent vectors to the regression surface at the maximum likelihood value. These tangent vectors X_p at the data point y^0 generate a plane parallel to the tangent plane at the maximum likelihood point $\hat{\eta}^0 = \eta(\hat{\theta}^0)$. As defined, the tangent vectors at the data point correspond to a first order ancillary, but analysis in Fraser & Reid (1993b) shows they are also tangent to a second order ancillary. For later use we also define

$$X_{p-1}(\theta) = \eta_\lambda(\theta) = \left\{ \frac{\partial \eta(\theta)}{\partial \lambda_1}, \dots, \frac{\partial \eta(\theta)}{\partial \lambda_{p-1}} \right\}.$$

4. TESTING ψ WITH NORMAL ERROR

We now apply the asymptotic results from Section 2 to the nonlinear regression model in Section 3, and for the moment restrict attention to the case of normal error with common variance σ_0^2 . The more general case with non-normal error and fixed scaling is discussed in Section 5.

First we derive the canonical parameter (2.8) of the tangent exponential model. The parameter is constructed by differentiating likelihood in the ancillary plane directions at the data point; using the vectors X_p , we obtain the corresponding nominal parameter as

$$\frac{d}{dX_p} \left[-\frac{1}{2\sigma_0^2} \sum \{y_i - \eta_i(\theta)\}^2 \right]_{y^0} = \frac{1}{\sigma_0^2} (\eta(\theta) - y^0)' X_p; \quad (4.1)$$

for simplicity we choose the affinely equivalent canonical parameter

$$\varphi(\theta) = \eta'(\theta) X_p \quad (4.2)$$

which gives coordinates for $\eta(\theta)$ as projected to the tangent plane relative to dual basis vectors; this parameterization has been examined by Hamilton, Bales & Watts (1982). Some geometrical simplicity is obtained if X_p in (4.2) is replaced by a set of orthonormal vectors that span the linear space generated by X_p .

Next we determine the scalar parameter $\chi = \sum a_i \varphi_i$ that gives the same local parameterization as ψ does in the neighbourhood of the parameter value $\widehat{\theta}_\psi^0$; for a related parameter see Hamilton (1986). Using Jacobian

$$\varphi_{\theta'}(\widehat{\theta}_\psi^0) = \left. \frac{\partial \varphi(\theta)}{\partial \theta'} \right|_{\widehat{\theta}_\psi^0} = X'_p(\widehat{\theta}_\psi^0) X_p$$

at $\widehat{\theta}_\psi^0$, we can write $d\varphi = (d\lambda, d\psi)\varphi_{\theta'}(\widehat{\theta}_\psi^0)$, and thus obtain $d\psi = d\varphi \varphi'^\psi(\widehat{\theta}_\psi^0)$ where $\varphi'^\psi(\widehat{\theta}_\psi^0)$ is the final column of $\{\varphi_{\theta'}(\widehat{\theta}_\psi^0)\}^{-1}$. Thus

$$\chi(\theta) = \varphi(\theta) w_p, \quad w_p = \frac{\varphi'^\psi(\widehat{\theta}_\psi^0)}{|\varphi'^\psi(\widehat{\theta}_\psi^0)|}, \quad (4.3)$$

and w_p is the unit vector such that the new coordinate $\chi(\theta)$ has a surface tangent to the surface for ψ at $\widehat{\theta}_\psi^0$.

Now consider the testing of a value for the parameter ψ . The signed likelihood ratio is

$$r = \text{sgn}(\widehat{\psi}^0 - \psi) \cdot \frac{1}{\sigma_0} \left[\sum \{y_i^0 - \eta_i(\widehat{\theta}_\psi^0)\}^2 - \sum \{y_i^0 - \eta_i(\widehat{\theta}^0)\}^2 \right]^{1/2}. \quad (4.4)$$

The parameter component from (4.2) that is equivalent to $d\psi$ at $\widehat{\theta}_\psi^0$ gives the departure

$$M = \text{sgn}(\widehat{\psi}^0 - \psi) \cdot |\{\varphi(\widehat{\theta}^0) - \varphi(\widehat{\theta}_\psi^0)\} w_p|; \quad (4.5)$$

for use in (2.10). This is the tangent plane departure perpendicular to the submodel at its maximum likelihood value. The related information expressions from (2.11) are

$$\begin{aligned} J_{p-1} &= \frac{1}{\sigma_0^{2(p-1)}} \left| \sum \{\eta_i(\widehat{\theta}_\psi^0) - y_i^0\} \eta_{i\lambda'\lambda}(\widehat{\theta}_\psi^0) + X'_{p-1}(\widehat{\theta}_\psi^0) X(\widehat{\theta}_\psi^0) \right| \cdot \left| X'_{p-1}(\widehat{\theta}_\psi^0) X_p \right|^{-2} \\ J_p &= \frac{1}{\sigma_0^{2p}} \left| \sum \{\eta_i(\widehat{\theta}^0) - y_i^0\} \eta_{i\theta'\theta}(\widehat{\theta}^0) + X'_p X_p \right| \cdot \left| X'_p X_p \right|^{-2}, \end{aligned} \quad (4.6)$$

where $\eta_{i\lambda'\lambda}(\theta)$ and $\eta_{i\theta'\theta}$ are the second derivative matrices of $\eta_i(\theta)$ with respect to λ and $\theta = (\lambda, \psi)$ respectively. Then the third order significance for testing ψ is given by (2.3) using the signed likelihood ratio r from (4.4) and standardized maximum likelihood quantity $Q = M(J_p/J_{p-1})^{1/2}$ from (4.5) and (4.6).

5. TESTING ψ WITH NON-NORMAL ERROR

Much of the simplicity of the normal error case follows from the canonical parameter being given by the projection of the regression surface to the tangent plane at the maximum likelihood point. More generally it seems appropriate to work directly from likelihood.

Consider the regression model (3.1) with known error components $f_i(e_i)$ and let $\ell_i(e_i) = \log f_i(e_i)$. Then the observed likelihood is

$$\ell(\theta; y^0) = \sum_{i=1}^{\eta} \ell_i(y_i^0 - \eta_i(\theta)) \quad (5.1)$$

and the locally defined canonical parameter is the row vector

$$\varphi(\theta) = \sum_{i=1}^n m_i(y_i^0 - \eta_i(\theta)) X_{(i)} \quad (5.2)$$

where $m_i(y) = d\ell_i(y)/dy$ and $X_{(i)}$ is the i th row of the tangent plane vector array $X(\hat{\theta}^0)$ from (3.3).

For testing a value for ψ in $\theta = (\lambda, \psi)$, the signed likelihood ratio is given by (2.8), and the standardized maximum likelihood departure Q is given by (2.10) with (2.11), where the parameter $\chi = \chi(\theta) = \varphi(\theta)w_p$ is defined by (4.3) with $\varphi'^{\theta}(\theta) = \{\varphi_{\theta'}(\theta)\}^{-1}$ and $\varphi_{\theta'}(\theta) = \partial\varphi(\theta)/\partial\theta'$ as obtained from (5.2). The third order significance for testing ψ is then obtained from (2.3) with the r and Q as just described.

6. EXAMPLES

For linear regression with normal error and known scaling the present method reproduces the usual normal analysis. For linear regression with non-normal error components the present methods reproduce the main approximation in DiCiccio, Field & Fraser (1990), but have the computational advantage of being directly and simply expressed in terms of likelihood. Both these normal and non-normal analyses in the literature are restricted to

the case of canonical or linear parameters. If we generalize from the case with canonical or linear parameters, we typically lose the sufficiency or conditionality reduction needed for the methods in the literature.

For the general case of nonlinear regression there are several issues concerning the calculation of approximate significance for a scalar component parameter ψ . First, there is the need for an approximate ancillary, a need attributable to the nonlinearity of the regression surface. Second, there is the effect of non-normality, if present, on the calculations; for this, the examples in DiCiccio, Field, & Fraser (1990) indicate high accuracy corresponding to the likelihood based approximation here. Third, there is the effect of curvature in the interest parameter ψ . Our primary concern then is to illustrate the first involving the construction of an ancillary and the third involving the presence of curvature in the interest parameter ψ .

The first three examples consider the effect of the construction method for the approximate ancillary and the effect of curvature in the interest parameter; in each case for comparison purposes there is an exact ancillary that allows the computation of exact significance values.

Example 1. We first consider a scalar interest parameter α with no nuisance parameter in the context of an exact ancillary: Fisher's normal distribution on the circle. The model is $y_1 = \rho \cos \alpha + e_1, y_2 = \rho \sin \alpha + e_2$ where e_i are independent standard normal. We consider inference for α with ρ assumed known.

The normal distribution is centered on a circle of radius ρ on the plane. Let (c, a) be new variables defined by $y_1 = c \cos a, y_2 = c \sin a$. Then c has the noncentral chi distribution on 2 degrees of freedom with noncentrality parameter ρ , and $(a - \alpha)$ given c is von Mises with precision $\kappa = c\rho$. Fisher (1957) recommended using the conditional distribution given c , as c is ancillary. The approximate ancillary direction of Section 3 is

tangent to this exact ancillary at the data point.

The signed root of the profile log likelihood ratio for testing α is

$$r = \text{sgn}(a - \alpha)[2c\rho\{1 - \cos(a - \alpha)\}]^{1/2}$$

The maximum likelihood departure using the tangent exponential model is $M = \sin(a - \alpha)$, which gives the standardized departure $Q = M(c\rho)^{1/2}$. The third order significance level for testing α is $\Phi(r, Q)$.

The exact significance level for testing α is $G_{c\rho}(a - \alpha)$ where $G_\kappa(a)$ is the distribution function on $(-\pi, \pi)$ for the von Mises (κ) distribution with density

$$f(a) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos a)$$

where I_0 is the Bessel function of order zero.

Table 1 compares the approximation with the exact value for various values of $(c\rho)^{1/2}$. For larger values of this quantity the approximation is quite good but for smaller values it degrades near the endpoints $-\pi$ and π of the distribution. This degradation in the approximation was noted in a different context in Barndorff-Nielsen (1990) and is roughly attributable to the wrap-around nature of the support of the distribution in comparison with the full real line implicit in the asymptotic analysis.

Table 1: Values of the approximate formulas LR(2.3) and BN(2.6) in comparison with the exact distribution function for various angular departures $a - \alpha$.

$(c\rho)^{1/2} = 2$				
$a - \alpha$	-1.5	-1.0	-0.5	-0.2
LR(2.3)	0.004504	0.032188	0.169107	0.349455
BN(2.6)	0.004501	0.032185	0.169106	0.349455
Exact	0.004932	0.033226	0.170512	0.350225
$(c\rho)^{1/2} = 5$				
$a - \alpha$	-1.0	-0.5	-0.2	-0.1
LR(2.3)	$9.34879 * 10^{-7}$	0.006922	0.160277	0.309492
BN(2.6)	$9.34879 * 10^{-7}$	0.006922	0.160277	0.309492
Exact	$9.37596 * 10^{-7}$	0.006927	0.160300	0.309509
$(c\rho)^{1/2} = 8$				
$a - \alpha$	-0.5	-0.3	-0.2	-0.1
LR(2.3)	$3.87098 * 10^{-5}$	0.008509	0.055445	0.212405
BN(2.6)	$3.87098 * 10^{-5}$	0.008509	0.055445	0.212405
Exact	$3.90059 * 10^{-5}$	0.008510	0.055449	0.212409

Example 2. Consider again the standard normal distribution on the plane with mean $(\rho \cos \alpha, \rho \sin \alpha)$. We examine α as the interest parameter with nuisance parameter ρ .

With a specified value for α the mean lies on the ray $\{(\rho \cos \alpha, \rho \sin \alpha) : \rho > 0\}$ from the origin. As a manifold, this has a boundary point at the origin and most third order asymptotic methods do not address such singularities; we avoid this issue here by allowing ρ to be positive and negative.

Let $x_1 = y_1 \cos \alpha + y_2 \sin \alpha$, $x_2 = -y_1 \sin \alpha + y_2 \cos \alpha$. In terms of (x_1, x_2) we have that x_1 is sufficient for the nuisance parameter ρ and has the normal $(\rho, 1)$ distribution; and x_2 is ancillary and has the normal $(0, 1)$ distribution. The obvious assessment for departure from α , which is free of the nuisance parameter ρ , is to examine x_2 or x_2 given x_1 ; the ancillary construction of Section 3 reproduces this conditioning. Since x_2 has a normal $(0, 1)$ distribution, The exact significance for α is thus

$$p(\alpha) = \Phi(x_2) = \Phi(-y_1 \sin \alpha + y_2 \cos \alpha) . \quad (6.1)$$

and the approximation of Section 4 reproduces this distribution, as we have $r = Q = x_2$.

Note that the test procedure involves only the specified value for α and takes no account of the form of the interest parameter otherwise: it uses only the particular line through the origin at angle α and no information on the partition of the parameter space generated by other values of the interest parameter. This happens generally with the present third order methods.

Example 3. Consider again the standard normal distribution on the plane with mean $(\rho \cos \alpha, \rho \sin \alpha)$. We now examine ρ as interest parameter with α as nuisance parameter.

As in Example 1 let (c, a) be defined by $y_1 = c \cos a, y_2 = c \sin a$: c has the noncentral chi distribution with noncentrality parameter ρ , and the conditional distribution of $(a - \alpha)$ given c is Von Mises with $\kappa = c\rho$. We thus have a marginal variable c measuring departure for ρ and a conditional variable a given c that measures the nuisance parameter. The exact significance for ρ is thus

$$p(\rho) = G_\rho(c) \tag{6.2}$$

where G_ρ is the distribution function for the noncentral chi distribution (ρ) with 2 degrees of freedom.

The signed root of the profile log likelihood ratio for testing ρ is

$$r = c - \rho . \tag{6.3}$$

The maximum likelihood departure is $M = c - \rho$. The full information $J_2 = 1$ and the nuisance parameter information at $(\rho, \hat{\alpha}_\rho)$ is $c\rho$ which gives

$$J_1 = c\rho \cdot \rho^{-2} = c/\rho \tag{6.4}$$

when calibrated to the canonical parameterization. Thus

$$Q = (c - \rho) \left(\frac{\rho}{c}\right)^{1/2} . \tag{6.5}$$

The third order asymptotic significance level is then $\Phi(r, Q)$ using (2.3) or (2.6) with (6.3) and (6.5).

Table 2 compares the approximate $\Phi(r, Q)$ to the exact $G_\rho(c)$. It should be noted that this corresponds to examining the asymptotic method for an $n = 1$ case. The approximation is very good for $c > \rho$ but less accurate for small r near the centre of curvature.

Table 2: Values of the approximate and exact distribution function for c with a noncentral chi distribution with noncentrality parameter ρ .

$\rho = 2$				
c	1	3	5	8
LR(2.3)	0.087783	0.786963	0.997791	1.000000
BN(2.6)	0.087059	0.787352	0.997795	1.000000
Exact	0.081892	0.785638	0.997779	1.000000
$\rho = 5$				
c	3	6	8	11
LR(2.3)	0.016665	0.818249	0.998258	1.000000
BN(2.6)	0.016680	0.818282	0.998259	1.000000
Exact	0.016616	0.818150	0.998257	1.000000
$\rho = 8$				
c	6	9	11	14
LR(2.3)	0.019133	0.826666	0.998395	1.000000
BN(2.6)	0.019136	0.826675	0.998395	1.000000
Exact	0.019124	0.826640	0.998394	1.000000

Example 4. A nonlinear regression model discussed by Gallant (1975b, 1987) has the response function

$$\eta(\theta; x_1, x_2, x_3) = \theta_1 + \theta_2 x_1 + \theta_3 e^{\theta_4 x_2}$$

where x_1 is an indicator variable for group 1 or 2. In the treatment group ($x_1 = 1$), the

data for x_2 and y are respectively

$$x_2 = (6.28, 9.11, 8.11, 6.58, 6.52, 9.86, 0.47, 4.07, 0.17, 4.39, 4.73, 8.90, 0.77, 4.51, 0.08)$$

$$y = (0.98610, 0.95482, 1.02324, 0.96263, 0.98861, 0.98982, 0.66768, 0.96822, 0.59759, \\ 1.01962, 1.04255, 0.97526, 0.80219, 0.95196, 0.50811)$$

and in the control group ($x_1 = 0$) the data for x_2 and y are

$$x_2 = (9.86, 8.43, 1.82, 5.02, 3.75, 7.31, 0.07, 4.61, 6.99, 0.39, 9.42, 3.02, 3.31, 2.65, 6.11)$$

$$y = (1.03848, 1.04184, 0.90475, 1.05026, 1.03437, 1.01214, 0.55107, 0.98823, 0.99418, \\ 0.69163, 1.04343, 1.04969, 1.01046, 0.97658, 0.91840).$$

A familiar estimate of σ^2 is the mean square error, which in this example is $0.03049/26 = 0.00117$; for the present analysis we use a normal model for the errors and take the standard deviation to have the estimated value $0.00117^{1/2}$.

The maximum likelihood estimate obtained by Marquart's method is

$$\hat{\theta} = (1.01567, -0.02589, -0.05049, -1.11600).$$

A parameter of particular interest is the decay rate $\psi = \theta_4$ with respect to age. We calculate the third order significance based on the local ancillary developed in Section 3 and using the two approximation formulas LR(2.3) and BN(2.6). For comparison purposes we also obtain the two first order significance values, $\Phi(r)$ based on the signed profile, and $\Phi(u)$ based on the ordinary standardized maximum likelihood departure defined after (2.7). The third order analysis produces a standardized maximum likelihood pseudo parameter; this lead also to a first order significance $\Phi(Q)$. We record these values for several values of the interest parameter:

Table 3: First and third order approximations to the significance level for testing various values of ψ in Example 4.

ψ	$\Phi(r)$	$\Phi(q)$	$\Phi(Q)$	LR(2.3)	BN(2.6)
-1.50	0.980771	0.991610	0.980683	0.980750	0.980750
-1.45	0.966407	0.981254	0.969866	0.966870	0.966870
-0.90	0.067725	0.089753	0.070099	0.068789	0.068789
-0.80	0.010266	0.024732	0.011821	0.010544	0.010544

In a plot of $\Phi(r)$, $\Phi(q)$, $\Phi(Q)$, LR(2.3), and BN(2.6) against ψ the two third order formulas coincide to graphing accuracy, and $\Phi(r)$ and $\Phi(Q)$ are close to the two third order formulas.

Testing for ψ is of particular interest analytically. If ψ is specified the model becomes an ordinary linear model, and the distinguishing feature of the present method lies in some effective contribution of separating out one degree of freedom from error to measure the departure from the specified ψ value. Thus the contribution here is more to linear model theory than to nonlinear model theory.

Example 5. A kinetics problem (Box & Lucas, 1959; Guttman & Meeter, 1965; Gallant, 1987; Seber & Wild, 1989) has become very prominent in the nonlinear regression literature, indeed Gallant (1987) calls it the standard pedagogical example. The response function $\eta(\theta; x) = C(x)$ is the solution of the differential equations

$$\begin{aligned}\frac{d}{dx} A(x) &= -\theta_1 A(x) \\ \frac{d}{dx} B(x) &= -\theta_1 A(x) - \theta_2 B(x) \\ \frac{d}{dx} C(x) &= \theta_2 B(x)\end{aligned}$$

with boundary conditions $A(x) = B(x) = C(x) = 0$ at time $x = 0$, and with $\theta_1 > \theta_2 > 0$.

The response function is

$$\eta(\theta; x) = 1 - (\theta_1 - \theta_2)^{-1} \{ \theta_1 \exp(-\theta_2 x) - \theta_2 \exp(-\theta_1 x) \}$$

and is viewed (Guttman & Meeter, 1965) as being highly nonlinear.

We examine a dataset of size 12, simulated by Guttman & Meeter (1965) using normal error with standard deviation 0.025 and with $\theta = (1.4, 0.4)$:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.137790 & 0.409262 & 0.639014 & 0.736366 & 0.786320 & 0.893237 & \\ 1 & 2 & 3 & 4 & 5 & 6 & \\ 0.163208 & 0.372145 & 0.599155 & 0.749201 & 0.835155 & 0.905845 & \end{array}$$

We take θ_1 to be the interest parameter and θ_2 to be the nuisance parameter.

For a first analysis we use a normal model with the given value $\sigma_0 = 0.025$ for the error scaling. The modified Gauss-Newton method gives the maximum likelihood estimates $\hat{\lambda} = \hat{\theta}_2 = 0.456427$, $\hat{\psi} = \hat{\theta}_1 = 1.145483$. We record the third order significance LR(2.3) and BN(2.6) for four selected values of $\psi = \theta_1$. We also record the first order significance values $\Phi(r)$, $\Phi(q)$, $\Phi(Q)$ as discussed in Example 4.

Table 4: First and third order approximations to the significance level for testing various values of $\psi = \theta_1$ in Example 5.

ψ	$\Phi(r)$	$\Phi(q)$	$\Phi(Q)$	LR(2.3)	BN(2.6)
0.93	0.828671	0.826459	0.830596	0.830799	0.830799
0.94	0.817560	0.815042	0.819381	0.819768	0.819768
1.61	0.037121	0.021335	0.039354	0.037815	0.037815
1.62	0.034555	0.019200	0.036745	0.035211	0.035211

In a plot of $\Phi(r)$, $\Phi(q)$, $\Phi(Q)$, LR(2.3), and BN(2.6) against $\psi = \theta_1$ we find that the third order formulas coincide to graphing accuracy. A 95% confidence interval for ψ obtained from the third order analysis is (0.7313, 1.5703) which can be compared with the intervals (0.7263, 1.5678), (0.7685, 1.5225), (0.7388, 1.5748) from the first order approximations for r , q , and Q .

For a second analysis we use a Student (6) error distribution rescaled by the factor 1.0903 so that 68.26% of the probability lies in the interval $(-1, +1)$; the scaling factor $\sigma_0 =$

0.025 is then applied to the preceding standardized Student distribution. For background details see Fraser (1976): the corresponding analyses tend to be close to the normal analysis if the errors are normally distributed, and are correct if the errors are from the Student (6).

The modified Gauss-Newton method gives the maximum likelihood estimates as $\hat{\lambda} = \hat{\theta}_2 = 0.466659$, $\hat{\psi} = \hat{\theta}_1 = 1.100710$. We record significance values for selected ψ values

ψ	$\Phi(r)$	$\Phi(q)$	$\Phi(Q)$	LR(2.3)	BN(2.6)
0.93	0.781095	0.781381	0.782261	0.783025	0.783025
0.94	0.767934	0.767720	0.769051	0.770098	0.770098
1.61	0.023010	0.010234	0.040180	0.026863	0.026860
1.62	0.021343	0.009060	0.038367	0.025005	0.025002

In a plot of $\Phi(r)$, $\Phi(q)$, $\Phi(Q)$, LR(2.3), and BN(2.6) against $\psi = \theta_1$, the two third order formulas coincide to graphing accuracy. And with this non-normal analysis we find that $\Phi(r)$ and $\Phi(Q)$ are not particularly close to the two third order formulas. An interpolated 10% confidence upper bound for ψ from the third order analysis is 1.5219.

7. REFERENCES

- Amari, S. I. (1982). Differential geometry of curved exponential families, curvature and information loss. *Ann. Statist.* **10**, 357–385.
- Amemiya, T. (1983). Non-linear regression models. (Eds: Griliches, Z. and Intriligator, M.D.), *Handbook of Econometrics*, Vol. I. 333-389. North-Holland: Amsterdam.
- Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–65.
- Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized, signed log likelihood ratio. *Biometrika* **73**, 307–22.
- Barndorff-Nielsen, O.E. (1990). Approximate interval probabilities. *J. Royal Statist. Soc. B* **52**, 485–96.

- Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557–563.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Royal Statist. Soc. B* **41**, 279–312.
- Bates, D.M. and Watts, D.G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. Royal Statist. Soc. B* **42**, 1–25.
- Bates, D.M. & Watts, D.G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Statist.* **9**, 1152–1167.
- Beale, E.M.L. (1960). Confidence regions in non-linear estimation (with discussion). *J. Royal Statist. Soc. B* **22**, 88–115.
- Bor, M.J. (1971). Bias in nonlinear estimation (with discussion). *J. Royal Statist. Soc. B* **33**, 171–201.
- Box, G.E.P. and Locas, H.L. (1959). The design of experiments in nonlinear situations. *Biometrika* **46**, 77–90.
- Cook, R.D. and Witmer, J.A. (1985). A note on parameter effect curvature. *J. Amer. Statist. Assoc.* **88**, 872–878.
- Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–650.
- 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.
- Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3**, 1189–1242.
- Fisher, R.A. (1957). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.
- Fraser, D.A.S. (1964). Local conditional sufficiency. *J. Royal Statist. Soc. B* **26**, 52–62.

- Fraser, D.A.S. (1976). Necessary analysis and adaptive inference. *J. Amer. Statist. Assoc.* **71**, 99–113.
- Fraser, D.A.S. (1990). Tail probabilities from observed likelihood. *Biometrika* **77**, 65–76.
- Fraser, D.A.S. (1991). Statistical inference: likelihood to significance. *J. Amer. Statist. Assoc.* **86**, 258–65.
- Fraser, D.A.S. and Massam, H. (1985). Conical tests: observed levels of significance and confidence regions. *Statistical Papers* **26**, 1–18.
- Fraser, D.A.S. and Massam, H. (1987). Second order inference for generalized least squares. *Can. J. Statist.* **15**, 21–30.
- Fraser, D.A.S., Monette, G., Ng, K.W. & Wong, A. (1992). Higher order approximations with generalized linear models. *Proceedings of the Symposium on Multivariate Analysis, Hong Kong*, to appear.
- Fraser, D.A.S. and Reid, N. (1993). Simple asymptotic connections between densities and cumulant generating function leading to accurate approximations for distribution functions. *Statist. Sinica.* **3** 67–82.
- Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third order significance, *Utilitas Mathematica***47** 33–53.
- Gallant, A.R. (1975a). The power of the likelihood ratio test of location in nonlinear regression models. *J. Amer. Statist. Assoc.* **70**, 198–203.
- Gallant, A.R. (1975b). Nonlinear regression. *Amer. Statist.* **29**, 73–81.
- Gallant, A.R. (1987). *Nonlinear Statistical Models*. New York: Wiley.
- Gillis, P.R. and Ratkowsky, D.A. (1978). The behaviour of estimators of the parameters of various yield density relationships. *Biometrika* **34**, 191–198.
- Guttman, I. and Meeter, D.A. (1965). On Beale’s measure of non-linearity. *Technometrics* **7**, 623–637.
- Hamilton, D.C. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika* **73**, 57-64.

- Hamilton, D.C., Bates, D.M., and Watts, D.G (1982). Accounting for intrinsic nonlinearity in nonlinear regression parametric inference regions. *Ann. Statist.* **10**, 386–393.
- Jennrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633–643.
- Kass, R.E. (1984). Canonical parametrizations and zero parameter-effects curvature. *J. Royal Statist. Soc. B* **46**, 88–92.
- Lugannani, R. and Rice, S.O (1980). Saddlepoint approximation for the distribution of the sums of independent random variables. *Adv. Appl. Prob.* **12**, 475–490.
- Malinvaud, E. (1970). The consistency of nonlinear regression. *Ann. Math. Statist.* **41**, 956–969.
- Ratkowsky, D.A. (1983). *Nonlinear Regression Modelling*. New York: Marcel Dekker.
- Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3**, 213–237.
- Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*. New York: Wiley.