

LOCATION REPARAMETERIZATION OF MULTIVARIATE MODELS

D.A.S. FRASER and Y. YI
Department of Statistics
University of Toronto
Toronto, Canada M5S 3G3

SUMMARY

Two asymptotic models which have variable and parameter of the same dimension and which agree at a data point to first derivative, conditional on an approximate ancillary, lead to the same p -values to third order for inferences concerning scalar interest parameters. With some given model of interest one can then choose a 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 gives the natural parameter of the location model that coincides with a given model at a data point. The derivation is in algorithmic form that is suitable for computer algebra. This extends the third order existence result and provides the basis for developing default priors for Bayesian analysis.

Keywords: Asymptotics, Likelihood analysis; Location model; Taylor series; Transformations.

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 $\ell(\theta; y) = \log f(y; \theta) - \log f(y; \hat{\theta})$ be the log-likelihood function from a point y .

For a data point 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 was location with canonical parameter $\beta(\theta)$. This is developed in Sections 4, 5 and earlier sections provide the necessary background. The location reparameterization is a natural parameterization for a flat prior for default or nonsubjective Bayesian analysis.

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 its gradient $\varphi(\theta)$ at the data point where

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

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 and 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

$$\ell(\theta) = \log f(y^0; \theta) - \log f(y^0; \theta^0) \quad (1.2)$$

and the observed likelihood gradient

$$\begin{aligned}\varphi(\theta) &= \frac{\partial}{\partial y} \left\{ \log f(y; \theta) - \log f(y; \hat{\theta}(y)) \right\} \Big|_{y^0} \\ &= \ell_{;y} \{ \ell(\theta; y) - \ell(\hat{\theta}; y) \} \Big|_{y^0} .\end{aligned}\tag{1.3}$$

More specifically, it shows that $\ell(\theta)$ and $\varphi(\theta)$ fully determine the third order model save 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

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

where \tilde{j} 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{j} 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 the given $\ell(\theta)$ and $\varphi(\theta)$ at y^0 ; 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

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

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

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

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 .

The exponential parameterization is important for approximate inference calculations with a data point y^0 ; the location parameterization is important for inference presentations from data y^0 (Fraser, Reid, Wu, 1998) and provides a 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 ; for some inference results based on this see Fraser & Reid (1995). The corresponding or tangent exponential model is given by (1.4).

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 $\beta(\theta)$. The existence of a third order expansion for $\beta(\theta)$ was established in Cakmak et al (1994). In this paper we develop a Taylor series expansion for $\beta(\theta)$ about $\theta^0 = \hat{\theta}(y^0)$; this is obtained in a form suitable for computer algebra and is presented in Sections 4 and 5. The usefulness of this for Bayesian default or nonsubjective priors will be discussed elsewhere.

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 i th derivative with respect to θ and the j th derivative with respect to y taken at the expansion point:

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

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

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

where $\hat{j} = -\ell_{\theta\theta}(\theta^0, y^0)$ 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 differentiation. 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

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

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} :

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

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

where $a = -(1/2) \log 2\pi$. As the reexpressions (2.3) 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 c records departure from exponential form and in the second case it records departure from location form. Also we have 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 :

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

this agrees with the expression in (1.1) and the second term here is to accommodate the present definition of $\ell(\theta; y)$ which does not include the standardization (1.2) at the maximum likelihood value.

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

$$\begin{aligned} \ell(\theta; y) = & a + a^i (y_i - y_i^0) \\ & + a_{ij} (\theta_i - \theta_i^0) (\theta_j - \theta_j^0) / 2! + a_i^j (\theta_i - \theta_i^0) (y_j - y_j^0) + \dots \end{aligned} \quad (3.1)$$

where tensor type summation is assumed over $\{1, 2, \dots, p\}$ and for example $a_{ij}^k = (\partial/\partial\theta_i)(\partial/\partial\theta_j)(\partial/\partial y_k)\ell(\theta; y)|_{(\theta^0, y^0)}$. 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

$$\begin{pmatrix} a & a^i & a^{ij} & a^{ijk} \dots \\ 0 & a_j^i & a_k^{ij} & \vdots \\ a_{ij} & a_{jk}^i & \vdots & \vdots \\ a_{ijk} & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (3.2)$$

we follow Cakmak et al (1994) for general notation.

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

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

chosen so that the new second order coefficients in columns 0 and 1 have Kroneker delta or identity matrix form,

$$\tilde{a}_{ij} = -\delta_{ij} , \quad \tilde{a}_j^i = \delta_j^i . \quad (3.4)$$

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

$$\ell(\theta) = -\frac{1}{2}\delta_{ij}\theta_i\theta_j + a_{ijk}\theta_i\theta_j\theta_k/6 + a_{ijkl}\theta_i\theta_j\theta_k\theta_l/24 + \dots \quad (3.5)$$

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

$$\ell^\alpha(\theta) = \delta_j^\alpha\theta_j + a_{jk}^\alpha\theta_j\theta_k/2 + a_{jkl}^\alpha\theta_j\theta_k\theta_l/6 + \dots \quad (3.6)$$

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

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

as we are centrally concerned with how likelihood itself determines an underlying density and model; in particular there is no constant term in the expression (3.6) and this is related to have some special choice of mode of expression for the variable involved.

Nonlinear reexpressions of the initial parameter and of the initial variable have the form indicated by (2.3) which 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, $\ell(\theta; y) = \log f(y - \theta)$, 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

$$d(\theta) = \{d'(\theta), \dots, d^p(\theta)\}' \quad (3.8)$$

where $d^i(\theta) = \ell^i(\theta) + \ell_{,i}(\theta)$ and

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

It follows that likelihood ratio describes a location model if and only if $d(\theta) \equiv 0$ for the standardized model satisfying (3.4) from (3.3).

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 $\tilde{\theta} = \beta(\theta)$ that changes a nonnull $d(\theta)$ for the initial model to a null $d(\theta)$ when calculated for the model reexpressed 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 $\{1, \dots, p\}$ we have the α -th score and α -th gradient

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

$$\ell^\alpha(\theta) = \frac{\partial \ell}{\partial y_\alpha} \Big|_{y=0} = \delta_j^\alpha \theta_j + a_{j k}^\alpha \theta_j \theta_k / 2 + a_{j k \ell}^\alpha \theta_j \theta_k \theta_\ell / 6 + \dots \quad (4.2)$$

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

$$\begin{aligned} d^\alpha &= (a_{\alpha j k} + a_{j k}^\alpha) \theta_j \theta_k / 2 + (a_{\alpha j k \ell} + a_{j k \ell}^\alpha) \theta_j \theta_k \theta_\ell / 6 + \dots \\ &= d_{j k}^\alpha \theta_j \theta_k / 2 + d_{j k \ell}^\alpha \theta_j \theta_k \theta_\ell / 6 + \dots \end{aligned} \quad (4.3)$$

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

Now first we consider a quadratic reparameterization

$$\theta_i = \tilde{\theta}_i + b_{j k}^i \tilde{\theta}_j \tilde{\theta}_k / 2 \quad (4.4)$$

that can make the recalculated quadratic discrepancy $d_{j k}^\alpha = 0$.

For this we modify the methods in Cakmak et al 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 $a_{\alpha j k}$ is

$$a_{\alpha j k} - b_{j k}^\alpha - 2b_{\alpha j}^k$$

and the new $a_{j k}^\alpha$ is

$$a_{j k}^\alpha + b_{j k}^\alpha$$

with the result that the new $d_{j k}^\alpha$ is

$$d_{j k}^\alpha - 2b_{\alpha k}^j. \quad (4.5)$$

We show that the $b_{\alpha k}^j$ 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

$$d_{11}^1 = 2b_{11}^1$$

which gives $b_{11}^1 = d_{11}^1/2$ and then more generally gives

$$b_{ii}^i = \frac{1}{2}d_{ii}^i . \quad (4.6)$$

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

$$d_{12}^1 = b_{12}^1 + b_{11}^2$$

$$d_{11}^2 = 2b_{12}^1$$

which gives $b_{12}^1 = d_{12}^1/2$ and $b_{11}^2 = d_{11}^2 - d_{12}^1/2$, and then more generally gives

$$b_{ij}^i = d_{ii}^j/2 , \quad b_{ii}^j = d_{ij}^i - d_{ii}^j/2 . \quad (4.7)$$

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

$$d_{23}^1 = b_{13}^2 + b_{12}^3$$

$$d_{13}^2 = b_{23}^1 + b_{12}^3$$

$$d_{12}^3 = b_{23}^1 + b_{13}^2$$

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

$$b_{j k}^i = \frac{1}{2}(d_{ik}^j + d_{ij}^k - d_{jk}^i) . \quad (4.8)$$

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

$$d_{j_1 j_2}^\alpha = 0, \dots, d_{j_1 \dots j_{m-1}}^\alpha = 0 \quad (4.9)$$

and we seek an m th order reparameterization

$$\theta_\alpha = \tilde{\theta}_\alpha + b_{j_1 \dots j_m}^\alpha \tilde{\theta}_{j_1} \cdots \tilde{\theta}_{j_m} / m! \quad (4.10)$$

so that the new $d_{j_1 \dots j_m}^\alpha$ are equal to zero. As a preliminary we note that the re-expression (5.2) has no effect on lower order discrepancies (5.1). The Appendix A outlines the argument that shows linear equations can be solved to give the b^α 's 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 development of the power series for $\beta(\theta)$ in terms of θ . Such a power series would be the multiparameter analog of an expansion for the scalar $\beta(\theta)$ recorded in (1.6).

Appendix A. We now show that the coefficients $b_{j_1 \dots j_m}^\alpha$ in the expansion (5.2) for θ can be solved for so that the m th order discrepancies $d_{j_1 \dots j_m}^\alpha$ are equal to zero.

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

$$a_{\alpha j_1 \dots j_m} - b_{j_1 \dots j_m}^\alpha - m b_{\alpha j_1 \dots j_{m-1}}^{j_m}$$

and the new $a_{j_1 \dots j_m}^\alpha$ is

$$a_{j_1 \dots j_m}^\alpha + b_{j_1 \dots j_m}^\alpha$$

with the result that the new m -th order discrepancies are

$$d_{j_1 \dots j_m}^\alpha - m b_{\alpha j_1 \dots j_m}^{j_m} . \quad (A.1)$$

Can we choose the $b_{j_1 \dots j_m}^\alpha$ so that these new m -th order discrepancies are all zero?

The discrepancies (A.1) appears 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

$$d_{1 \dots 1}^\alpha = m b_{1 \dots 1 \alpha}^1$$

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

$$d_{1 \dots 123}^\alpha = (m-2) b_{1 \dots 123 \alpha}^1 + b_{1 \dots 13 \alpha}^2 + b_{1 \dots 12 \alpha}^3$$

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

$$d_{j_1 \dots j_r}^\alpha = m_1 b_{j_1 \dots j_r \alpha}^{j_1} + \dots + m_r b_{j_1 \dots j_r \alpha}^{j_r} \quad (A.2)$$

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, \dots , j_r appear m_r times where $j_1 < \dots < j_r$ and $\sum m_i = m + 1$. We consider the r different equations that use this collection of integers for superscript and subscripts. Specifically we take $\alpha = j_1$ in the $d_{j_1 \dots j_r}^\alpha$ equation, \dots , $\alpha = j_r$ in the $d_{j_1 \dots j_r}^\alpha$ equation; the Jacobian of the r equations in the r different b 's is

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

Accordingly we can solve for the b 's:

$$b_{j_1 \dots j_r}^{j_i} = \frac{J_i}{J} \quad (A.4)$$

where J_i is the determinant J but with the i th 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 $r = 1, \dots, m + 1$ and various integers $j_1 < \dots < j_r$ with various frequencies m_1, \dots, m_r with $\sum m_i = m + 1$ we determine the m th order parameter adjustment to give the m th 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**, 343-65.
- Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N. & 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. & Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning & Inference* **66**, 211-222. By processing error this was also published as Cakmak et al (1995).
- Cakmak, S., Fraser, D.A.S., and Reid, N. (1994). Multivariate asymptotic model: exponential and location approximations. *Utilitas Mathematica* **46**, 21-31.
- Cox, D. and Snells, E.J. (1991). Applied Statistics. London: Chapman and Hall.
- 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., Monette, G., Ng, K.W., 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., Reid, N. and Wu, J. (1999). A simple general formula for frequentist and Bayesian inference, *Biometrika* **86**, 249-264.
- 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.