JIRSS (2002)
Vol. 1, Nos. 1-2, pp 55-78

Location Reparameterization and Default Priors for Statistical Analysis

D. A. S. Fraser[1], **Grace Yun Yi**[2]

[1]Department of Statistics, University of Toronto, 100 St George Street, Toronto, Canada M5S 3G3. (dfraser@utstat.toronto.edu)
[2]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. (yyi@uwaterloo.ca)

Abstract. This paper develops default priors for Bayesian analysis that reproduce familiar frequentist and Bayesian analyses for models that are exponential or location. For the vector parameter case there is an information adjustment that avoids the Bayesian marginalization paradoxes and properly targets the prior on the parameter of interest thus adjusting for any complicating nonlinearity; the details of this vector Bayesian issue will be investigated in detail elsewhere. As in wide generality a statistical model has an inference component structure that is approximately exponential or approximately location to third order, this provides general default prior procedures that can be described as reweighting likelihood in accord with a Jeffreys' prior based on observed information.

Two asymptotic models, that have variable and parameter of the same dimension and agree at a data point to first derivative conditional on an approximate ancillary, produce the same p-values to third order for inferences concerning scalar interest parameters. With

---

Received: May 2002
Key words and phrases: Asymptotics, default prior, likelihood analysis, location model, Taylor series, transformations.

some given model of interest there is then the opportunity to choose some second model to best assist the calculations or best achieve certain inference objectives. Exponential models are useful for obtaining accurate approximations while location models present possible parameter values in a direct measurement or location manner. We derive the general construction of the location reparameterization that gives the natural parameter of the location model coinciding with the given model to first derivative at a data point; the derivation is in algorithmic form that is suitable for computer algebra. We then define a general default prior based on this location reparameterization; this gives third order agreement between frequentist p-values and Bayesian survivor values; in the vector case however, an adjustment factor is needed for component parameters that are not linear in the location parameterization. The general default prior can be difficult to calculate. But if we choose to work only to the second order, a Jeffreys' prior based on the observed information function gives second order agreement between the frequentist p-values and Bayesian survivor values; again adjustments are needed for parameters nonlinear in the vector location parameter; the adjustment is a ratio of two nuisance information determinants, one for the nuisance parameter as given and one for the locally equivalent linear nuisance parameter.

Preamble

A location model $f(y-\theta)$ on the real line is clearly the most basic and important model in statistics, one for which all inference approaches should be in agreement. Elementary statistics texts however typically restrict attention to the special case of normal densities $f(e)$ for the error $e = y - \theta$ and then develop a wealth of techniques based on feasibility for that special case. A few advanced monographs of course do present the inference theory for the general error case, but the theory and methods are certainly not widely available. In a parallel way various foundational approaches to statistics start with a model-data unit, where the model together with the observed data are viewed as the primary input ingredients for inference; and in practice this does not include any particular topology for the parameterization.

Recent likelihod theory shows that the terminal inference model is locaton to the second order and location to the third order in the special model data sense just described. We exmamine the form of this local model structure and show that Bayesian and frequentist methods are in agreement for linear parameters and that an easily constructed adjustment gives agreement more generally. Thus location model structure far from being a specialized anomaly is in fact the statistical substance of parameter-variable relationship.

Our theme examines this general location parameterization and the related equivalence of Bayesian and frequentist theories.

1 Introduction

Consider a continuous statistical model $f(y;\theta)$ with variable y and parameter θ both of dimension p and with asymptotic properties inherited from some antecedent model whose dimension n becomes large. Let $\ell(\theta;y^0) = \log f(y^0;\theta) - \log f(y^0;\hat{\theta})$ be the log-likelihood function coming from a data value y^0, where $\hat{\theta}$ is the maximum likelihood estimate.

For a specific data value y^0 we develop in this paper a location reparameterization $\beta(\theta)$ that allows third order inference to be presented as if the original model were location with canonical parameter $\beta(\theta)$. This is developed in Sections 3, 4, and Section 2 provides some needed background. The location reparameterization is a natural parameterization for a flat prior for default or nonsubjective Bayesian analysis. Section 5 discusses this third order default prior (5.2) that gives third order agreement between frequentist and Bayesian procedures, and Section 6 discusses a second order default prior (6.1) that gives second order frequentist-Bayes agreement and is much easier to calculate. Section 7 provides a brief overview.

Recent likelihood results show that for third order inference from a data point y^0 we need to have available only the observed likelihood $\ell(\theta)$ and observed likelihood gradient $\varphi(\theta)$ where

$$\ell(\theta) = \ell(\theta;y^0)\ ,\ \varphi(\theta) = \frac{\partial}{\partial y}\ell(\theta;y)|_{y^0} \tag{1.1}$$

Thus any model that provides a first derivative approximation to the likelihood for some given model near a data point y^0 will lead to the same inference results. We refer to such a model that provides a first derivative likelihood approximation as a tangent model to the given model at the data point y^0. For a recent summary of the background methodology see Fraser, Reid & Wu (1999).

For the scalar case with $p = 1$ the asymptotic form of a model at a data point y^0 was investigated (Fraser & Reid, 1993) by a Taylor expansion of $\log f(y;\theta)$ about y^0 and the corresponding $\hat{\theta}(y^0) = \theta^0$. The theory shows that third order inference depends only on the observed likelihood

$$\ell(\theta; y^0) = \log f(y^0;\theta) - \log f(y^0;\theta^0) \tag{1.2}$$

and the observed likelihood gradient

$$\varphi(\theta; y^0) = \frac{\partial}{\partial y}\left\{\log f(y;\theta) - \log f(y;\hat{\theta}(y))\right\}|_{y^0} \tag{1.3}$$

More specifically, it shows that $\ell(\theta)$ and $\varphi(\theta)$ fully determine the third order model except for a fourth order Taylor coefficient, quadratic in variable and quadratic in parameter. And more importantly, it shows that the p-value or probability left of the data is independent of the quadratic-quadratic fourth order coefficient. These results thus establish that the third-order p-value for testing any chosen parameter value is dependent only on $\ell(\theta)$ and $\varphi(\theta)$. An exponential model with the given $\ell(\theta)$ and $\varphi(\theta)$ was developed in Fraser & Reid (1993), and related results were obtained in Cakmak et al (1995, 1998). The model is third order and has the form

$$f_E(y) = \frac{c}{(2\pi)^{p/2}} \exp\left\{\ell\{\theta(\varphi)\} + \varphi'(y - y^0)\right\}|\tilde{\jmath}|^{-1/2} \tag{1.4}$$

where $\tilde{\jmath}$ is the information for φ obtained from the observed likelihood $\ell\{\theta(\varphi)\}$, $c = 1 + O(n^{-1})$ is constant to order $O(n^{-3/2})$, and for discussion now $p = 1$ and the bars on $\tilde{\jmath}$ are not needed. This extends Barndorff-Nielsen's (1983) p^* formula in the sense that p^* gives the density at a point y^0 with related $\ell(\theta)$ while f_E gives the density expression for a general y as propagated in an exponential manner from y^0 using the given $\ell(\theta)$ and $\varphi(\theta)$; in structure f_E can be viewed as a first derivative extension of the p^* formula at the given point. We can view $\varphi(\theta)$ as the canonical parameter of the fitted exponential model at the data point y^0.

A location model with the given $\ell(\theta)$ and $\varphi(\theta)$ was developed in Cakmak et al (1995, 1998). The model as developed is third order and has the form

$$f_L(y) = \frac{c}{(2\pi)^{p/2}} \exp\{\ell\{\theta(\beta - y + y^0)\}|\hat{\jmath}^0|^{-1/2} \tag{1.5}$$

where $\beta(\theta)$ is the location parameter and when $p = 1$ is given as

$$\beta(\theta) = \int_{\theta^0}^{\theta} -\frac{\ell_\theta(\theta)}{\varphi(\theta)} d\theta \tag{1.6}$$

where $\ell_\theta(\theta) = (\partial/\partial\theta)\ell(\theta)$ is the score parameter and $\varphi(\theta)$ is the canonical exponential parameterization. Thus we view $\beta(\theta)$ as the canonical parameter of the fitted location model at the data point y^0. It is straightforward to show that if the initial model is location $f\{x-\gamma(\theta)\}$ and has continuous derivatives then the extracted $\beta(\theta)$ is some version of the location parameter, that is, it is an affine function of γ.

The exponential parameterization is important for approximate inference calculations with a data point y^0 (Fraser, Reid & Wu, 1999); the location parameterization is important for inference presentations from data y^0 (Fraser, Reid & Wu, 1998) and provides the basis for default Bayesian priors.

Now consider the vector parameter case with general dimension p. Formula (1.3) is valid for vector y and θ and gives the canonical exponential parameterization $\varphi(\theta)$ for the approximating exponential model at the data point y^0. If the dimension of y is greater than that of θ then the differentiation in (1.1) and (1.3) needs to be in directions

$$V = (v_1, ..., v_p) \tag{1.7}$$

tangent to an approximate ancillary (Fraser & Reid, 2000). The corresponding tangent exponential model is given by (1.4). For inference results based on this see for example, Fraser & Reid (1995), and Fraser, Reid & Wu (1999).

Also for the vector case, an approximating location model would have the form (1.5) and would allow the presentation of third order inference results as if the original model were location with canonical parameter $\beta(\theta)$; for some discussion see Fraser, Reid & Wu (1998). The existence of $\beta(\theta)$ to the third order was established in Cakmak et al (1994); in Sections 3 and 4 we develop a Taylor series expansion for $\beta(\theta)$ in terms of the observed likelihood $\ell(\theta)$ and observed likelihood gradient $\varphi(\theta)$. This is the vector parameter version of (1.6).

2 Background: The scalar case

Consider a statistical model $f(y;\theta)$ with scalar variable and parameter and asymptotic properties as some external parameter n becomes large: we assume that $\log f(y;\theta)$ is $O(n)$ and that the maximum likelihood value $\hat{\theta}$ is unique and is $O_p(n^{-1/2})$ about θ as discussed for example in DiCiccio, Field & Fraser (1990).

Fraser & Reid (1993) examined the two dimensional Taylor expansion of $\ell(\theta;y) = \log f(y;\theta)$ about $(\hat{\theta}(y^0), y^0)$, where y^0 is an arbitrary point, typically an observed data point and $\hat{\theta}(y^0) = \theta^0$ is the corresponding maximum likelihood value. Further results on this and an expansion about $(\theta_0, \hat{y}(\theta_0))$ were examined by Cakmak et al (1995, 1998), where $\hat{y}(\theta_0)$ is the maximum density value for some chosen parameter value θ_0. We will be concerned with the first type of expansion here and let a_{ij} designate the Taylor coefficient for the ith derivative with respect to θ and the jth derivative with respect to y taken at the expansion point:

$$\log f(y;\theta) = \Sigma a_{ij}(\theta - \theta^0)^i (y - y^0)^j / i!j! \qquad (2.1)$$

The log density is examined in a moderate deviations range about the expansion point by using standardized coordinates $\tilde{\theta}$, $\tilde{y}$,

$$\tilde{\theta} = (\theta - \theta^0)\hat{\jmath}^{1/2}\ , \quad \tilde{y} = (y - y^0)\hat{k}\hat{\jmath}^{-1/2} \qquad (2.2)$$

where $\hat{\jmath} = -\ell_{\theta\theta}(\theta^0, y^0)$ is the observed information, $\hat{k} = \ell_{\theta;y}(\theta^0; y^0)$ is the observed gradient of the score, and $\ell(\theta;y)$ here is taken to be the log density, $\log f(y;\theta)$, with subscripts denoting differentation. The asymptotic properties show that the new coefficients $\tilde{a}_{ij}$ are $O(1)$ with $i+j=2$, are $O(n^{-1/2})$ with $i+j=3$, and are $O(n^{-1})$ with $i+j=4$; we neglect terms of order $O(n^{-3/2})$ and higher. For simplicity of notation we choose to write these modified variables and coefficients as just θ, y, and a_{ij}.

A reexpression of the original θ and y has the following pattern in terms of the standardized variables

$$\tilde{\theta} = \theta + b_1\theta^2/2n^{1/2} + b_2\theta^3/6n\ , \quad \tilde{y} = y + c_1 y^2/2n^{1/2} + c_2 y^3/6n \qquad (2.3)$$

where the initial coefficients are unity as a consequence of (2.2). The reexpressions can be chosen to give special structure to the reexpressed model. For exponential model and for location model structure appropriate transformations give the following two matrix arrays of Taylor coefficients a_{ij}:

$$\begin{pmatrix} a+\frac{3\alpha_4-5\alpha_3^2-12c}{24n} & \frac{-\alpha_3}{2\sqrt{n}} & -\{1+\frac{\alpha_4-2\alpha_3^2-5c}{2n}\} & \frac{\alpha_3}{\sqrt{n}} & \frac{\alpha_4-3\alpha_3^2-6c}{n} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \frac{c}{n} & - & - \\ \frac{-\alpha_3}{\sqrt{n}} & 0 & - & - & - \\ \frac{-\alpha_4}{n} & - & - & - & - \end{pmatrix} \qquad (2.4)$$

$$\begin{pmatrix} a+\frac{3\alpha_4-5\alpha_3^2-12c}{24n} & 0 & -\{\frac{1-5c}{2n}\} & \frac{\alpha_3}{\sqrt{n}} & \frac{-\alpha_4-6c}{n} \\ 0 & 1 & \frac{-\alpha_3}{\sqrt{n}} & \frac{\alpha_4}{n} & - \\ -1 & \frac{\alpha_3}{\sqrt{n}} & \frac{-\alpha_4+c}{n} & - & - \\ \frac{-\alpha_3}{\sqrt{n}} & \frac{\alpha_4}{n} & - & - & - \\ \frac{-\alpha_4}{n} & - & - & - & - \end{pmatrix} \tag{2.5}$$

where $a = -(1/2)\log 2\pi$. As the reexpressions (2.3) of the given model are different for the two model types, we have that the structure parameters α_3, α_4, and c are in general different in (2.4) and (2.5): in the first case α_3 and α_4 describe latent exponential structure and c records departure from exponential form; and in the second case α_3 and α_4 describe latent location structure and c records departure from location form. Also we have the remarkable fact that in each case the first row is determined by the remaining rows, an important property underlying the development of the approximations (1.4) and (1.5); in other words a density is available from a likelihood inversion, a rather important extension of the more familiar Fourier or saddlepoint inversion.

The reparameterization that gives the exponential approximation (1.4) is available from the second column of (2.4) as the first derivative of likelihood with respect to y:

$$\varphi(\theta) = \left(\frac{\partial}{\partial y}\ell(\theta; y) - \frac{\partial}{\partial y}\ell(\hat{\theta}; y) \right)\Bigg|_{y^0}$$

this agrees with the expression in (1.1) and the second term here is to accommodate the present definition of $\ell(\theta; y)$ which need not include the standardization at the maximum likelihood value $\hat{\theta}(y)$.

In the scalar parameter case, the reparameterization that gives the location approximation (1.5) is recorded as (1.6); for details see Cakmak et al (1995, 1998).

3 Background: Multivariate case

Consider a statistical model $f(y; \theta)$ with p-dimensional variable and parameter, and asymptotic properties as described in Section 2. We consider a Taylor expansion of $\ell(\theta; y) = \log f(y; \theta)$ about (θ^0, y^0), where $\theta^0 = \hat{\theta}(y^0)$ is the maximum likelihood value corresponding to a data value y^0 of interest. This gives

$$\ell(\theta; y) = a + a^i(y_i - y_i^0) + a_{ij}(\theta_i - \theta_i^0)(\theta_j - \theta_j^0)/2! + a_i^j(\theta_i - \theta_i^0)(y_j - y_j^0) + \cdots \tag{3.1}$$

where tensor type summation is assumed over $\{1,2,\ldots,p\}$ and for example $a^k_{ij} = (\partial/\partial\theta_i)(\partial/\partial\theta_j)(\partial/\partial y_k)\ell(\theta;y)|_{(\theta^0,y^0)}$. Note the change of notation from the preceding section where i gave the order of a derivative while now it designates a coordinate of θ or y. The coefficients can be recorded as a general matrix type array

$$\begin{pmatrix} a & a^i & a^{ij} & a^{ijk}\cdots \\ 0 & a^i_j & a^{ij}_k & \vdots \\ a_{ij} & a^i_{jk} & \vdots & \vdots \\ a_{ijk} & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{3.2}$$

following Cakmak et al (1994).

The log density is examined in a moderate deviations region by using location-scale standardized coordinates

$$\tilde{\theta}_i = c_{ij}(\theta_j - \theta^0_j) , \quad \tilde{y}_i = d_{ij}(y_j - y^0_j) \tag{3.3}$$

chosen so that the new second order coefficients in columns 1 and 2 have Kroneker delta or identity matrix form,

$$\tilde{a}_{ij} = -\delta_{ij} , \quad \tilde{a}^i_j = \delta^i_j \tag{3.4}$$

It follows that the new coefficients with three indices are $O(n^{-1/2})$ and with four indices are $O(n^{-1})$, as in Section 2; here we incorporate this dependence within the coefficients. The resulting log-likelihood ratio function at $y = 0$ is

$$\ell(\theta) = -\frac{1}{2}\delta_{ij}\theta_i\theta_j + a_{ijk}\theta_i\theta_j\theta_k/6 + a_{ijk\ell}\theta_i\theta_j\theta_k\theta_\ell/24 + \ldots \tag{3.5}$$

and the gradient of this log-likelihood ratio at $y = 0$ has α-th coordinate

$$\ell^\alpha(\theta) = \delta^\alpha_j\theta_j + a^\alpha_{jk}\theta_j\theta_k/2 + a^\alpha_{jk\ell}\theta_j\theta_k\theta_\ell/6 + \ldots \tag{3.6}$$

In these expressions we have again omitted the tildes for ease of notation. Also the expressions (3.5) and (3.6) are based on log-likelihood ratio

$$\ell(\theta;y) = \log f(y;\theta) - \log f(y;\hat{\theta}) \tag{3.7}$$

which is consistent with our general concern as to how likelihood itself determines an underlying density and model; in particular there is no constant term in the expression (3.6) and this is a consequence of having a particular mode of expression for the variable near zero.

Nonlinear reexpressions of the initial parameter and of the initial variable have the form indicated by (2.3) when presented in terms of the location scale standardized variables. Reexpressions can then in turn be chosen to give for example an approximating exponential model analogous to (2.4). We do not develop here the coefficients of the corresponding Taylor array but do note that the related exponential model which has the c-type array equal to zero can be written generally as (1.4) in terms of the original variables; the canonical parameter is given by (1.3) and the observed log-likelihood by (1.2).

Our primary interest is the location model approximation analogous to (2.5) which then has the form (1.5). In particular we seek the location reparameterization $\beta(\theta)$ that gives the vector generalization of (1.6).

For this we follow Cakmak et al (1994) and examine to what degree the statistical model departs from being of location model form. In particular, if the model as it stands is location, say $\ell(\theta; y) = \log f(y-\theta)$, then the first derivative property at $y = 0$ is $\ell_\theta(\theta; 0) + \ell_{;y}(\theta; 0) = 0$. Accordingly for a general model we define a nonlocation measure at y^0 by

$$d(\theta) = \{d^1(\theta), \ldots, d^p(\theta)\}' \tag{3.8}$$

where $d^i(\theta) = \ell^i(\theta) + \ell_i(\theta)$ and

$$\ell^i(\theta) = \frac{\partial}{\partial y_i}\ell(\theta; y)|_{y^0}\ , \quad \ell_i(\theta) = \frac{\partial}{\partial \theta_i}\ell(\theta; y)|_{y^0} \tag{3.9}$$

when expressed in terms of the standardized coordinates (3.3) with (3.4) . It follows that likelihood ratio describes a location model if and only if $d(\theta) \equiv 0$ for the standardized model satisfying (3.4) based on (3.3). We do note that the sandardization (3.4) does match particular y coordinates with corresponding θ coordinates.

As an illustration of the nonlocation discrepancy functions $d(\theta)$ we record some steps in the analysis of an example (Fraser et al, 1994) that does have an underlying location model.

Example 3.1. Consider independent $y_1, \ldots, y_n$ where y_i has density $f(y_i; \theta_i) = \exp\{-y_i\theta_i + \log\theta_i\}$, and where mean life

$$E(y_i) = \theta_i^{-1} = \exp\{\alpha + \beta(x_i - \bar{x})\}$$

is assumed to be log linear in α and β and where x is a concomitant variable. The observed log likelihood is

$$\ell(\alpha,\beta) = -\Sigma \exp\{-\alpha - \beta(x_i - \bar{x})\} y_i^0 - n\alpha$$

and the score parameters are

$$\begin{aligned} \ell_1(\alpha,\beta) &= \Sigma \exp\{-\alpha - \beta(x_i - \bar{x})\} y_i^0 - n \ , \\ \ell_2(\alpha,\beta) &= \Sigma \exp\{-\alpha - \beta(x_i - \bar{x})\} y_i^0 (x_i - \bar{x}) \end{aligned}$$

Tangent directions to a second order ancillary are obtained as

$$\begin{aligned} \frac{d\mathbf{y}}{d\alpha}\Big|_{(y^0,\hat{\theta}^0)} &= (y_1^0, \ldots, y_n^0)' \ , \\ \frac{d\mathbf{y}}{d\beta}\Big|_{(y^0,\hat{\theta}^0)} &= \{y_1^0(x_1 - \bar{x}), \ldots, y_n^0(x_n - \bar{x})\}' \end{aligned}$$

from a full dimensional pivotal quantity $(z_1, \ldots, z_n)$, where $z_i = \theta_i y_i$, and $dy_i/d\alpha = y_i^0$ and $dy_i/d\beta = y_i^0(x_i - \bar{x})$ are calculated for fixed pivotal at $(y^0, \hat{\theta}^0)$. Then as $d\ell(\theta; y)/dy_i = -\exp\{-\alpha - \beta(x_i - \bar{x})\}$ we obtain

$$\varphi = \frac{\partial \ell}{\partial V'}\Big|_{y^0} = -\Sigma \exp\{-\alpha - \beta(x_i - \bar{x})\} \cdot \begin{pmatrix} y_i^0 \\ y_i^0(x_i - \bar{x}) \end{pmatrix}$$

and

$$\begin{aligned} \begin{pmatrix} \ell^1(\alpha,\beta) \\ \ell^2(\alpha,\beta) \end{pmatrix} &= \varphi(\alpha,\beta) - \hat{\varphi}(\alpha,\beta) \\ &= -\Sigma \exp\{-\alpha - \beta(x_i - \bar{x})\} \begin{pmatrix} y_i^0 \\ y_i^0(x_i - \bar{x}) \end{pmatrix} + \begin{pmatrix} n \\ 0 \end{pmatrix} \end{aligned}$$

We then have immediately that the discrepancies are

$$\begin{aligned} d^1(\theta) &= \ell^1(\theta) + \ell_1(\theta) = 0 \\ d^2(\theta) &= \ell^2(\theta) + \ell_2(\theta) = 0 \end{aligned}$$

where the scale adjustment indicated for (3.4) is implicit in the model. This shows the existence of a local location parameterization. Of course here the full model is location in α and β and accordingly the conditional model is also location in α and β, as just verified by the current methods.

4 Location reparameterization

Consider further the asymptotic model $f(y;\theta)$ with p dimensional parameter and p dimensional variable; also we continue with the redefinition of $\ell(\theta;y) = \log f(y;\theta) - \log f(y;\hat{\theta})$ as log-likelihood ratio and use the location scale standardized version as obtained from the chosen transformations (3.3) with (3.4) .

We seek a reparameterization $\beta(\theta)$ for the statistical model $f(y;\theta)$ so that the model has location form with respect to $\beta(\theta)$ to first derivative at a data point y^0. For this in Section 3 we defined a nonlocation measure $d(\theta)$ based on the first derivative structure of the model at the point y^0.

If the model is location at $y = y^0 = 0$ then $d(\theta)$ in (3.8) is equal to zero. More generally we seek a transformation, typically nonlinear, of θ to say $\tilde{\theta} = \beta(\theta)$ so that a nonnull $d(\theta)$ for the initial model changes to a null $d(\theta)$ when recalculated for the model expressed in terms of the new $\tilde{\theta}$. For this we examine the form of the nonlocation measure as expanded about the centered maximum likelihood value $\tilde{\theta}^0 = 0$. For α in $\{1,\ldots,p\}$ we have the α-th score and α-th gradient

$$\ell_\alpha(\theta) = \frac{\partial \ell}{\partial \theta_\alpha}|_{y=0} = -\delta_{\alpha j}\theta_j + a_{\alpha jk}\theta_j\theta_k/2 + a_{\alpha jk\ell}\theta_j\theta_k\theta_\ell/6 + \ldots \quad (4.1)$$

$$\ell^\alpha(\theta) = \frac{\partial \ell}{\partial y_\alpha}|_{y=0} = \delta^\alpha_j\theta_j + a^\alpha_{jk}\theta_j\theta_k/2 + a^\alpha_{jk\ell}\theta_j\theta_k\theta_\ell/6 + \ldots \quad (4.2)$$

where we have incorporated the usual $n^{-1/2}$, $n^{-1}, \ldots$ into the coefficients. It follows that the nonlocation measure (3.8) has α-th coordinate given by

$$\begin{aligned} d^\alpha &= (a_{\alpha jk} + a^\alpha_{jk})\theta_j\theta_k/2 + (a_{\alpha jk\ell} + a^\alpha_{jk\ell})\theta_j\theta_k\theta_\ell/6 + \ldots \\ &= d^\alpha_{jk}\theta_j\theta_k/2 + d^\alpha_{jk\ell}\theta_j\theta_k\theta_\ell/6 + \ldots \end{aligned} \quad (4.3)$$

where say $d^\alpha_{jk} = a_{\alpha jk} + \alpha^\alpha_{jk}$ is a sum of a first column element and a second column element one row higher.

Now consider a quadratic reparameterization

$$\theta_i = \tilde{\theta}_i + b^i_{jk}\tilde{\theta}_j\tilde{\theta}_k/2 \quad (4.4)$$

that can make the recalculated quadratic discrepancy $d^\alpha_{jk} = 0$. For this we modify the methods in Cakmak et al (1994) so as to obtain a pattern that can be generalized to eliminate the higher order terms in (4.3).

If the initial parameter θ is replaced by the quadratic reexpression (4.4), then the new $a_{\alpha jk}$ is

$$a_{\alpha jk} - b^{\alpha}_{jk} - 2b^{k}_{\alpha j}$$

and the new a^{α}_{jk} is

$$a^{\alpha}_{jk} + b^{\alpha}_{jk}$$

with the result that the new d^{α}_{jk} is

$$d^{\alpha}_{jk} - 2b^{k}_{\alpha j} \tag{4.5}$$

We show that the $b^{k}_{\alpha j}$ can be chosen so that the new quadratic discrepancy is zero.

To show that the equations (4.5) are not quite as trivial as they might first appear in tensor notation we record representatives. If $\{\alpha, j, k\} = \{1, 1, 1\}$ we obtain

$$d^{1}_{11} = 2b^{1}_{11}$$

which gives $b^{1}_{11} = d^{1}_{11}/2$ and then more generally gives

$$b^{i}_{ii} = \frac{1}{2} d^{i}_{ii} \tag{4.6}$$

If $\{\alpha, j, k\} = \{1, 1, 2\}$ we obtain

$$d^{1}_{12} = b^{1}_{12} + b^{2}_{11} \quad , \quad d^{2}_{11} = 2b^{1}_{12}$$

which gives $b^{1}_{12} = d^{2}_{11}/2$ and $b^{2}_{11} = d^{1}_{12} - d^{2}_{11}/2$, and then more generally gives

$$b^{i}_{ij} = d^{j}_{ii}/2 \quad , \quad b^{j}_{ii} = d^{i}_{ij} - d^{j}_{ii}/2 \tag{4.7}$$

If $\{\alpha, j, k\} = \{1, 2, 3\}$ we obtain

$$d^{1}_{23} = b^{2}_{13} + b^{3}_{12} \quad , \quad d^{2}_{13} = b^{1}_{23} + b^{3}_{12} \quad , \quad d^{3}_{12} = b^{1}_{23} + b^{2}_{13}$$

which can be solved giving more generally for different i, j, k

$$b^{i}_{jk} = \frac{1}{2}(d^{j}_{ik} + d^{k}_{ij} - d^{i}_{jk}) \tag{4.8}$$

Note reassuringly here that (4.8) also reproduces (4.6) and (4.7) by letting various indexes be equal. It is of interest as background for the higher order terms to record the Jacobian determinants for the successive groups of linear equations

$$|2| = 2 \ , \quad \begin{vmatrix} 1 & 1 \\ 2 & 0 \end{vmatrix} = -2 \ , \quad \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2$$

We have shown that a quadratic representation from θ to a candidate $\beta(\theta) = \tilde{\theta}$ can eliminate the quadratic terms in the nonlocation measure. Now suppose that an $(m-1)$st order reparameterization has eliminated the $(m-1)$st order terms with preceding terms already eliminated. Then for a given parameterization θ we have that the discrepancies

$$d^{\alpha}_{j_1 j_2} = 0, \ \ldots, \ d^{\alpha}_{j_1 \cdots j_{m-1}} = 0 \tag{4.9}$$

and we seek an mth order reparameterization

$$\theta_\alpha = \tilde{\theta}_\alpha + b^{\alpha}_{j_1 \cdots j_m} \tilde{\theta}_{j_1} \cdots \tilde{\theta}_{j_m}/m! \tag{4.10}$$

so that the new $d^{\alpha}_{j_1 \cdots j_m}$ are equal to zero. As a preliminary, we note that the reexpression (4.10) has no effect on lower order discrepancies (4.9). Appendix A outlines the argument that shows linear equations can be solved to give the b^α in terms of the d^α. Of course this shows the existence of the location reparameterization but it also gives a procedure for the computer algebra implementation of the power series for $\beta(\theta)$ in terms of θ. Such a power series would be the multi-parameter analog of an expansion for the scalar $\beta(\theta)$ recorded in (1.6).

Example 4.1. Consider the scalar parameter case and suppose the $\ell(\varphi)$ is available as a function of the exponential parameter φ. Then using the standardized expression from (2.4) we obtain

$$\begin{aligned} \ell(\varphi) &= -\frac{1}{2}\varphi^2 - \alpha_3\varphi^3/6n^{1/2} - \alpha_4\varphi^4/24n \\ -\ell_\varphi(\varphi) &= \varphi + \alpha_3\varphi^2/2n^{1/2} + \alpha_4\varphi^3/6n \\ \beta(\varphi) &= \varphi + \alpha_3\varphi^2/4n^{1/2} + \alpha_4\varphi^3/18n \end{aligned} \tag{4.11}$$

giving the start of the expansion for the location parameter in terms of the exponential parameter.

As a particular case consider the simple exponential model $\varphi \exp\{-\varphi x\}$ with data value say $x = 1$. The observed likelihood is

$$\ell^0 = -\varphi + log\varphi = -\frac{\varphi^2}{2} + \frac{\varphi^3}{3} - \frac{\varphi^4}{4}$$

giving $\alpha_3 = -2$ and $\alpha_4 = 6$. The expression for β from (4.11) is then

$$\beta = \varphi - \frac{\varphi^2}{2} + \frac{\varphi^3}{3}$$

which records the lead terms in the log function and is consistent with $\log \varphi$ as a version of the location parameter of the model.

Theorem 4.1. *If $f(y;\theta)$ has location form $f\{y-\gamma(\theta)\}$ and if f and γ have continuous derivatives, then $\beta(\theta)$ is an affine function of $\gamma(\theta)$ and is thus an alternate version of the location parameter.*

Proof. Let $\gamma(\theta)$ be the parameter θ for the calculations in Example 4.1. The nonlocation measure d^α is identically zero and all the $d^\alpha_{i_1\ldots i_k}$ are equal to zero. The resulting $b^\alpha_{i_1\ldots i_k}$ are then equal to zero and $\beta(\theta) = \theta$.

5 Frequentist inference and default priors

Much of inference theory until the 1950's tended to seek procedures that satisfied optimality criteria. The supporting arguments were persuasive but the theory in many cases did not produce answers for the wealth of more complicated statistical models that were emerging. The Bayesian approach made prominent by Savage (1954) and others provided answers to many problems, by using the likelihood function together with an objective or subjective choice of prior probability measure. The strict Bayesian view would argue that such answers were definitive but the less committed would be at least skeptical. Of course, the frequentists could have used likelihood as a central inference ingredient and modified it by suitable pragmatic weight functions and obtained the same wealth of answers, but the focus on optimality somehow prevented this.

Our viewpoint here is that observed data together with an appropriate statistical model entails certain inference results or presentations, and that these exist apart from any proposed sequence of such data-model combinations; in short, the probability component is provided by the model and an assumed valid data value. Of course, conditioning may be involved, see for example Fraser (2001).

Recent likelihood analysis as initiated by Barndorff-Nielson (1983, 1986) provides inference for scalar parameters in a wide variety of contexts with continuous variables; some extension to discrete variables is in development. The details of development show these results to be definitive and highly accurate in wide generality. The results use only the observed likelihood function $\ell(\theta)$ from (1.2) and the observed likelihood gradient $\varphi(\theta)$ from (1.3) or from the extensions using vectors V in (1.7).

In particular for inference concerning a value ψ for a scalar parameter $\psi(\theta)$, the essential p-value $p(\theta)$ is obtained from the signed likelihood root $r(\psi)$ together with a special maximum likelihood type departure $q(\psi)$ (for example: Barndorff-Nielsen, 1986; Fraser & Reid, 1995; Fraser, Reid & Wu, 1999) using a combining formula such as the Lugannani & Rice (1980) formula, the Barndorff-Nielsen(1986) formula or modifications thereof.

The Barndorff-Nielsen formula has the form

$$p(\psi) = \Phi(r^*) = \Phi(r - r^{-1} log \frac{r}{q}) \tag{5.1}$$

where Φ is the standard normal distribution function and $r^*(\psi)$ is a corrected likelihood root implicitly defined by (5.1).

For some recent discussion and formulas see Fraser, Reid & Wu (1999). This reference also develops a third order formula for the Bayesian survivor function $s(\psi)$ which can be presented as (5.1) but with q replaced by a special score type departure $q_B(\psi)$.

We will be concerned with Bayesian and frequentist results that agree to third order. In the simple case of data y^0 from a location model $f(y - \theta)$ the frequentist p-value

$$p(\theta) = \int_{-\infty}^{y^0} f(y - \theta) dy$$

and the Bayesian survivor function

$$s(\theta) = \int_{\theta}^{\infty} f(y^0 - \theta) d\theta$$

based on a flat prior are obviously equal. This has long been acknowledged often from widely different viewpoints. And it provides a framework for various arguments supporting the general use of the flat prior. For a vector location model $f(y - \theta)$, the equality holds for parameter components $\psi(\theta) = a'\theta + b$ that are linear in the location parameter. But in general the equality is not available for curved parameter components. See Fraser & Reid (2001) for discussion and various details concerning curved parameters, as well as an adjustment factor to target a prior on a component curved parameter of interest.

As a default prior for general Bayesian use we propose

$$d\beta(\theta) = \pi_{\text{def}}(\theta)d\theta \tag{5.2}$$

where $\pi_{\text{def}}(\theta) = |d\beta/d\theta|$; in the scalar parameter case the default prior can be expressed simply using (1.6) as

$$\pi_{\text{def}}(\theta) = |\frac{\ell_\theta(\theta)}{\varphi(\theta)}| \tag{5.3}$$

which uses (1.2) and (1.3) or its more general version derived from (1.7).

Theorem 5.1. *If $f(y;\theta)$ is a scalar model with location form*

$$f\{x(y) - \gamma(\theta)\}h(y)$$

then the default prior from (5.2) or (5.3) gives

$$\pi_{\text{def}}(\theta) = |\frac{d\gamma(\theta)}{d\theta}|$$

which is the flat prior in the location parameter $\gamma(\theta)$.

Proof. This is a trivial consequence of Theorem 4.1 for the scalar variable and scalar parameter case.

In general applications with continuous variables the default prior $\pi_{\text{def}}(\theta)$ gives third order agreement between frequentist and Bayesian theory for linear location parameters. The frequentist theory uses appropriate conditioning, and the Bayesian theory uses the flat prior from the third order location model approximation to the given model. Both of these follow viewpoints that are acceptable to particular areas of inference theory. Subjective priors can then in turn be presented relative to the default prior $\pi_{\text{def}}(\theta)$; thus the default prior $\pi_{\text{def}}(\theta)$ can be viewed as a reference prior. Bernardo's (1979) reference priors have of course this objective; they do however involve sample space averages and full model information functions. By contrast, the present development is based on the large sample form of the model and uses only the observed likelihood and how such likelihood would change in the neighbourhood of the data point; this is closer to some Bayesian ideal, unless the data point were treated as a mathematical point with no record of local sensitivity towards likelihood. In some oral presentations the present default priors have been referred to as deference priors, deferring to model form. Curved component parameters however do need a targetting adjustment (Fraser & Reid, 2001), given as the ratio of nuisance information determinants.

6 Second order default priors

We have discussed aspects of the third order asymototic model and given various references to background material. In particular the theory conforms to reasonable conditioning requirements (for some recent discussion, see Fraser, 2001), and produces the exponential reparameterization $\varphi(\theta)$ and the location reparmeterization $\beta(\theta)$ for the given statistical model, at least, for the continuous variable case. We have also noted that the exponential parameterization $\varphi(\theta)$ is easily computed using (1.1) and (1.3) with (1.7), but that the location parameterization is not readily available, being obtained as an iteratively defined power series (Section 4). This latter puts substantial constraints on the direct use of (5.2) for Bayesian default analysis.

Now suppose we restrict our attention to second order asymptotic analysis. The available theory as cited in preceding sections shows that the statistical model is exponential to the second order in a moderate deviations region and is location to the second order in a moderate deviations range. Also the construction of these approximation models shows that the exponential model and the location model are in fact the same model, just differing in the modes of expression for the variable and the parameter. This has substantial implications.

For an exponential model a second-order default prior is given by Jeffreys' (1946) prior; indeed Welch & Peers (1963) show in effect that the frequentist p-value, $p(\theta)$, and the Bayesian survivor value $s(\theta)$ based on Jefferys' prior are equal to the second order.

For the more general asymptotic context here, we can note that Jeffreys' prior is parameterization invariant and thus can be calculated in the location model reexpression of the model where of course it is a flat prior. It follows then that to second order the default prior $\pi_{\text{def}}(\theta)$ can be expressed in terms of the information function from the exponential reexpression of the model. It follows that to secord order the default prior is given by

$$\pi_{\text{def}}(\theta)d\theta = d\beta(\theta) = |j_{\varphi\varphi}(\theta; y^0)|^{1/2} d\varphi(\theta) \tag{6.1}$$

where $j_{\varphi\varphi}(\theta; y^0)$ is the observed information function

$$j_{\varphi\varphi}(\theta; y^0) = -\frac{\partial}{\partial\varphi}\frac{\partial}{\partial\varphi'}\ell(\theta; y^0)$$

which is in fact the expected information for the approximating second order exponential model.

If the initial model is exponential then this reproduces Jeffreys' prior. And if the initial model is location then this present method gives the flat prior in that parameterization. From a more general viewpoint it is just Jeffreys' prior for the visible statistical model, visible to the second order. Also in passing we can note that the analysis in Section 2 with our present discussion gives an alternate proof of the Welch & Peers (1963) result.

7 Some overview

We have noted that recent likelihood analysis has been remarkably successful for statistical inference, at least for the case with continuous observable variables. The model is sectioned in accord with appropriate conditioning, the nuisance parameters are eliminated by subsequent marginalization, and the p-values are obtained by likelihood inversion, which is a generalization of third order Fourier inversion. And all of this has clear uniqueness aspects, at least to the third order.

A pragmatic check is to examine the extremes, of small sample sizes and difficult distributions (Fraser, Reid & Wu, 1999). For example a sample of one from a location Cauchy gives (Fraser, 1990) acceptable p-values using (5.1). However, as noted by Prof. Chris Field the asymptotic methods can have difficulies with a location model with two modal points, but even the maximum likelihood itself has difficulties with such models; two mode models it should be noted can arise easily with a sample of two from the location Cauchy. Nevertheless the methodology is widely applicable and reliable, as indicated by simulations, even though there are fringes where issues can arise.

In our present context it does however seem appropriate to examine the proposed default priors in the presence of an extreme such as the location Cauchy. The third order prior (5.3) gives the seemingly appropriate flat prior (Theorem 5.1). The second order flat prior (6.1) is flat near the maximum likelihood value, drops to zero towards infinity, but has a singularity at $\theta = \hat{\theta}^0 \pm 1$. This is not surprising given that the exponential parameterization

$$\varphi(\theta) = 2(\theta - \hat{\theta}^0)/\{1 + (\theta - \hat{\theta}^0)^2\}^{-1}$$

is redescending giving a reparameterization that is not one-one; this feature seems not to upset the third order p-values or the third order default prior but does cause moderate perturbations with the second order Bayes methods.

With moderate data and reasonable models we can expect likelihoods that are convex downwards and are amenable to likelihood analysis. The direct integration of likelihood with or without a weight or prior functions does have risks but the present proposal is to start with a recalibrated likelihood that reflects how the variable in the neightbourhood of the data is measuring the parameter of the model.

8 Acknowledgements

The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada and the helpful suggestions and comments by readers and discussants of preliminary presentations of this material. The following discussion responds to some of the suggestions and comments.

Recent likelihood theory for the continuous variable case has made extensive use of conditioning down, so that the free variables have the dimension of the parameter. While in general this is not available exactly, it is available widely to third order which is the level of accuracy for the recent theory (see for example, Fraser & Reid, 2000, 2001).

Sufficiency is not mentioned in the present discussions, which address models with continuous observable variables. Suppose now that there is a sufficient statistic $s(y)$ having the same dimension p as the parameter. Also suppose for ease of argument that the conditioned variable given $s(y)$, say $t(y)$, has constant dimension which would then be $n-p$. It follows that the distribution of $t(y)$ given $s(y)$ is parameter free: let $u(y)$ be a multivariate probability integral transformation of $t(y)$ as obtained from the conditional distribution given $s(y)$. Then the conditional distribution of $u(y)$ given $s(y)$ is free of s and thus u and s are independent. It follows that $f(s;\theta)$ and $f(s|t;\theta)$ are equal, that is, the marginal and conditional models are equal. This says that an analysis using sufficiency can be duplicated by a conditional analysis. For a simple example consider (y_1, y_2) from the normal (θ, σ_0^2). The model for $\bar{y}$ is $N(\theta, \sigma_0^2/2)$: the conditional model for $\bar{y}$ given the configuration $y_2 - y_1$ is also $N(\theta, \sigma_0^2/2)$. If, however, we depart from normality then sufficiency is quite generally not available, but the conditional analysis remains available and is routine. Accordingly we follow the conditional approach and suggest that there is no need for sufficiency methods for inference in the continuous case.

The data dependence of a statistical procedure is closely related to the use of conditional statistical procedures. But the statistical literature has certainly not embraced the use of conditional methods; for example, issues raised in Cox (1958) are not yet part of core statistical thinking. Of course there are some important conditional methods for eliminating nuisance parameters; but these can all be handled by a more general marginalization method. For some recent discussion see Fraser (2001). For a Bayesian illustration derived from Cox (1958) suppose that z has a Bernoulli (1/2) distribution and $y|z$ has a model with information $I_i(\theta)$ if $z = i$. Should the Bayesian use Jeffreys' prior which is $I(\theta) = \{I_1(\theta) + I_2(\theta)\}/2$ or should he use the conditional Jeffreys' which is $I_z(\theta)$ where z has the observed value of the indicator. The use of the conditional Jeffreys' does seem persuasive and two distinguished Bayesian colleagues would seem to agree. Our viewpoint here is that Bayesian analysis should work from the measurements actually made and not invoke a model that describes measurements that were not made.

The proposals in Section 5 go beyond this somewhat standard use of conditional methods. The recent likelihood theory shows that inference to third order uses only observed likelihood and the sensitivity of the likelihood function at the observed data point. This led to the use of tangent models at an observed data point; it also led to the viewpoint that data with a model entails certain inference results or presentations apart from any proposed sequence of such data model combinations (Section 5). This could not be viewed as a generally accepted inference approach at the present time. It does have reasonable grounds as discussed above and the counter argument would seem to require a distribution on such model data combinations. Our viewpoint is that inference from a model data combination itself deserves direct attention, and many larger issues can be addressed from this.

Are sample space derivatives an unattractive feature for Bayesian adoption? Certainly they depart from subjective and recent default Bayesian approaches. Is this convention? Or does the use of an observed likelihood not suggest that sensitivity of likelihood to the data input is of some relevance? And it may be preferable to the sample space averages involved in the definition of expected information. The success of such likelihood methods in inference theory does suggest that it can be of value to the Bayesian approach.

Normality is widely used as an approximation for sums of independent variables. Recent theory establishes that sums of log densities provides an approximate exponential model, for which accurate p-values are readily available. This later is useful for frequentist inference and the present viewpoint is that it can be useful for Bayesian inference, taking it to second order accuracy from a frequent first order accuracy.

Appendix A. Iteration for the location parameter

We now show that the coefficients $b^{\alpha}_{j_1\cdots j_m}$ in the expansion (4.10) for θ can be solved for so that the mth order discrepancies $d^{\alpha}_{j_1\cdots j_m}$ are equal to zero.

If we substitute (4.10) in (3.5) we find that the new $a_{\alpha j_1\cdots j_m}$ is

$$a_{\alpha j_1\cdots j_m} - b^{\alpha}_{j_1\cdots j_m} - m b^{j_m}_{\alpha j_1\cdots j_{m-1}}$$

and the new $a^{\alpha}_{j_1\cdots j_m}$ is

$$a^{\alpha}_{j_1\cdots j_m} + b^{\alpha}_{j_1\cdots j_m}$$

with the result that the new mth order discrepancies are

$$d^{\alpha}_{j_1\cdots j_m} - m b^{j_m}_{\alpha j_1\cdots j_m} \qquad (A.1)$$

Can we choose the $b^{\alpha}_{j_1\cdots j_m}$ so that these new mth order discrepancies are all zero?

The discrepancies (A.1) appear as the coefficient of $\tilde{\theta}_{j_1}\cdots\tilde{\theta}_{j_m}/m!$ and thus must be symmetrized as in the quadratic case following (4.5). For example, the coefficient of $\tilde{\theta}_1^m/m!$ for the α-th coordinate gives immediately the equation

$$d^{\alpha}_{1\cdots 1} = m b^{1}_{1\cdots 1\alpha}$$

but the coefficient of $\tilde{\theta}_1^{m-2}\tilde{\theta}_2\tilde{\theta}_3/(m-2)!$ gives the equation

$$d^{\alpha}_{1\cdots 123} = (m-2)b^{1}_{1\cdots 123\alpha} + b^{2}_{1\cdots 13\alpha} + b^{3}_{1\cdots 12\alpha}$$

where each item has m subscripts. More generally if j_1 appears m_1 times, ..., j_r appears m_r times with $\Sigma m_i = m$ and $j_1 < \cdots < j_r$ then the symmetrized form of (A.1) is

$$d^{\alpha}_{j_1\cdots j_r} = m_1 b^{j_1}_{j_1\cdots j_r\alpha} + \cdots + m_r b^{j_r}_{j_1\cdots j_r\alpha} \qquad (A.2)$$

where each term has m subscripts and a j_i as superscript requires one less j_i as subscript with the missing j_i replaced by α.

Now consider the full set of integers that appear in an (A.2) type of equation and let j_1 appear m_1 times, ..., j_r appear m_r times where $j_1 < \cdots < j_r$ and $\Sigma m_i = m+1$. We consider the r different equations that use this collection of integers for superscript and subscripts. Specifically we take $\alpha = j_1$ in the $d^{\alpha}_{j_1\cdots j_r}$ equation, ..., $\alpha = j_r$ in the $d^{\alpha}_{j_1\cdots j_r}$ equation; the Jacobian of the r equations in the r different b's is

$$J = \begin{vmatrix} m_1 - 1 & m_2 & \cdots & m_{r-1} & m_r \\ m_1 & m_2 - 1 & \cdots & m_{r-1} & m_r \\ \vdots & \vdots & & \vdots & \vdots \\ m_1 & m_2 & \cdots & m_{r-1} - 1 & m_r \\ m_1 & m_2 & \cdots & m_{r-1} & m_r - 1 \end{vmatrix} = (-1)^{r-1} m \qquad (A.3)$$

Accordingly we can solve for the b's:

$$b^{j_i}_{j_1\cdots j_r} = \frac{J_i}{J} \qquad (A.4)$$

where J_i is the determinant J but with the ith column replaced by the vector of d's that are recorded in sequence above. For this we note that the total number of j_i's is m_i and thus that the subscript array on the left side is different for each j_i as superscript. With various values of $r = 1, \ldots, m+1$ and various integers $j_1 < \cdots < j_r$ with various frequencies $m_1, \ldots, m_r$ with $\Sigma m_i = m+1$ we determine the mth order parameter adjustment to give the mth order location property.

Iteration on m then determines the power series representation for the location parameter $\beta(\theta)$ in terms of the original parameter θ.

References

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

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

Bernardo, J. M. (1979), Reference posterior distributions for Bayesian inference. J. Royal Statistics B, **41**, 113-147.

Cakmak, S., Fraser, D. A. S., McDunnough, P., Reid, N., and Yuan, X. (1995), Likelihood centered asymptotic model: exponential and location model versions. Int. J. Math. & Statist. Sc., **4**, 1-12.

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 & Inference, **66**, 211-222.

Cakmak, S., Fraser, D. A. S., and Reid, N. (1994), Multivariate asymptotic model: exponential and location approximations. Utilitas Mathematica, **46**, 21-31.

Cox, D. R. (1958), Some problems connected with statistical inference. Math. Stat., **29**, 357-372.

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

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

Fraser, D. A. S. (2001), Ancillaries and conditional inference. Statistical Science, To appear.

Fraser, D. A. S., Monette, G., Ng. K. W., and Wong, A. (1994), Higher order approximations with generalized linear models. IMS Lecture Notes - Monograph Series, **24**, 253-262.

Fraser, D. A. S. and Reid, N. (1993), Simple asymptotic connections between densities and cumulant generating 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, **47**, 33-53.

Fraser, D. A. S. and Reid, N. (2000), Ancillary information for statistical inference. Empirical Bayes and likelihood Inference, Eds: Ahmed, S. E., Reid, N., New York: Springer, 185-210.

Fraser, D. A. S. and Reid, N. (2001), Strong matching of frequentist and Bayesian inference. Journal of Statistical Planning and Inference, To appear.

Fraser, D. A. S., Reid, N., and Wu, J. (1998), On the informative presentation of likelihood. Applied Statistical Science III, Commack, New York: Nova Science Publishers, 253-266.

Fraser, D. A. S., Reid, N., and Wu, J. (1999), Regression Analysis, Nonlinear or nonnormal: simple and accurate p-values from likelihood analysis. Jour. Amer. Statist. Association, **94**, 1286-1295.

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

Jeffreys, H. (1946), An invariant form for the prior distribution in estimation problems. Proc. Royal Soc., Ser. A, **186**, 453-461.

Lugannani, R. and Rice, S. (1980), Saddlepoint approximation for hte distribution of the sum of independent random variables. Adv. Appl.Prob., **12**, 475-490.

Savage, L. J. (1954), The Foundations of Statistics. New York: John Wiley.

Welch, B. L. and Peers, H. W. (1963), On Formulae for confidence points based on integrals of weighted likelihood. J. of the Royal Stat. Soc., Ser. B, **25**, 318-329.