Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 2, pp. 335-346
Bangladesh

PARAMETER CURVATURE REVISITED AND THE BAYES-FREQUENTIST DIVERGENCE

A.M. FRASER
Department of Mathematics, University of British Columbia, Vancouver, Canada V6T 1Z2
Email: afraser@math.ubc.ca

D.A.S. FRASER*
Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3
Email: dfraser@utstat.toronto.edu

M.J. FRASER
Department of Computer Science, University of Chicago, Chicago, IL 60637
Email: maia@cs.uchicago.edu

SUMMARY

Parameter curvature was introduced by Efron (1975) for classifying curved exponential models. We develop an alternative definition that describes curvature relative to location models. This modified curvature calibrates how Bayes posterior probability differs from familiar frequency based probability. And it provides a basis for then correcting Bayes probabilities to agree with the reproducibility traditional to mainstream statistics. The two curvatures are compared and examples are given.

Bayes calibration; Bayes-frequentist discrepancy; Efron curvature; Exponential model approximation; Location model approximation.

AMS Classification: 62F12, 62F05, 62F15.

1 Introduction

Consider a statistical model $f(y_1, y_2; \mu_1, \mu_2)$ centered at (μ_1, μ_2) with standard Normal errors (z_1, z_2). An interest parameter say $\psi_1(\mu_1, \mu_2) = \mu_1$ has contours or level curves that are straight lines on the space $\{(\mu_1, \mu_2)\}$. And an interest parameter $\psi_2(\mu_1, \mu_2) = \mu_1 + \gamma\mu_2^2/2$ has contours that are curves: the standard curvature of such a contour at a point with $\mu_2 = 0$ is the second derivative $(\partial^2/\partial\mu_2^2)\psi|_{\mu_2=0} = \gamma$. This is the reciprocal of the radius of curvature $\rho = 1/\gamma$ of the best fitting circle to that contour at the point with $\mu_2 = 0$; see the Appendix for some further details. The example also displays the first-order asymptotic form of a two-parameter statistical model relative to some antecedent sample size n. For such models, Efron (1975) introduced a measure of curvature to describe how the ψ-fixed model differs from pure exponential model form, where Efron notes "nice properties for estimation, testing, and other inference" methods are available. For this simple example the preceding measures of curvature are equal. And for extensive details on the slightly more complicated Normal (μ, σ) case see Fraser & Sun (2010), which provided the stimulus for the developments in this paper.

We develop an alternative curvature measure that assesses how the ψ-fixed model differs from pure location model form; and in doing this, the new measure determines how posterior probability departs from the reproducibility traditional in main stream statistics. For the simple normal example just mentioned, the two definitions of curvature are in agreement, but they typically differ when the standardized error is not Normal. For purposes here we also include a sign with the curvature, a sign that is positive if the contour is blunt in the direction of increasing ψ and is negative in the reverse case; the sign indicates whether the standard Bayes computation overshoots or undershoots the usual frequentist reproducibility.

The exponential model form considered by Efron, where "good statistical properties" for estimation and inference are available, is **not now** a limitation on the availability of good inference procedures: indeed recent saddlepoint-type analysis has widely extended inference theory to the general likelihood setting and has provided highly accurate p-values for general interest parameters; for recent discussion and extension to the general discrete contingency table context, see Reid & Fraser (2010), Fraser, Wong & Sun (2009), Davison, Fraser & Reid (2006).

An overt challenge in inference theory arises, however, when Bayes calculations of probability are in direct conflict with routine confidence reproducibility, and yet the power of the word probability, here deemed inappropriate in the Bayes setting, eclipses the less aggressive term confidence; indeed other arguments are even introduced citing Lindley's (1958) view that confidence distributions are wrong if they do not correspond to just likelihood combination as with Bayes.

Consider a parameter ρ for a variable r with distribution function $F(r; \rho)$. The confidence p-value from data r^0 is $p(\rho) = F(r^0; \rho)$, and it records just the %-age position of the data with respect to the parameter value ρ; and when the p-value is then used to form confidence quantiles they have the usual Neyman (1937) interpretation under repetitions. Meanwhile for the Bayes case, the posterior survivor value $s(\rho)$ is the right tail distribution function for the posterior of ρ using a suitable default prior. We would of course hope that the Bayes alleged probability bears some sensible connection with frequentist reproducibility.

Now consider the initial example involving a Normal on the plane. For the obviously linear parameter μ_1 the p-value is obtained immediately from the corresponding variable y_1 and is given as $p(\psi) = \Phi(y_1^0 - \mu_1)$ where Φ is the standard Normal distribution function. For a Bayes calculation we would need of course an appropriate default prior; the obvious prior fully in accord with Bayes (1763) is the flat prior on the plane. This immediately says that the posterior for (μ_1, μ_2) is the standard Normal located at the data point (y_1^0, y_2^0) and then that the posterior for μ_1 is the standard Normal located at y_1^0, which leads routinely to the survivor probability $s(\mu_1) = \Phi(y_1^0 - \mu_1)$. This Bayes survivor value is of course equal to the frequentist p-value.

But now for a curved parameter say $\rho = (\mu_1^2 + \mu_2^2)^{1/2}$ with curvature $\gamma = 1/\rho$ we have the obvious $r = (y_1^2 + y_2^2)^{1/2}$ which measures ρ. The p-value for assessing ρ is then $H_2\{(r^0)^2; \rho^2\}$ where $H_2(r^2; \rho^2)$ is the non-central chi square distribution function with degrees of freedom 2 and non-centrality ρ^2. By contrast the Bayes survivor value is $1 - H_2\{\rho^2; (r^0)^2\}$: The p-value $p(\rho)$ and the Bayes value $s(\rho)$ are not equal. Indeed the second is uniformly larger than the first (Fraser & Reid, 2002; Fraser, 2010); this is easily verified by starting with the linear case where they are equal and then, with $\rho - r$ fixed, morphing into the second case with changing γ and noting that the sample space sets on the 2-dimensional space are decreasing for the p-value calculation and increasing for the s-value calculation.

A location model $f\{y - \beta(\theta)\}$ has linearity for the parameter on the sample space, and an exponential model $exp\{\varphi'(\theta)s(y) - k(\theta)\}h(y)$ has linearity in the exponent over the sample space, where the p-dimensional vectors φ, s are combined linearly. The mapping between the two spaces is not linear and accordingly linearity of parameters is different in the two spaces. We develop the curvature measure for the sample space somewhat following the Efron route for parameters in the exponent. In either case a local standardization is needed: Efron uses expected information but we follow current likelihood directions and use observed information calculated in the locally defined canonical parameterization. In exponential models these informations are in agreement; and in non exponential models the use of the special observed information function provides the critical ingredient for third-order inference accuracy.

For a model with data, let $\ell(\theta)$ be the observed log-likelihood and let $\varphi(\theta)$ be the log-likelihood gradient in directions tangent to an exact or approximate ancillary having surface dimension equal to the dimension say p of the parameter (for example, Fraser, Fraser & Staicu, 2010). And then let $\{\ell(\theta); \varphi(\theta)\}$ designate the exponential model:

$$g(s;\theta) = \exp\{\ell(\theta) + \varphi'(\theta)s\}h(s) = \frac{\exp\{k/n\}}{(2\pi)^{p/2}} \exp\{\ell(\varphi) - \ell(\hat{\varphi})\}|\jmath_{\varphi\varphi}(\hat{\varphi})|^{-1/2},$$

where $\hat{\theta}$ and $\hat{\varphi} = \varphi(\hat{\theta})$ depend on s and are calculated from the tilted likelihood in the middle expression and where the right hand expression is the saddlepoint approximation for the middle expression involving a constant k and the observed information $\jmath_{\varphi\varphi}(\hat{\varphi}) = (\partial/\partial\varphi)(\partial/\partial\varphi')\ell(\theta)|_{\hat{\varphi}(s)}$ where the derivatives are of $\ell(\theta)$ with respect to $\varphi(\theta)$; the related observed $s^0 = 0$ derives from having the score variable s centered at the observed data point. This exponential model is a first derivative approximation to the original model and yet provides full third order inference for arbitrary scalar interest parameters; for recent discussion see, for example, Reid & Fraser (2010). We work entirely within this exponential model; it agrees with the original model if the original is exponential and more generally it is a first derivative approximation to that model but retains full third order inference reliability.

The data-dependent canonical parameterization $\varphi(\theta)$ is the essential ingredient for extending saddlepoint technology from cumulant generating function contexts to the general asymptotic context (Fraser & Reid, 1995; Reid & Fraser, 2010; and an overview in Fraser, Wong & Sun, 2009). The tangents to an approximate ancillary are given by the $n \times p$ array

$$V = (v_1, \ldots, v_p) = \left. \frac{dy(\theta, u)}{d\theta} \right|_{y^0, \hat{\theta}^0},$$

where y is written in quantile form $y(\theta, u)$ as a function of the parameter θ and the vector u of p-values; the vector $u = F(y; \theta)$ of coordinate distribution functions gives the vector of p-values and its inverse is the vector of quantiles $y(\theta, u)$. See for example Fraser, Fraser & Staicu, (2010). This gives the formal definition of the approximating canonical parameter as $\varphi(\theta) = (d/dV \ell(\theta; y))|_{y^0}$, as the gradient in essential directions V of the log-model at the observed data; this can be calculated easily and accurately by differencing if not by the more obvious differentiation.

In our initial Normal error example we considered two scalar parameters, one linear and one curved, and allowed that, with a change in the parameterization, the curvature of a parameter could change. For the more general asymptotic model we then have a similar result and will be interested in curvature at the observed maximum likelihood value, say $\hat{\theta}^0$ in the initial parameterization. For this we will use the standardized metric provided by the observed information $\jmath^0_{\theta\theta} = \jmath_{\theta\theta}(\hat{\theta}^0)$, that is,

$$ds^2 = (\theta - \hat{\theta}^0)' \jmath^0_{\theta\theta} (\theta - \hat{\theta}^0) = d\theta' \jmath^0_{\theta\theta} d\theta,$$

which records local departure standardized to observed information and presented here in the given parameterization. For other parmeterizations this transforms by change of variable, that is, from the θ parameterization to the new parameterization. It then follows that the connection between two standardizations is given locally by an orthogonal transformation. And in the $p = 2$ case if the standardizations are aligned with a scalar interest parameter, as will be discussed, then the orthogonal parameterization is either an identity transformation or a reflection for the second coordinate. As a consequence any change in the curvature of a scalar parameter is attributable entirely to the second derivative array between the standardized parameterizations.

In Section 3 we show that the aligned and standardized second derivative arrays, say W^{ij} from parameterization say θ^i to θ^j, satisfy the convenient property $W^{ik} = W^{ij} + W^{jk} = -W^{ki}$. And in addition if W is the array from parameterization θ^1 to θ^2 and is aligned with the interest parameter ψ then the change in curvature from γ_1 to γ_2 for the interest parameter ψ is given as $\gamma_2 = \gamma_1 - w^1_{22}$, where w^1_{22} in W is the second partial of the first new coordinate with respect to the second old coordinate; this agrees directly with our calculations for the two interest parameters in the initial example. For some background on the Bayes-frequentist divergence and the feasibility of the present curvature implementation, see Fraser & Sun (2010).

2 Change of Parameterization

Consider a statistical model $f(y;\theta)$ having an alternative parameterization $\varphi(\theta)$, and suppose we are interested in the curvature of the scalar interest parameter $\psi(\theta)$. Curvature gets more complicated quickly with larger parameter dimension p (Fraser, Fraser & Staicu, 2010), so we give details for just the primary case $p = 2$. In an application, curvature at the observed maximum likelihood value can be of central concern; accordingly, we investigate the change of parameterization from θ to $\varphi(\theta)$ in the neighborhood of the maximum likelihood value $\theta^0 = \hat{\theta}(y^0)$, and we do this in standardized units relative to observed information; this observed information is properly calculated from the tangent exponential model $g(s;\theta)$ and thus from the log-likelihood $\{\ell(\theta) + \varphi'(\theta)s\}$ and not from the original model; but at $s = 0$ corresponding to the maximum for $\ell(\theta)$ the results are the same. Our objective is to find out how the directly calculated curvature in the initial parameterization gets modified by the change to the new parameterization; this is of substantial importance in the calibration of Bayes procedures relative to repetition validity.

For some given parameterization say θ, we first center at the data-indicated maximum likelihood value $\theta^0 = \hat{\theta}^0$ and thus address the departure

$$\bar{\theta} = \begin{pmatrix} \theta_1 - \theta_1^0 \\ \theta_2 - \theta_2^0 \end{pmatrix}.$$

We then align this with the scalar interest parameter ψ so that the first new coordinate changes locally like the scalar ψ; this gives us a linear transformation say

$$\vec{\theta} = A\bar{\theta} = \begin{pmatrix} \psi_1^0 & \psi_2^0 \\ -\psi_2^0 & \psi_1^0 \end{pmatrix} \bar{\theta},$$

where $\psi_1^0 = \partial\psi/\partial\theta_1|_{\theta^0}$, $\psi_2^0 = \partial\psi/\partial\theta_2|_{\theta^0}$ are the partial derivatives of ψ at the value θ^0. And then for standardization let

$$T = \begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

be the positive lower triangular **right** root of the observed information $\jmath_{\vec{\theta}\vec{\theta}} = T'T$ in the parameterization $\vec{\theta}$, and define the standardized parameter departure

$$\tilde{\theta} = T\vec{\theta} = TA\bar{\theta} = B\bar{\theta},$$

where $B = TA$; this retains the centering and alignment but in addition gives an identity observed information for the modified $\tilde{\theta}$. Also we let B_1 be the first row vector of B.

Now for the scalar parameter $\psi(\theta)$ we obtain in a similar way the centered and standardized version $\tilde{\psi}(\theta) = t_{11}(\psi(\theta) - \hat{\psi}^0)$. Then for the scalar parameter $\psi_2(\theta)$ in the initial example of Section 1 we obtain the curvature of ψ_2 with respect to θ as $\gamma = (\partial^2\tilde{\psi}/\partial\tilde{\theta}^2)$. Also for some alternative parameterization say $\varphi(\theta)$ we can similarly center, align and standardize to go from $\bar{\varphi}(\theta)$ to $\tilde{\varphi}(\theta)$ by replacing $\bar{\theta}$ in the preceding steps by

$$\varphi_\theta^{-1}\begin{pmatrix}\varphi_1 - \varphi_1^0 \\ \varphi_2 - \varphi_2^0\end{pmatrix},$$

where φ_θ is the Jacobian $\partial\varphi/\partial\theta|_{\theta^0}$, and then using the consequent $\tilde{\theta}$ as the centered, aligned and standardized $\tilde{\varphi}$. These transformation A, T, B, and φ_θ^{-1} are constant and linear and thus semi-transparent for second derivative computations; further details are given in Section 6.

3 Taylor Expansion between Parameterizations

Consider an initial parameterization θ and some alternative parameterization $\varphi(\theta)$; we examine how these relate to each other in the neighborhood of the maximum likelihood value θ^0. We of course center, align and standardize as just described; accordingly we then work with the modified versions developed in the preceding section; and for convenience of notation we also omit the tildes and thus use the notation θ as a simple substitute for $\tilde{\theta}$. We Taylor expand φ in terms of θ and obtain

$$\begin{pmatrix}\varphi_1 \\ \varphi_2\end{pmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix} + \frac{1}{2n^{1/2}}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}'\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22}\end{bmatrix}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix} + \cdots,$$

where the v_i and w_{ij} in the $[v_i]$ and $[w_{ij}]$ arrays are vectors in R^2 and are combined linearly by the indicated matrix multiplication. In particular the v array designated say V takes the form of a 2×2 identity matrix due to the standardization. And the w array say W records second derivatives w_{ij} of the new vector parameter φ with respect to coordinates θ_i and θ_j of the the initial parameter θ; thus $w_{ij} = (\partial^2\varphi(\theta)/\partial\theta_i\partial\theta_j)|_{\hat{\theta}^0}$, but all are in the aligned and standardized version of the coordinates; in particular W could be presented as a $2 \times 2 \times 2$ array; the observed information used in the standardization is of order n as part of dependence on a deemed antecedent sample size n and when used to standardize the first derivatives leads to an effect $n^{-1/2}$ in each of the second derivatives; we occasionally make this explicit in the equations. And we do emphasize that these calculations are in terms of the standardized coordinates.

Consider a first parameterization say θ^1 and a second parameterization say θ^2; the corresponding Taylor expansion as above would have a second derivative array say W^{12} for the change of parameterization. And suppose we have a third parameterization say θ^3 with an array W^{23} for the standardized quadratic array relative to θ^2. The simple substitution and retention of terms of order $O(n^{-1/2})$ gives $W^{13} = W^{12} + W^{23}$ and similarly $W^{12} = -W^{21}$.

We return to the initial example in the Introduction and the curved interest parameter $\psi(\theta_1, \theta_2) = \theta_1 + \gamma\theta_2^2/2$. We can picture ψ increasing as we move positively along the axis for θ_1; and then with positive γ we would have contours that are blunt nosed to the right, and with negative γ we would have blunt nosed to the left. Now suppose we make a change of parameterization involving a curvature change array W. The initial curve through the origin can be written in explicit form as

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} -\gamma t^2/2n^{1/2} \\ t \end{pmatrix}$$

using a mathematical parameter t. If we substitute this in the Taylor series above we obtain

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} -\gamma t^2/2n^{1/2} \\ t \end{pmatrix} + w_{22}\frac{t^2}{2n^{1/2}} = \begin{pmatrix} -\gamma t^2/2n^{1/2} + w_{22}^1 t^2/2n^{1/2} \\ t + w_{22}^2 t^2/2n^{1/2} \end{pmatrix}, \tag{3.1}$$

where w_{22}^1 and w_{22}^2 are the first and second coordinates of the second derivative vector w_{22}, and again t is a free parameter that generates the contour. The calculations omit terms of order $O(n^{-1})$ that arise in expanding the matrix array; and the quadratic addition to t of course alters the free parameter but does not alter the curvature to the order $O(n^{-1})$. It follows that the new curvature is just $\gamma_2 = \gamma - w_{22}^1$; this does include the usual orthogonalization adjustment applied in the calculation of acceleration vectors. Thus we see that the first coordinate w_{22}^1 of the aligned and standardized vector w_{22} provides the correction to the curvature calculated in the initial parameterization, but for this the parameters need to be in the aligned and standardized form.

In other words we can calculate the curvature in any convenient parameterization and afterwards correct to that in the particular standardized parameterization of interest; and this just needs the second derivative array to the new parameterization. In the next two sections we consider the two parameterizations of particular interest here, the exponential parameterization discussed by Efron (1975) and the linear parameterization needed for understanding the positive and negative aspects of Bayes methodology.

4 The Exponential Parameterization

Consider an exponential model $\exp\{\varphi'(\theta)s(y) - k(\theta)\}h(y)$ as discussed in the Introduction; this can be the actual model in an application or it can be an approximate exponential model as also described there. With a canonical parameter $\varphi(\theta)$ of dimension 2 and a primary parameter θ of dimension 1 we have a $(2,1)$-exponential model; Efron (1975) examined curvature for such models. For these models the canonical parameterization $\varphi(\theta)$ extends naturally to give a full 2 dimensional parameter called the natural parameter of the model, and the extended model is a $(2,2)$-exponential model. There may also be a convenient extension of the initial parameter θ; in such a case we might calculate curvature in the extended initial parameterization and then correct it using the second derivative array between parameterizations; this would give us the Efron curvature but would also give a sign for the curvature, a sign that may not be of immediate use for the "estimation, testing, and other inference" methods addressed by Efron.

A crucial step for determining the exponential parameterization φ outside a pure exponential model involves the sample space directions V that establish approximate conditioning. These directions V are essential for going beyond second order inference unless the model has the symmetry of the Normal or closely related special models. The sample space directions are also crucial for determining location parameterization which in turn allows the determination of the bias typical in familiar Bayesian calculations: the sign of the curvature indicates whether the Bayes calculation will give values that exceed or fall short of the reproducibility implied by the term probability. We next develop the location parameterization.

5 The Location Parameterization

For a location model $f(y_1 - \theta_1, y_2 - \theta_2)$ we consider how a change δ in the first variable at its observed value y_1^0 relates to the form of the model: a parallel increase δ in θ_1 at any value whatsoever leaves the model unchanged,

$$f\{y_1^0 + \delta - (\theta_1 + \delta), y_2 - \theta_2\} = f(y_1^0 - \theta_1, y_2 - \theta_2);$$

and the same also holds to first derivative at any arbitrary value $y_1 = y_1'$. This presents θ_1 as a location parameter for the first variable. We extend this in approximate form to regular models as considered in the Introduction. For this we use the quantile form $y(\theta, u)$ of the model; this presents the response y in terms of the parameter and p-value and provides an mathematical equivalent to the use of coordinate distribution functions.

To find out how θ affects the data at the observed value y^0 we differentiate $y = y(\theta, u)$ with respect to θ and at the observed data y^0 obtain

$$V(\theta) = \frac{d}{d\theta} y(\theta, u)|_{y^0}.$$

and thus $dy = V(\theta)d\theta$. And then in turn to find out how y affects $\hat{\theta}(y)$ we differentiate the score equation $\ell_\theta(\theta; y) = 0$ obtaining

$$\ell_{\theta\theta'}(\hat{\theta}^0; y^0)\partial\hat{\theta} + \ell_{\theta;y'}(\hat{\theta}^0; y^0)\partial y = 0,$$

where the differentials $d\hat{\theta}$ and dy are respectively $p \times 1$ and $n \times 1$ and $\ell_{\theta;y'}(\hat{\theta}^0; y^0) = (\partial/\partial\theta)(\partial/\partial y')\ell(\theta; y) = H'$ is the gradient of the score function at the data point. Then solving for $d\hat{\theta}$ we obtain

$$d\hat{\theta} = \hat{\jmath}^{-1} H' dy,$$

where $\hat{\jmath} = j(\hat{\theta}^0; y^0) = -\partial^2\ell(\hat{\theta}^0; y^0)/\partial\theta\partial\theta'$ is the observed Fisher information. Combining these gives

$$d\hat{\theta} = \hat{\jmath}^{-1}H'V(\theta)d\theta = M(\theta)d\theta = \{m_1(\theta), m_2(\theta)\}d\theta, \tag{5.1}$$

where in the final expression we are restricting consideration to the $p = 2$ dimensional case. This records how change $d\theta$ in the parameter θ at various θ values influences the maximum likelihood value $\hat{\theta}$ at the observed data. These results build on analysis in Fraser & Reid (1995) and Fraser, Fraser & Staicu (2010).

We now examine the equation (5.1) to determine how a change in $\hat{\theta}$ at its observed value relates to change in the parameter at various θ values. More specifically we seek to integrate the right side from the observed $\hat{\theta}^0$ to a point θ in moderate deviations about the observed and do this to second order. And for this we follow related analysis (Fraser, Fraser & Staicu, 2010) where parameter change generates a Taylor series approximation to an intrinsic second order ancillary contour on the sample space. Accordingly we first express the right side $M(\theta)d\theta$ of (5.1) in a first order Taylor series in the departure $\delta = \theta - \hat{\theta}^0$ from the observed $\delta^0 = 0$, and then formally substitute in (5.1) to obtain the following expression for a location parameterization $\beta(\theta)$:

$$\beta(\theta) = \begin{bmatrix} m_1(\hat{\theta}^0) & m_2(\hat{\theta}^0) \end{bmatrix} (\theta - \hat{\theta}^0) + \frac{1}{2n^{1/2}}(\theta - \hat{\theta}^0)' \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} (\theta - \hat{\theta}^0) + \cdots ,$$

where $m_{ij} = (\partial m_i(\theta)/\partial\theta_j)|_{\hat{\theta}^0}$, and the m_i-vectors are given by (5.1).

A direct integration of (5.1) from $\theta = \hat{\theta}^0$ to θ in moderate deviations depends on the integration path chosen, and the expression just recorded corresponds to radial integration from $\delta = 0$ to a general δ in moderate deviations, thus from $(0,0)$ to $t(\theta_1 - \hat{\theta}_1^0, (\theta_2 - \hat{\theta}_2^0)$ using t going from 0 to 1. This involves a parameter curvature array $M = \{m_{ij}\}$ which can be symmetrized and this array for the location parameterization here will usually be different from the curvature array say $W = \{w_{ij}\}$ for exponential properties developed in Section 4. The present array is then a second derivative matrix of the change from the θ parameterization to the location parameterization $\beta(\theta)$. And this will give the curvature correction for adjusting an initial curvature of a scalar parameter ψ to the curvature relative to β parameterization. There is the possibility of path dependency but this does not affect the curvature obtained here using the radial integration path. We thus obtain the curvature characteristic m_{22}^1 for adjusting the curvature γ in the initial parameterization θ to the curvature $\gamma - m_{22}^1$ in the developed location parameterization as discussed after (3.1). An example is discussed in full detail in Fraser & Sun (2010).

6 Calculating Curvature

Consider a scalar interest parameter $\psi(\theta)$ together with a regular statistical model $f(y;\theta)$ having observed data y^0 and 2-dimensional parameter θ. We need of course the observed maximum likelihood value $\hat{\theta}^0$ and then the centered, aligned and standardized version $\tilde{\theta} =$

$B(\theta - \hat{\theta}^0)$ of the parameter where B is the aligning and standardizing matrix for θ recorded in Section 2. We can then express $\psi(\theta)$ in terms of the modified $\tilde{\theta}$ as $\psi(\hat{\theta}^0 + B^{-1}\tilde{\theta})$ and then standardize it obtaining $\tilde{\psi}(\tilde{\theta}) = B_1\psi(\hat{\theta}^0 + B^{-1}\tilde{\theta})$. The curvature γ of ψ in the given θ parameterization is then the second derivative of $\tilde{\psi}$ with respect to $\tilde{\theta}_2$ at $\tilde{\theta} = 0$.

For the Efron curvature but based on observed rather than expected information we need first the canonical parameterization $\varphi(\theta)$. This is available as the gradient $\varphi(\theta) = \{d/dV\}\ell(\theta; y)|_{y^0}$ of log-likelihood at the observed data y^0, calculated in the quantile movement directions $V = dy(\theta, u)/d\theta|_{(y^0,\hat{\theta}^0)}$. This can be centered, aligned and standardized as $\tilde{\varphi}(\theta) = B\varphi_\theta^{-1}\{\varphi(\theta) - \hat{\varphi}^0\}$ where φ_θ is the Jacobian $\partial\varphi/\partial\theta|_{\theta^0}$ evaluated at the observed $\hat{\theta}^0$. We then determine the second derivative array say $\tilde{W}_{12}$ of $\tilde{\varphi}(\theta)$ with respect $\tilde{\theta}$. As the connection is linear, we can calculate this from the second derivatives w_{ij} for the initial parameters :

$$\tilde{W}_{12} = \frac{\partial^2\tilde{\varphi}(\theta)}{\partial\tilde{\theta}^2}\Big|_{\hat{\theta}^0} = (B^{-1})' \begin{bmatrix} B\varphi_\theta^{-1}w_{11} & B\varphi_\theta^{-1}w_{12} \\ B\varphi_\theta^{-1}w_{21} & B\varphi_\theta^{-1}w_{22} \end{bmatrix} B^{-1}.$$

The exponential curvature γ^{exp} is then the curvature γ obtained in Section 3 and adjusted $\gamma_2^{\text{exp}} = \gamma - w_{22}^1$ using the first coordinate $w_{22}^1$of the aligned and standardized vector w_{22}.

For our proposed location curvature we have derived a location parameter contour in Section 5; it uses two vectors $m_1(\theta), m_2(\theta)$ given in (5.1). The basic location curvature array (m_{ij}) is the gradient with respect to θ of these two vectors evaluated at $\hat{\theta}^0$ and recorded after (5.1). The array is typically different from that for the the exponential parameterization array just considered. We then need the second derivative array say $\tilde{M}_{12}$ of $\tilde{\beta}(\theta)$ with respect $\tilde{\theta}$. As the connection is linear, we can calculate this from the second derivatives m_{ij} obtained after (5.1):

$$\tilde{M}_{12} = \frac{\partial^2\tilde{\beta}(\theta)}{\partial\tilde{\theta}^2}\Big|_{\hat{\theta}^0} = (B^{-1})' \begin{bmatrix} B\varphi_\theta^{-1}m_{11} & B\varphi_\theta^{-1}m_{12} \\ B\varphi_\theta^{-1}m_{21} & B\varphi_\theta^{-1}m_{22} \end{bmatrix} B^{-1}.$$

The location curvature γ^{loc} is then the curvature γ obtained in Section 5 and adjusted $\gamma_2^{\text{loc}} = \gamma - m_{22}^1$ using the first coordinate $m_{22}^1$of the aligned and standardized vector m_{22} derived in Section 5.

7 Examples

The feasibility of the present material was explored in Fraser & Sun (2010) and was exhibited for the familiar reference example, the Normal (μ, σ^2) sampling model. In that reference the parameters μ and σ are shown to be linear and thus have $\gamma = 0$; also any linear combination of them is also shown to be linear. To encounter a curved parameter we only need to look to the familiar first exponential canonical parameter μ/σ^2, which we now designate ψ.

For illustration consider a data point y^0, with $\hat{\mu} = \bar{y}^0 = 0.975442, \hat{\sigma} = \sqrt{(n-1)/n}s_y^o =$ 1.226137 and with $n = 3$. The maximum likelihood value of $\hat{\mu}/\hat{\sigma}^2 = \hat{\psi}$ is 0.6488188. For present purposes, it would be helpful to be able to write the model in term of a location parameterization, say as (β_1, β_2), but what would be the basis for such a parameterization? Large sample analysis (Cakmak, Fraser & Reid 1994) indicates that there is a location model as in Section 1 but such does not address the regression type structure with our present Normal (μ, σ^2) example; even with this change there are consequences that do not follow the direct location model pattern; for some differential consequences indicated by the sample to parameter space mapping function $M(\theta)$; see Fraser & Reid (1995).

For the present data the observed curvature $\hat{\gamma}^0 = 0.1346722$ and the radius of curvature $\hat{\rho}^0 = 7.425437$. For some further discussion including iterative computation for the present curvature see Fraser & Sun (2010). Continuing research will explore the use of this curvature measure as a device to correct routine Bayes calculations to give them repetition validity.

8 Acknowledgements

Support from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. Some portions of this material were reported at the Canadian Mathematical Society Winter Meeting in Windsor Canada on December 6, 2009; support from the organizer Prof. S. Ejaz Ahmed is deeply appreciated.

References

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

[2] Davison, A.C., Fraser, D.A.S. and Reid, N. (2006). Improved likelihood inference for discrete data. *J. Royal Statist. Soc. Series B* **68**, 495-508.

[3] Efron, B. (1975). Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency). *Annals of Statistics,* **6**, *1189-1242.*

[4] Fraser, D.A.S. (2010). Is Bayes posterior just quick and dirty confidence? *Statistical Science*; in review.

[5] Fraser, A.M., Fraser, D.A.S. and Staicu, A.-M. (2010). The second order ancillary: A differential view with continuity. *Bernoulli*, to appear.

[6] Fraser, D.A.S., and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

[7] Fraser, D.A.S., and Reid, N. (2002). Strong matching of frequentist and Bayesian parametric inference. *Journal of Statistical Planning and Inference.* **103**, 263-285.

[8] Fraser, D.A.S., Reid, N., Marras, E., and Yi, G.Y. (2010). Default priors for Bayesian and frequentist inference. *J. Royal Statist. Soc. B*, to appear.

[9] Fraser, D.A.S., Wong, A. and Sun, Ye (2009). Three enigmatic examples and inference from likelihood. *Canadian Journal of Statistics*, **37**, 1-21.

[10] Fraser, D.A.S. and Sun, Y. (2010). Some corrections for Bayes curvature. *Parkistan J. of Statist.*; **25**, 351-370.

[11] Lindley, D.V. (1958). Fiducial distribution and Bayes theorem. *J. Roy. Statist. Soc.* B **20**, 102-107.

[12] Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil.Trans. Roy. Soc. A***237**, 333-380.

[13] Reid, N, and Fraser, D.A.S. (2010). Mean likelihood and higher order approximations. *Biometrika* **97**, 159-170.

A Radius of curvature example

Consider the circle with center $(-\rho, 0)$ and radius ρ; it corresponds to the following function taking the value ρ

$$R = \{(\rho + y_1)^2 + y_2^2\}^{1/2} = \rho\{1 + \frac{y_1}{\rho} + \frac{y_2^2}{2\rho^2} + \cdots\} = \rho + y_1 + \frac{y_2^2}{2\rho} + \cdots,$$

where y_1, y_2 are viewed as $O(n^{-1/2})$, and this agrees locally with the interest parameter ψ_2 at the beginning of Section 1.