219v7:2002.07.05

# LIKELIHOOD FOR COMPONENT PARAMETERS

By D.A.S. FRASER

Department of Statistics, University of Toronto

100 St. George Street, Toronto, Canada M5S 3G3.

dfraser@utstat.toronto.edu

## SUMMARY

For a statistical model with data, likelihood for the scalar or vector full parameter $\theta$, of dimension $p$ say, is typically well defined and easily computed. In this paper, we investigate likelihood for a component parameter $\psi(\theta)$ of dimension $d < p$ and make use of the recent likelihood theory that has been successful in producing highly accurate third-order $p$-values for scalar parameters of continuous models. The theory leads under moderate regularity to a definitive third-order determination of likelihood for a component parameter $\psi(\theta)$ of dimension $d$, where $1 \leq d \leq p$. We use the simple location model on the plane with standard normal errors to motivate the development. The example exhibits most of the key characteristics of the general case and the recent theory then extends the determination of likelihood to the general context. For the scalar interest parameter case with $d = 1$, the usual determinations are typically of second-order accuracy; the example indicates how the new determination achieves third-order accuracy. The implementation is straightforward and uses familiar ingredients to other determinations, such as the full maximum likelihood value $\hat{\theta}$, the constrained value $\hat{\theta}_\psi$ given $\psi(\theta) = \psi$, and the observed information $j_{\lambda\lambda}(\hat{\theta}_\psi)$ for a complementing nuisance parameter $\lambda(\theta)$. It does however require a special version of the nuisance information $j_{(\lambda\lambda)}(\hat{\theta}_\psi)$, a version calibrated relative to a symmetric choice of the exponential-type reparameterisation $\varphi(\theta)$ underlying the recent theory, but this is easily computed. Various examples are given and the motivating example is discussed in detail.

**Some key words.** Ancillary; Conditional likelihood; Conditioning; Interest parameter; Likelihood; Marginal likelihood; Modified likelihood; $p$-value.

## 1. INTRODUCTION

For the full parameter $\theta$ of a standard statistical model, likelihood is well defined and describes how probability density at a data point of interest varies with different values for the parameter. For a component parameter, however, likelihood may not always be obviously defined, and various determinations for the scalar component parameter case have been proposed; these are discussed in §2. In this section we record the new determination (1.2) of likelihood for a component parameter $\psi(\theta)$ of dimension $d$ and give with some needed background information on the exponential parameterisation. The determination is derived in §3 using the simple symmetric normal on the plane as the motivating pattern. It has uniqueness properties at third-order accuracy, as discussed at point (xi) in §6. And it maintains this third-order accuracy in general contexts where other determinations drop to second-order accuracy due to inappropriate target likelihoods: see the final two paragraphs in §4.

Consider a continous statistical model $f(y;\theta)$ with $p$-dimensional full parameter and $d$-dimensional interest parameter $\psi(\theta)$; let $\ell(\theta;y)$ be the loglikelihood for the full model. Formulae are simpler if we have available a complementing nuisance parameterisation $\lambda(\theta)$ so that $\theta$ is one-one equivalent to $(\psi, \lambda)$. For given data $y$ let $\hat{\theta}$ be the overall maximum likelihood value, let $\hat{\theta}_\psi$ be the constrained value given $\psi(\theta) = \psi$, and let $j_{\lambda\lambda}(\hat{\theta}_\psi)$ be the observed nuisance information obtained by evaluating the matrix $-\ell_{\lambda\lambda}(\psi, \lambda; y) = -(\partial/\partial\lambda)(\partial/\partial\lambda')\ell(\psi, \lambda; y)$ at the constrained maximum likelihood value $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$. For any likelihood assessment of a component $\psi(\theta)$ the profile loglikelihood function,

$$\ell_{\mathrm{P}}(\psi) = \ell(\hat{\theta}_\psi) = \ell(\psi, \hat{\lambda}_\psi), \tag{1.1}$$

can be viewed as a necessary first step; it has first-order properties based on asymptotic normality of the maximum likelihood estimate.

Recent likelihood theory either explicitly or implicitly makes use of an exponential type reparameterisation $\varphi(\theta)$; some background on this is recorded now and more in §6. The third-order determination of loglikelihood for $\psi(\theta)$ is then given as

$$\begin{aligned} \ell^*(\psi) &= \ell_{\mathrm{P}}(\psi) + \frac{1}{2}\log|j_{(\lambda\lambda)}(\hat{\theta}_\psi)|, \\ &= \ell_{\mathrm{P}}(\psi) + \frac{1}{2}\log|j_{\lambda\lambda}\hat{\theta}_\psi| - \log|\partial(\lambda)/\partial\lambda|_{\hat{\lambda}_\psi}, \end{aligned} \tag{1.2}$$

where the asterisk indicates the third- order property and the parentheses enclosing $\lambda$ are to indicate that the nuisance parameter has been re-expressed in terms of a special symmetric choice $\bar{\varphi}(\theta) = \hat{\jmath}^{1/2}_{\varphi\varphi}\varphi(\theta)$ of exponential parameterisation that has $\hat{\jmath}_{\bar{\varphi}\bar{\varphi}} = I$. The reparameterisation affects the nuisance information only through the derivative $\partial(\lambda)/\partial\lambda$ at $\hat{\lambda}_\psi$ and the effect is available from the $p \times d$ Jacobian $X = \bar{\varphi}_{\lambda'} = \partial\bar{\varphi}(\theta)/\partial\lambda' = \hat{\jmath}^{1/2}_{\varphi\varphi}\partial\varphi/\partial\lambda'$ of $\bar{\varphi}$ with respect to $\lambda'$ evaluated at the observed $\hat{\theta}^0_\psi$. The derivative $|\partial(\lambda)/\partial\lambda|_{\hat{\lambda}_\psi}$ is then available as $|X| = |X'X|^{1/2}$, where $X'X = \varphi'_\lambda(\hat{\theta}_\psi)\hat{\jmath}_{\varphi\varphi}\varphi_{\lambda'}(\hat{\theta}_\psi)$, and the re-expressed nuisance information is given as

$$|\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)| = |\jmath_{\lambda\lambda}(\hat{\theta}_\psi)||\varphi'_\lambda(\hat{\theta}_\psi)\hat{\jmath}_{\varphi\varphi}\varphi_{\lambda'}(\hat{\theta}_\psi)|^{-1}. \tag{1.3}$$

For notation we write for example $\varphi'_\lambda = (\varphi_{\lambda'})'$ as the Jacobian of the row vector $\varphi'$ with respect to the column vector $\lambda$.

The recent likelihood theory gives an essentially unique third-order probability differential for assessing the parameter $\psi(\theta)$. We find however that the probability differentials for various $\psi$ values come in fact from different distributions. This is associated with long recognised difficulties and is seen clearly in the example in §3. Our approach is to show that an appropriate transformation of the initial variable produces in fact a common reference distribution for the various probability differentials. This then gives the loglikelihood (1.2) for $\psi$. Details are discussed in §3 and point (x) in §6.

As background we mention that an exponential parameterisation $\varphi(\theta)$ allows a statistical model to be treated to third order as a full exponential model with canonical parameter $\varphi(\theta)$. Such models admit a wide range of third-order inference methods, in particular $p$-value formulae for scalar parameters (Barndorff-Nielsen, 1986; Fraser, 1990; Fraser & Reid, 1995; Fraser, Reid & Wu, 1999) where the last reference is specifically organised to present the inference methods in terms of the related exponential model.

An exponential parameter is obtained as the gradient of loglikelihood,

$$\varphi'(\theta) = \frac{d}{dV}\ell(\theta; y)\big|_{y^0} = \ell_{;V}(\theta; y^0) \ , \tag{1.4}$$

calculated in $p$ linearly independent directions $(v_1, \ldots, v_p) = V$ tangent to an approximate or exact ancillary at the observed data $y^0$. The derivative in a direction $v_i$ is given as $(d/dv_i)\ell(\theta; y)\big|_{y^0} = (d/dx)\ell(\theta; y^0 + xv_i)\big|_{x=0}$ and the chain rule can be used to give

$$\varphi'(\theta) = \sum_j \ell_{;y_j}(\theta; y^0)(v_{j1} \cdots v_{jp}) = \ell_{;y'}(\theta; y^0)V; \tag{1.5}$$

in terms of coordinate by coordinate derivatives of likelihood.

In the presence of an exact or approximate ancillary $a(y)$ Barndorff-Nielsen (1986) uses in effect

$$\varphi'(\theta) = \frac{\partial}{\partial\hat\theta'}\ell(\theta;\hat\theta^0,a^0); = \ell_{;\hat\theta'}(\theta;\hat\theta^0,a^0) \tag{1.6}$$

as an implicit reparameterisation for the development of $p$-values.

For any given parameterisation say $\theta$ we have a corresponding change of variable from $y$ to $(\hat\theta, a)$ and then at any data point $y^0$ we can find vectors $V$ tangent to the ancillary surface that provide a local basis for the coordinates $\hat\theta$. Similarly for any tangent vectors $V$ at a data point there exists a parameterisation, $\theta$ say, so that $\hat\theta$ provides coordinates locally with respect to the tangent vectors $V$. This gives the equivalence of (1.4) and (1.6), and it follows that they can be used interchangeably, provided the $V$ and the $\theta$ are understood to be in correspondence; in particular we would then have $\ell_{;V}(\theta;y^0) = \ell_{;\hat\theta'}(\theta;y^0)$. Also if $\theta$ is rescaled to have an identity information at the data point we would obtain

$$|X| = |\ell_{\lambda;\hat\theta}(\hat\theta_\psi;y^0)| = |\tilde\ell_{\lambda;\hat\theta'}|, \tag{1.7}$$

where the tilde denotes evaluation at $\hat\theta_\psi$. As part of this we have that any full affine function of $\varphi$ is an equivalent exponential parameterisation and can be obtained by either of the routes (1.4) and (1.6). Details concerning the equivalence of third-order $p$-value formulae based on the two parameterisations (1.4) and (1.6) may be found in §2.3 of Fraser, Reid & Wu (1999), where it is seen that an affine change in $\varphi$ has direct compensatation within the formulae.

Tangent vectors $V$ for an approximate ancillary are available (6.1) in wide generality. They can be derived from a full-dimensional pivotal quantity and some details are recorded at point (ii) in §6. For certain specialised models, however, such as the location-scale and transformation models, an exact ancillary $a(y)$ of full dimension $n-p$ may be available and the change of variable from $y$ to $(\hat{\theta}, a)$ may be feasible. In such cases the exponential parameter version (1.6) may be more convenient but the tangent vectors $V$ are also readily available and thus permit the alternative use of (1.4); for details concerning the calculation of vectors $V$ from an available ancillary see point (ii) in §6. For the general cases, however, where the change of variable is not easy or feasible or in the typical absence of an explicit ancillary it is straightforward, see (6.1), to calculate $V$ and to obtain the exponential parameter using (1.4).

The new likelihood determination is presented in terms of continuous statistical models but the results apply equally for full exponential models where $\varphi(\theta)$ is taken to be the corresponding canonical parameter.

## 2. BACKGROUND: LIKELIHOOD FOR A COMPONENT

Consider a parameter component $\psi(\theta)$, where for notational convenience we assume that $\theta' = (\psi', \lambda')$. The maximum likelihood value $\hat{\psi}$ can have serious bias problems when the dimension of the nuisance parameter $\lambda$ is large. This was made prominent by Neyman & Scott (1948) and in turn points to general difficulties with the direct use of the profile likelihood (1.1) for inference. Ongoing efforts have focussed on finding corrections or improvements to profile likelihood, to make it more nearly a true likelihood in the sense of describing probability at an observed value.

Transformation methods were used in Fraser (1967, 1968, 1972) to obtain determinations of marginal likelihood for special parameters in transformation models; properties and comparison of these are discussed in Skovgaard et al. (2001). Factorisation techniques were used in Kalbfleisch & Sprott (1973) to obtain marginal likelihoods for parameters. In most cases it is relatively straightforward to obtain a probability differential involving only the interest parameter but a difficulty arises in finding an appropriate support differential that is free of the parameter being tested; the present approach standardises with respect to the corresponding reference distribution and resolves this difficulty.

Barndorff-Nielsen (1983) used asymptotic properties to obtain determinations of likelihood for a scalar component $\psi(\theta)$. The original derivation worked from the assumed presence of a variable $t$ whose marginal distribution depends only on $\psi$ and is thereby free of the nuisance parameter $\lambda$. The resulting modified profile loglikelihood is

$$\begin{aligned}\ell_{\mathrm{BN}}(\psi) &= \ell_P(\psi) - \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat\theta_\psi)| + \log|\partial\hat\lambda/\partial\hat\lambda_\psi| \\ &= \ell_P(\psi) + \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat\theta_\psi)| - \log|\ell_{\lambda;\hat\lambda}(\hat\theta_\psi)|, \end{aligned} \tag{2.1}$$

where $\partial\hat\lambda/\partial\hat\lambda_\psi$ is taken for fixed $\hat\psi$ and the second expression is obtained by differentiating $\ell_\lambda(\psi,\hat\lambda_\psi;\hat\psi,\hat\lambda)=0$ with respect to $\hat\lambda$ holding $\hat\psi$ fixed. Barndorff-Nielsen (1983, 1985, 1987) subsequently gave various conditional derivations of (2.1). For cases where an exact or approximate ancillary conditioning is not readily accessible, various estimation procedures for the last term in (2.1) have been proposed; see for example §8.3 in Barndorff-Nielsen and Cox (1994).

By contrast, Cox & Reid (1987) worked from the assumed presence of a variable $a$ for which the conditional distribution depends only on $\psi$; the conditional distribution is then used to obtain a conditional loglikelihood for $\psi$,

$$\ell_{\mathrm{CR}}(\psi) = \ell_p(\psi) - \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat\theta_\psi)|, \tag{2.2}$$

where the parameter $\lambda$ is required to be orthogonal to $\psi$.

Recent likelihood theory gives third-order $p$-values for testing scalar interest parameters $\psi(\theta)=\psi$. The $p$-value $p(\psi)$ is obtained from the signed likelihood root

$$r=\operatorname{sgn}(\hat{\psi}-\psi)\left[2\{\ell(\hat{\theta})-\ell(\hat{\theta}_{\psi})\}\right]^{1/2} \tag{2.3}$$

and an appropriate maximum likelihood departure $Q$; these are combined using the Lugannani & Rice (1980) formula or the Barndorff-Nielsen (1986) formula

$$p(\psi)=\Phi(r^*)=\Phi\{r-r^{-1}\log(r/Q)\}\ , \tag{2.4}$$

where $\Phi$ is the standard normal distribution function. From the latter, Barndorff-Nielsen (1994) proposed an extended likelihood for $\psi$,

$$\ell_{\mathrm{EL}}(\psi)=-\frac{1}{2}(r^*)^2+\log|\partial r^*/\partial\hat{\psi}|, \tag{2.5}$$

where the partial derivative is taken for fixed ancillary and for fixed $\hat{\lambda}_{\psi}$; this treats $r^*$ as a variable with distribution on the contour with $\hat{\lambda}_{\psi}$ fixed.

For these determinations the variables $t$ and $a$ are assumed to be free of $\psi$. In many applications however the possible candidates for such variables do in fact depend on $\psi$; this difficulty arises even for the $N(\mu,\sigma^2)$ with interest in $\mu$, where the familiar departure measure $t$ involves $\mu$. Fraser & Reid (1989) adjusted for this dependence on the interest parameter by using a constant information metric as the support for calculating probability density at the observed data. The development assumed the initial variable was sufficient and of the same dimension as the full parameter, and the computations required explicit expressions for the particular conditioning or marginalising variable. The use here of the exponential parameterisation by-passes these restrictions.

For the special cases where the appropriate conditioning or marginal variable does not dependent on the parameter, we have that the determinations (2.1), (2.2) and (2.5) are all third-order accurate. More generally however the appropriate variable does depend on the parameter, the difficulty does arise and the accuracy typically falls to second order; this is discussed in §4 where examples show how the new determination (1.2) attains third-order accuracy.

An extensive literature discusses relationships among these likelihood determinations and provides other modifications for likelihood for component parameters; see for example McCullagh & Tibshirani (1980), Fraser & Reid (1989), Barndorff-Nielsen & McCullagh (1993), Severini (1998), Skovgaard et al. (2001) and references therein.

### 3. MOTIVATING EXAMPLE AND DERIVATION

Consider the normal distribution on the plane for the variable $y = (y_1, y_2)'$ with mean $\theta = (\theta_1, \theta_2)'$ and rotationally symmetric error. For asymptotic properties we take the variance to be of order $O(n^{-1})$ or equivalently we take $\theta$ to be at a distance of order $O(n^{1/2})$ from the origin; however, for simplicity of presentation here we work with $n = 1$. We use polar coordinates and write $y = (r \cos a, r \sin a)'$ and $\theta = (\rho \cos \alpha, \rho \sin \alpha)'$, and take $\alpha$ to be the interest parameter and $\rho$ the nuisance parameter; for the case with $\rho$ as interest parameter the difficulty does not arise. A hypothesised value for $\alpha$ puts the mean of the distribution on a line $\mathcal{L}_\alpha$ through the origin with angle $\alpha$ to the first axis. For coordinates rotated through the angle $\alpha$ we have $(\hat{\rho}_\alpha, d_\alpha)$, where

$$\begin{aligned} \hat{\rho}_\alpha &= y_1 \cos \alpha + y_2 \sin \alpha = r \cos(\hat{\alpha} - \alpha), \\ d_\alpha &= -y_1 \sin \alpha + y_2 \cos \alpha = r \sin(\hat{\alpha} - \alpha) \end{aligned} \tag{3.1}$$

are respectively $N(\rho, 1)$ and $N(0, 1)$. We see that $d_\alpha$ records directed distance from the hypothesised line and with moderate continuity is essentially the unique variable with distribution free of the nuisance parameter $\rho$. The probability element for $d_\alpha$ is $(2\pi)^{-1/2}\exp(-d_\alpha^2/2)dd_\alpha$ which gives the loglikelihood

$$\ell(\alpha) = -\frac{1}{2}d_\alpha^2 = -\frac{1}{2}r^2\sin^2(\hat{\alpha}-\alpha) \tag{3.2}$$

as the direct assessment of $\alpha$ from data $(y_1, y_2)$; see Fieller(1954).

Now suppose we allow a normal error distribution with variance matrix $Cn^{-1}$, and with no loss of generality take $|C| = 1$ and for simplicity again take $n = 1$. This can be viewed as a linearly transformed version of the previous symmetric normal model and the structure of the parameter $\alpha$ is respected. For testing $\alpha$ we have as before that $d_\alpha$ in (3.1) has uniqueness as a variable with distribution free of $\rho$; the probability element for $d_\alpha$ is now $(2\pi\sigma_\alpha^2)^{-1/2}\exp(-d_\alpha^2/2\sigma_\alpha^2)dd_\alpha$, where

$$\sigma_\alpha^2 = c_{11}\sin^2\alpha - 2c_{12}\sin\alpha\cos\alpha + c_{22}\cos^2\alpha.$$

We thus see that the distribution for evaluating an observed $d_\alpha$ changes when $\alpha$ changes. In order to assess properly the probability differential for $d_\alpha$ we need to relate it to a distribution that does not change with $\alpha$. This can be accomplished by standardising the reference distribution or equivalently by using $d_\alpha/\sigma_\alpha$ as the departure measure. This gives the loglikelihood

$$\ell(\alpha) = -\frac{1}{2}\frac{d_\alpha^2}{\sigma_\alpha^2} \tag{3.3}$$

for $\alpha$. This is of course just the appropriately re-expressed version of the likelihood (3.2) as obtained with the initial symmetric version of the problem; in effect we are acknowledging sample space invariance. Also these loglikelihoods coincide with the loglikelihood $\ell^*(\alpha)$ given by (1.2), and provide the arguably definitive determination of likelihood for this normal location context.

Next suppose that $\psi$ specifies a $d$-dimensional subspace through the origin. The preceding arguments for the rotationally symmetric normal remain valid and $d_\psi$ becomes a vector of $p-d$ independent standard normal departures. This gives the loglikelihood $-\frac{1}{2}|d_\psi|^2$ with a $(p-d)$-dimensional standard normal as reference distribution; this again coincides with the general determination (1.2).

For the general case, we have that the reparameterisation $\varphi(\theta)$ defines the exponential model that produces general third order inference at the data point. For assessing $\psi(\theta)=\psi$ the departure is recorded essentially uniquely on $\hat{\theta}_\psi = \hat{\theta}^0_\psi$ which is a $d$-dimensional plane in terms of canonical coordinates. The observed log-density is recorded as (6.3) in Fraser & Reid (1995) and it gives the likelihood $\ell_{\rm P}(\psi)+\frac{1}{2}\log|j_{\lambda\lambda}(\hat{\theta}_\psi)|$, where here we take the nuisance information to be calibrated just in terms of the given $\varphi$. We see, however, that the observed probability differentials for different $\psi$ values come typically from differently scaled statistical models, this being the substance of the difficulty mentioned in §2 and noted for the skewed normal earlier in this section. If we then replace the initial exponential parameterisation by the symmetric version $\bar{\varphi}(\theta)$, we find that the reference statistical models are in fact third-order equivalent as expressed in terms of the score variables; some technical details are discussed in points (iv)-(x) in §6. The observed likelihood for $\psi$ is then given by (1.2) using the symmetrised $\bar{\varphi}$ and is seen to be coming from a common model to third order as $\psi$ varies.

### 4. COMPARISON OF LIKELIHOOD DETERMINATIONS

We now compare the likelihood determinations $\ell^*(\psi), \ell_{\rm BN}$ and $\ell_{\rm EL}(\psi)$ given by (1.2), (2.1) and (2.5). Each uses in effect the same third-order calculation of probability differential at the observed data and each is parameterisation invariant. The difficulty that we discussed earlier arises in the search for a support differential that is appropriately free of the parameter $\psi$. The new determination $\ell^*(\psi)$ is obtained by requiring the same reference statistical model for each of the various probability differentials for testing the $\psi$ values. This was motivated in §2 by examining the symmetric location normal on the plane and finding that the Fieller likelihood was the correct likelihood; the equivalent normal skewed model then demonstrated the need to standardize the likelihood differential against a common reference model.

For the present comparisons we continue with the rotationally symmetric normal and then indicate in §6 how the general case after standardisation has the same rotational symmetry with third-order formulas carrying the minor extension from rotational normal form to general rotational form. The new determination (1.2) gives the Fieller loglikelihood $\ell^*(\alpha) = -(1/2)d_\alpha^2/\sigma_\alpha^2$.

The determination (2.1) for testing the angle $\alpha$ has $j_{\rho\rho}(\hat{\theta}_\alpha) = 1$ and $\partial\hat{\rho}_\alpha/\partial\hat{\rho} = \cos(\hat{\alpha} - \alpha)$ for given $\hat{\alpha}$; thus

$$\ell_{\text{BN}} = \ell^*(\alpha) - \log\cos(\hat{\alpha} - \alpha). \tag{4.1}$$

The extra term is $O(n^{-1})$ and applies an $O(n^{-1})$ quadratic flattening $(\hat{\alpha} - \alpha)^2/2$ to the third-order likelihood. This flattening is in sharp disagreement with the Fieller solution.

The determination (2.5) uses $r^*$, which for the symmetric normal is just $d_\alpha = r\sin(\hat{\alpha} - \alpha)$, in agreement with the Fieller solution. We then have that $|\partial d_\alpha/\partial\hat{\alpha}| = r/\cos(\hat{\alpha} - \alpha)$ when calculated for fixed $r\cos(\hat{\alpha} - \alpha)$. This gives

$$\ell_{\text{EL}}(\alpha) = \ell^*(\alpha) - \log\cos(\hat{\alpha} - \alpha), \tag{4.2}$$

which has the same $O(n^{-1})$ quadratic flattening $(\hat{\alpha} - \alpha)^2/2$ added to the third-order likelihood. We also see from the extra term in (4.2) that the $r^*$ probability element is not being assessed against the appropriate Euclidean element supporting the standard normal density.

Now consider the comparisons more generally. For this it is convenient to work with a version of $\theta' = (\psi', \lambda')'$ that has identity information $\hat{\jmath}_{\theta\theta} = I$ at the data point and then let $\varphi(\theta)$ be the corresponding exponential parameterisation (1.6); if follows that $\theta$ and $\varphi$ are equivalent to first derivative at the observed maximum likelihood value. Simple calculations from (1.2) and (1.7) then give

$$\begin{aligned}\ell^*(\psi) &= \ell_P(\psi) + \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat{\theta}_\psi)| - \log|\tilde{\ell}_{\lambda;\hat{\theta}'}| \\ &= \ell_P(\psi) + \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat{\theta}_\psi)| - \frac{1}{2}\log|\tilde{\ell}_{\lambda;\hat{\theta}'}\tilde{\ell}'_{\lambda;\hat{\theta}'}| \\ &= \ell_P(\psi) + \frac{1}{2}\log|\jmath_{\lambda\lambda}(\hat{\theta}_\psi)| - \frac{1}{2}\log|\tilde{\ell}_{\lambda;\hat{\psi}'}\tilde{\ell}'_{\lambda;\hat{\psi}'} + \tilde{\ell}_{\lambda;\hat{\lambda}'}\tilde{\ell}'_{\lambda;\hat{\lambda}'}|.\end{aligned}$$

It follows that

$$\ell_{\mathrm{BN}} = \ell^*(\alpha) - \frac{1}{2}\log\frac{|\tilde{\ell}_{\lambda;\hat{\lambda}'}\tilde{\ell}'_{\lambda;\hat{\lambda}'}|}{|\tilde{\ell}_{\lambda;\hat{\theta}'}\tilde{\ell}'_{\lambda;\hat{\theta}'}|},$$

which presents the departure from third order again as $-\log\cos$ of an angle between two subspaces coresponding to given $\hat{\psi}$ and given $\psi$.

For the extended likelihood (2.5) we have in a similar manner

$$\ell_{\mathrm{EL}}(\alpha) = \ell^*(\alpha) + \log|\frac{\partial(\psi)}{\partial\psi}|_{\hat{\psi}}, \tag{4.3}$$

where the partial derivative is taken for fixed ancillary and for fixed $\hat{\lambda}_\psi$ and $(\psi)$ is $\psi$ recalibrated for fixed $\hat{\lambda}_\psi$ in terms of the special symmetric exponenential parameterisation. As $\theta$ and the special $\varphi$ coincide to first derivative at the maximum likelihood value we have that any second derivative difference provides the departure of (2.5) from the Fieller solution. Also, by reversing (4.3) we see that (2.5) can be modified towards $(\hat{\psi})$ rather than $\hat{\psi}$ to give third-order accuracy:

$$\ell^*_{\mathrm{EL}}(\psi) = -\frac{1}{2}(r^*)^2 + \log|\partial r^*/\partial(\hat{\psi})|. \tag{4.4}$$

Barndorff-Nielsen & Cox (1994, p.267) discuss the modified profile likelihood (2.1) in relation to three possible factorisations to isolate $\psi$ from the model density $f(\hat{\psi}, \hat{\lambda}; \psi, \lambda)$ conditional on an appropriate ancillary: (a) a marginal factorisation $f(\hat{\psi}; \psi)g(\hat{\lambda}|\hat{\psi}; \psi, \lambda)$ as in the original derivation for (2.1), (b) a conditional factorisation $g(\hat{\lambda}; \psi, \lambda)f(\hat{\psi}|\hat{\lambda}; \psi)$, (c) a generalised conditional factorisation $g(\hat{\lambda}_\psi; \psi, \lambda)f(\hat{\psi}|\hat{\lambda}_\psi; \psi)$, where in each case the $f$ factor is used for the sought-after likelihood. They note that "... in all three of the above cases the modified profile likelihood approximates the target likelihood to $O(n^{-3/2})$ in regions of the form $\{\psi : |\psi - \hat{\psi}| \leq cn^{-1/2}\}$". The order of the preceding approximations seems not to be at issue but the appropriate target likelihood is.

For this is suffices to examine the rotationally symmetric normal with the Fieller (1954) angle $\alpha$, as this presents all the complications of the general asymptotic case. The factorisations (a) and (b) are not available for assessing the angle $\alpha$ but a version of (c) is available. The conditional distribution given $\hat{\lambda}_\psi$ or specifically given $\hat{\rho}_\alpha = r\cos(\hat{\alpha} - \alpha)$ is $N(0,1)$ in the Euclidean coordinates of the problem. But the conditional distribution using the indicated $\hat{\alpha}$ coordinates has probability element

$$(2\pi)^{-1/2}e^{-d_\alpha^2/2}r\cos^{-1}(\hat{\alpha} - \alpha)d\hat{\alpha}$$

which gives the target log-likelihood $-d_\alpha^2/2 - \log\cos(\hat{\alpha} - \alpha)$ and coincides with (4.1) but is not in agreement with Fieller. Thus we see that the target likelihood based on the factorisation (c) is inappropriate and leads to the $O(n^{-1})$ discrepancy.

## 5. SOME EXAMPLES OF COMPONENT LIKELIHOOD

**Example 1. Normal location-scale model.** Consider the simple case of a sample $y_1, \ldots, y_n$ from the $N(\mu, \sigma^2)$ distribution and suppose we are interested in the likelihood for the component parameter $\mu$. Almost any set $V = (v_1, v_2)$ of linearly independent vectors not tangent to the maximum likelihood surface $\hat{\theta} = \hat{\theta}^0$ will extract a version of the exponential parameter, but a familiar version is available immediately as the model itself is a (2.2) exponential model with canonical parameter $\varphi = (\mu/\sigma^2, 1/\sigma^2)'$. The profile likelihood is

$$L_{\mathrm{P}}(\mu) = \left(\frac{\hat{\sigma}_\mu^2}{\hat{\sigma}^2}\right)^{-\frac{n}{2}} = (1+T^2)^{-\frac{n}{2}}, \tag{5.1}$$

where $T^2$ is the ratio of the usual sum of squares with $s^2 = \Sigma(y_i - \bar{y})^2$ and $T = \sqrt{n}(\bar{y} - \mu)/s = (n-1)^{-1/2}t$ giving the connection to the usual $t$ quantity. The quantity $T$ is $O(n^{-1/2})$ but does provide simpler expressions as with (5.1). We also see that (5.1) is proportional to the usual Student density function when $T$ is expressed in terms of $t$.

The model has location-scale properties and thus it suffices to use a special data point having $\bar{y} = 0$ and $s^2 = \Sigma(y_i - \bar{y})^2 = n$ and yet obtain general results. This choice gives $\hat{\sigma}^2 = 1$ and $\hat{\sigma}_\mu^2 = (1+T^2) = 1 + \mu^2$ with $\mu = O(n^{-1/2})$ for the moderate deviations range. The observed information for $\varphi$ is

$$\hat{\jmath}_{\varphi\varphi} = \begin{pmatrix} n & 0 \\ 0 & n/2 \end{pmatrix},$$

and the observed information for the modified $\bar{\varphi} = (\sqrt{n}\mu/\sigma^2, \sqrt{n}/\sqrt{2}\sigma^2)'$ is equal to the identity. The constrained nuisance information is $\jmath_{\sigma^2\sigma^2}(\hat{\theta}_\mu) = n/2\hat{\sigma}_\mu^4$ and the recalibrated nuisance information is $\jmath_{(\sigma^2\sigma^2)}(\hat{\theta}_\mu) = (1+T^2)^2/(1+2\mu^2) = 1+O(n^{-2})$. This gives $\ell^*(\mu) = \ell_{\mathrm{P}}(\mu)$, which is the profile (5.1) and is the very natural $t(n-1)$ density expression that describes the location $\mu$.

The modified profile likelihood (2.1), the approximate conditional likelihood (2.2) and the extended likelihood (2.5) all lead to $(1+T^2)^{-(n-2)/2}$, which corresponds to a $t(n-3)$ density for $t$; in a certain technical sense this underestimates the available precision by 2 degrees of freedom and corresponds to the quadratic flattening mentioned in the paragraphs after (4.1) and (4.2).

**Example 5.2. Normal regression model.** Consider the normal regression model $y = X\beta + \sigma e$, where $e$ is a sample from the $N(0,1)$ distribution, and suppose that $\beta_1$ is the parameter of interest. We assume that $\beta$ has dimension $r$ and thus $p = r + 1$. The details follow closely those for the preceding example and it suffices to use a canonical version of the design matrix with say $X = (I\ 0)'$. Let $s^2 = \Sigma(y_i - \hat{y}_i)^2 = \Sigma y_{r+i}^2$ and let $T^2 = (y_1 - \beta_1)^2/s^2$, which is the ratio of the usual sums of squares; then

$$\hat{\sigma}^2 = \frac{s^2}{n}\ , \quad \hat{\sigma}^2_{\beta_1} = \frac{s^2}{n}(1+T^2)\ , \quad \frac{\hat{\sigma}^2_{\beta_1}}{\hat{\sigma}^2} = 1 + T^2\ .$$

Again because of invariance properties it suffices to take a special point with $y_1 = 0$ and $s^2 = n$, which give $\hat{\sigma}^2 = 1$ and $\hat{\sigma}^2_{\beta_1} = (1 + T^2)$. The profile likelihood is $L_{\mathrm{P}}(\beta_1) = (1 + T^2)^{-\frac{n}{2}}$. The nuisance parameter can be taken equal to $\lambda' = (\beta_2, \ldots, \beta_r, \sigma^2)$ with information determinant $|\jmath_{\lambda\lambda}(\hat{\theta}_{\beta_1})| = n/2(\hat{\sigma}^2_{\beta_1})^{r-1+2}$ and recalibrated information determinant $|\jmath_{(\lambda\lambda)}(\hat{\theta}_{\beta_1})| = (\hat{\sigma}^2_{\beta_1})^{(r-1)+2}(1+2\beta_1^2/n)^{-1}$. The likelihood for $\beta_1$ is then obtained from (1.2) using (1.3): $L^*(\beta_1) = (1 + T^2)^{-\{n-(r-1)\}/2}$ to the third order, where we have again used the simplification $(1+T^2)^2(1+2T^2)^{-1} = 1 + O(n^{-2})$ . The new likelihood corresponds to the seemingly appropriate $t(n-r)$ density. For comparison we note that the modified profile likelihood (2.1) and the extended likelihood (2.5) correspond to the $t(n - r - 2)$ density and thus have the familiar $O(n^{-1})$ quadratic flattening. The avoidance of the 2-degrees-of-freedom shortfall is due partly to the use of the canonical parameterisation and partly to the symmetrisation of the observed information for the chosen canonical parameterisation.

Now suppose the interest parameter is $\beta_{(1)} = (\beta_1, \ldots, \beta_d)'$. Again we let

$s^2 = \Sigma(y_i - \hat{y}_i)^2$ and have $T^2 = \sum_1^d (y_i - \beta_i)^2/s^2$, which is the ratio of the usual sums of squares for testing the hypothesis concerning the interest parameter. Calculations as above then lead to the third-order likelihood $L^*(\beta_{(1)}) = (1+T^2)^{-\{n-(r-d)\}/2}$. This corresponds to the rather natural multivariate Student density with the degrees of freedom equal to that for the corresponding multivariate $t$-type test statistic.

As noted in §2 the difficulty in determining likelihood for components arises when the apparent support differential involves the interest parameter as in the simple normal Example 1, the underlyng structural feature being the rotation of the line $\mathcal{L}_\alpha$ with change in the interest parameter $\alpha$. In an exponential model context this commonly occurs when the parameter of interest is a ratio of canonical parameters. A similar phenomenon arises in the general case as indicated in §4 and point (vi) of §6. Without such rotation of the interest parameter on the canonical parameter space the difficulties with the various determinations seem not to arise

## 6. RECENT THEORY AND THE DERIVATION

Recent likelihood theory has worked primarily with continuous statistical models and has produced highly accurate third-order $p$-values for scalar parameter components. We summarise some relevant results from this theory.

(i) *Reduction by ancillarity.* Asymptotic theory has elicited the patterns in an asymptotic model and produced a reduction from an initial variable of dimension $n$ to a conditioned variable of the same dimension $p$ as the parameter; the conditioning uses an exact or approximate ancillary, $a(y)$ say, of dimension $n-p$ (Fraser & Reid, 1995, 2000). In the special cases where a full sufficient statistic is available, the conditioning duplicates the more familiar sufficiency result (Fraser, 2002).

(ii) *Tangents to the ancillary.* In the presence of independent scalar coordinates or in the presence of independent vector coordinates each with an associated pivotal quantity, reasonable continuity shows that conditioning is essentially unique for third-order inference and that only the observed likelihood $\ell^0(\theta)$ and the observed likelihood gradient $\varphi(\theta)$ from (1.4) or (1.5) are needed for third-order accuracy. In the presence of a full $n$-dimensional pivotal say $z(y)$ the vectors $v$ are available (Fraser & Reid, 2000) as

$$V = -z_{y'}^{-1}(y^0; \hat{\theta}^0) z_{;\theta'}(y^0; \hat{\theta}^0). \tag{6.1}$$

Also in the presence of an exact full ancillary $a(y)$ the vectors $V$ are available as a basis for the orthogonal complement of the ancillary gradient $(\partial/\partial y)a'(y)$; this can be obtained as the last $p$ row vectors of $A^{-1}$, where $A = (\partial/\partial y)\{a'(y), \ell_{\theta'}(\theta; y)\}$ evaluated at $(y^0, \hat{\theta}^0)$.

(iii) *Tangent exponential model.* Various possible conditional distributions as discussed in (i) can have the same ingredients $\ell^0(\theta)$, $\varphi(\theta)$ and yet differ by $O(n^{-1})$ but still give the same third-order inference. With these various possible models it is convenient to work with a familiar type, the exponential which can be expressed explicitly as

$$f(s; \varphi) = (2\pi)^{-p/2} e^k \exp\{\ell^0(\varphi) - \ell^0(\hat{\varphi}^0) + s'(\varphi - \hat{\varphi}^0)\} \hat{\jmath}_{\varphi\varphi}^{-1/2}, \tag{6.2}$$

where $k$ is constant of order $O(n^{-1})$, $\ell^0(\theta)$ has been reexpressed as $\ell^0\{\theta(\varphi)\}$ in terms of the canonical $\varphi$, and $s$ is the score variable with $s^0 = 0$; this is called the tangent exponential model (Fraser & Reid, 1993, 1995); see also Fraser, Wong & Wu (1999).

(iv) *Marginal model for testing* $\psi$. For testing $\psi(\varphi) = \psi$ of dimension $d$ and using the exponential model (6.2) we have (Fraser & Reid, 1995) that the relevant testing distribution is essentially unique as presented on the score plane

$\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$ . This plane intersects the contours of the ancillary for that $\psi$ value and accordingly the probability on the ancillary contours can be projected for fixed ancillary value and examined on the intersecting score plane. This gives the usual conditional exponential model referred to now as the profile support combined with a nuisance information factor obtained by integrating over the contours given the ancillary (Pierce & Peters, 1992; Fraser & Reid, 1995; Cheah et al, 1995). The plane passes through $s^0 = 0$ and is defined (Efron, 1978) by the span of the $d$ vectors $\partial\psi'(\varphi)/\partial\varphi|_{\hat{\varphi}^0_\psi}$ .

(v) *Simplified notation.* Now for ease of presentation we assume that both $\psi$ and $\lambda$ are scalars. Let $\varphi(\theta)$ in (1.4) be the canonical parameter defined at the data $y^0$ and suppose that it has been centred so that $\hat{\varphi}^0 = 0$ and scaled so that $\hat{j}^0_{\varphi\varphi} = I$ . For the interest parameter $\psi$ suppose that it has been centred so that $\hat{\psi}^0 = 0$ and scaled so that $\partial\psi/\partial\varphi_1 = 1$ , $\partial\psi/\partial\varphi_2 = 0$ at $\hat{\varphi}^0$ .

(vi) *Nonlinearity of* $\psi(\varphi)$ . We expand $\psi(\varphi)$ at $\hat{\varphi}^0 = 0$ ,

$$\psi(\varphi) = \varphi_1 + c_{11}\varphi_1^2/2n^{1/2} + c_{12}\varphi_1\varphi_2/n^{1/2} + c_{22}\varphi_2^2/2n^{1/2},$$

and we reexpress it so that $\psi - c_{11}\psi^2/2n^{1/2}$ becomes just $\psi$ ; this gives

$$\psi(\varphi) = \varphi_1 + c_{12}\varphi_1\varphi_2/2n^{1/2} + c_{22}\varphi_2^2/2n^{1/2}. \tag{6.3}$$

Then, with $\ell^0(\varphi) = -(\varphi_1^2 + \varphi_2^2)/2 + O(n^{-1/2})$ , we obtain $\hat{\varphi}_\psi = (\psi, c_{12}\psi^2/n^{1/2})'$ and the unit gradient vector to $\psi(\varphi)$ at $\hat{\varphi}_\psi$ is $u_\psi = -(1, c_{12}\psi/n^{1/2})'$ to order $O(n^{-1})$ . It follows (Fraser & Reid, 1995) that the distribution for assessing $\psi$ is on the line $\mathcal{L}(u_\psi)$ in score space which is at an angle $c_{12}\psi/n^{1/2}$ to the first axis $s_2 = 0$ . The vector $u_\psi$ plays a crucial role in the definition (Fraser & Reid, 1995; Fraser, Wong & Wu, 1999) of $Q$ for the $p$-value (2.4), and the squared term with coefficient $c_{22}$ in (6.3) is intrinsic to the Bayesian-frequentist difference as discussed in Fraser & Reid (1996, 2002).

(vii) *The profile support.* The log density from (6.2) can be expanded (Fraser & Reid, 1993; Abebe et al., 1995) in terms of re-expressed coordinates $\tilde{s}$ about the maximum density point, where $\tilde{s} = s$ to order $O(n^{-1})$. If we examine along the line $\mathcal{L}(0,1) = \mathcal{L}(u_{\hat{\psi}})$, the cubic term is $a_{111}s_1^3/6n^{1/2}$, and along the rotated line $\mathcal{L}(u_\psi)$ the cubic term is

$$(a_{111} + 3\varphi_1' c_{12} a_{112}/n^{1/2})\tilde{s}_1^3/6n^{1/2}, \tag{6.4}$$

where we have replaced $\psi$ by $\varphi_1$ using (6.3); the quartic term is unaffected to order $O(n^{-3/2})$ by the rotation, and also the norming constant analogous to $k$ in (6.2) is unaffected by the rotation. Of course the primary quadratic term is unaffected by the rotation because of the initial symmetrisation of the full information.

(viii) *The nuisance adjustment.* The nuisance effect adjustment comes from the integration mentioned in (iv) above and appears as the square root of a ratio of information determinants. At issue is whether or not there is a further factor derived from the rotation; in (vii) it was shown that the rotation has no effect using standardised coordinates and thus the information determinants carry the essential information.

(ix) *Rotationally symmetric model at the data point.* The rotation discussed in (vii) adds a $\varphi_1$ term to the cubic term as seen in (6.4). This added term can be moved to the tilt term in (6.2) and then $\tilde{s}_1$ can be redefined as $\tilde{s}_1 + c_{12}a_{112}s_1^4/n$. This gives a rotationally symmetric density subject to an exponential tilt.

(x) *The marginal likelihood.* The marginal density for $\tilde{s}_1$ at the data value based on the symmetrised model then gives the likelihood (1.2) using (6.3) in Fraser & Reid (1995).

(xi) Uniqueness of the third-order marginal likelihood. The derivations in this paper have used (a) conditioning given an appropriate ancillary, (b) marginalization to obtain a probability differential that describes the interest parameter $\psi$, and (c) symmetrisation of the canonical parameter to obtain a support differential that is free of $\psi$. Under moderate regularity the conditioning (a) is unique (Fraser & Reid, 2000) with independent scalar coordinates or with independent vector coordinates and a pivotal to describe how the coordinates measure the parameter. The marginalization (b) gives a unique probability differential as described in Fraser & Reid (1995). The symmetrisation (c) gives a unique target likelihood for the third-order theory. The complication mentioned in §2 of finding an appropriate support differential for assessing an interest parameter has thus been linked to the presence of non-parallel contours for the interest parameter as expressed in the canonical parameter space; the ratio of canonical parameters in an exponential model provides the prominent example.

Some additional detail concerning the examples and the derivation is available from `ftp://utstat.toronto.edu/pub/dfraser/219vx.pdf` in the form of earlier versions of this paper where `x` is a version number.

## 7. DISCUSSION

Recent likelihood asymptotics has largely resolved the search for $p$-values for scalar parameters, at least for continuous statistical models. The search for comparable likelihoods for component parameters has however been more elusive and key contributions have assumed the presence of some single variable that provides either a marginal or conditional assessment of the interest parameter. In many cases however the natural variable for this is found to depend on the parameter being tested as is the case with the $t$-variable for assessing a normal mean.

Our approach here has been to examine the large sample approximate form of the model that provides the assessment of a value $\psi$ and to note that this model typically depends on $\psi$. We then standardise the support variable and obtain a common model form; the probability differential describing $\psi$ then gives the loglikelihood function (1.2) which conforms to third-order calculations. It avoids some familiar anomalies and leads to the correct degrees of freedom for various examples where there is a likelihood otherwise viewed as acceptable.

**ACKNOWLEDGEMENT**

The Natural Sciences and Engineering Research Council of Canada has provided support for this research. Special appreciation is accorded to the associate editor for very substantial and insightful recommendations that greatly improved the revised version.

## REFERENCES

Abebe, F., Cakmak, S., Cheah, P.H., Fraser, D.A.S., Kuhn, J., McDunnough, P., Reid, N. & Tapia, A. (1995) Third order asymptotic model: exponential and location type approximations. *Parisankhyan Samikkha* **2**, 25-33.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-65.

Barndorff-Nielsen, O.E. (1985). Properties of modified profile likelihood. In *Contributions to Probability and Statistics in Honour of Gunnar Blom*, Ed. J.Lanke and G.Lindgren, pp 25-38. Lund: Lund University.

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

Barndorff-Nielsen, O.E. (1987). Discussion of 'Parameter orthogonality and approximate conditional inference' by D.R. Cox and N. Reid. *J. R. Statist. Soc.* B **49**, 18-20.

Barndorff-Nielsen, O.E. (1994). Adjusted version of profile likelihood and directed likelihood and extended likelihood. *J. R. Statist. Soc.* B **56**, 125-40.

Barndorff-Nielsen, O.E. & McCullagh, P. (1993). A note on the relation between modified profile likelihood and the Cox-Reid adjusted profile likelihood. *Biometrika* **80**, 321-8.

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

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

Cheah, P.K., Fraser, D.A.S., & Reid, N. (1995). Adjustment to likelihood and densities; calculating significance. *J. Statist. Res.* **29**, 1-13.

Efron, B. (1978). The geometry of exponential families. *Ann. Statist.* **6**, 362-76.

Fieller, E.C. (1954). Some problems in interval estimation. *J. R. Statist. Soc.* B **16**, 175-85.

Fraser, D.A.S. (1967). Data transformations and the linear model. *Ann. Math. Statist.* **38**, 1456-65.

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

Fraser, D.A.S. (1972). The determination of likelihood and the transformed regression model. *Ann. Math. Statist.* **43**, 898-916.

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

Fraser, D.A.S. (2002). Ancillaries and conditional inference. *Statist. Sci.*, to appear.

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

Fraser, D.A.S. & 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. & Reid, N. (1995). Ancillaries and third-order significance. *Utilitas Math.* **47**, 33-53.

Fraser, D.A.S. & Reid, N. (1996). Bayes posteriors for scalar interest parameters. In *Bayesian Statistics 5*, Ed. J.M. Bernardo, J.O. Berger, A.P.Dawid and A.F.M. Smith, pp 581-5. Oxford: Clarendon Press.

Fraser, D.A.S. & Reid, N. (2000). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference.* Ed. S.E. Ahmed and N. Reid, pp. 185-207. New York: Springer.

Fraser, D.A.S. & Reid, N. (2002). Strong matching of frequentist and Bayesian inference. *J. Statist. Plan. Infer.*, to appear.

Fraser, D.A.S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249-64.

Fraser, D.A.S., Wong, A. & Wu, J. (1999). Regression analysis, Nonlinear on non-normal: simple and accurate $p$-values from likelihood analysis. *J. Am. Statist. Assoc.* **94**, 1286-95.

Kalbfleisch, J.D. & Sprott, D.A. (1973). Marginal and conditional likelihood. *Sankhya* A **35**, 311-28.

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

McCullagh, P. & Tibshirani, R. (1980). A simple method for the adjustment of profile likelihoods. *J. R. Statist. Soc.* B **52**, 325-44.

Neyman, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1-32.

Pierce, D.A. & Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with Discussion). *J. R. Statist. Soc.* B

**54**, 701-37.

Severini, T.A. (1998). An approximation to the modified profile likelihood function. *Biometrika* **85**, 403-11.

Skovgaard, I.M., Martinussen, T. & Andersen, S.W. (2001). On the use of a parameter dependent group structure for inference. In *Data Analysis from Statistical Foundations*, Ed. A.K.Md.E. Saleh, pp. 53-64. New York: Nova Science.