Chapter 6

Curvature Interpretations

In this section some of the fundamental quantities encountered in Chapter 4 will be interpreted in terms of certain basic model curvatures. The case where $r = 1$, that is ψ univariate will be considered. This is in keeping with much of the current literature, and facilitates comparisons with Efron (1985). Since ψ is univariate, the previous need for a superscript notation to denote the components of the projective scores is redundant since U_0 has only a single component.

6.1 Asymptotic bias interpretation

The bias term in the asymptotic distribution of the $U_1(\hat{\lambda})$ estimating equation is (see Theorems (4.2) and (4.3)),

$$-\tfrac{1}{2}\mathrm{tr}(T^* W^{-1}),$$

and the aim of this section is to provide a simple interpretation for this bias.

Since $\{V_{ij}^*\}$ is a second order covariant tensor with respect to invertible transformations of λ, so too is $T_{ij}^* = E[U_0 V_{ij}^*]$. Furthermore, $W^{-1} = W^{i,j}$ is a second order contravariant tensor. Therefore, $\mathrm{tr}(T^* W^{-1}) = T_{ij}^* W^{j,i}$ is a contraction of a tensor to a scalar, and hence *invariant* under reparameterization of λ. To demonstrate this invariance explicitly, consider a transformation from λ to μ, and let J denote the Jacobian of the transformation. The Bhattacharyya basis functions in the transformed coordinate system will be denoted by a superscript $\tilde{}$. It is straight forward to show that the following relationships hold between the Bhattacharyya functions in the original and transformed coordinate systems:

$$
\begin{aligned}
\widetilde{\mathbf{V}}_1 &= J\mathbf{V}_1. \\
\widetilde{\mathbf{V}}_2^* &= J\mathbf{V}_2^* J^T.
\end{aligned}
$$

In the new coordinate system,

$$
\begin{aligned}
\mathrm{tr}(\widetilde{T}^*\widetilde{W}^{-1}) &= \mathrm{tr}(JT^*J^T(JWJ^T)^{-1}) = \mathrm{tr}(JT^*J^TJ^{T-1}W^{-1}J^{-1}) \\
&= \mathrm{tr}(JT^*W^{-1}J^{-1}) = \mathrm{tr}(T^*W^{-1}J^{-1}J) = \mathrm{tr}(T^*W^{-1}),
\end{aligned}
$$

which establishes the invariance of the trace under reparameterization.

Now treating ψ as fixed, consider the hyper-surface H, in the $s+1$ dimensional Riemannian Manifold, M_{s+1} given by the function

$$
H(\lambda) = \begin{pmatrix} E_\lambda[U_1(\lambda_0)] \\ W(\lambda_0)^{-\frac{1}{2}} E_\lambda[\mathbf{V}_1(\lambda_0)] \end{pmatrix}.
$$

It is very useful to introduce the local reparameterization given by

$$
\lambda \to \mu \quad \text{where} \quad \mu = W(\lambda_0)^{-\frac{1}{2}} E_\lambda[\mathbf{V}_1(\lambda_0)]. \tag{6.1}
$$

It will be assumed that the hyper-surface is of class $r \geq 2$ (see Kreyszig (1968, p.277) for a definition of "class r"). The condition amounts to assuming an adequate degree of smoothness of the hyper-surface.

The Jacobian of the transformation from λ to μ is

$$
\left[\frac{\partial \lambda}{\partial \mu}\right] = \left[\frac{\partial \mu}{\partial \lambda}\right]^{-1} = \left[W(\lambda_0)^{-\frac{1}{2}} \frac{\partial}{\partial \lambda} E_\lambda[\mathbf{V}_1(\lambda_0)]\right]^{-1},
$$

but

$$
\frac{\partial}{\partial \lambda} E_\lambda[\mathbf{V}_1(\lambda_0)]|_{\lambda=\lambda_0} = W(\lambda_0),
$$

so that the Jacobian is simply $W^{-\frac{1}{2}}$, (note that this is a symmetric matrix). Since the Jacobian is a positive definite matrix, it follows that its determinant is positive, which means that the sense of the surface remains unaltered by the transformation. The relevance of this is that the transformation does not alter the sign of the principle curvatures of the surface at μ_0.

The importance of the transformation lies in the simplicity and interpretability that it gives to the extended information matrix. In the transformed coordinate system

$$\widetilde{W} = E[\widetilde{\mathbf{V}}_1 \widetilde{\mathbf{V}}_1^T] = E[J\mathbf{V}_1\mathbf{V}_1^T J^T] = JWJ = I,$$

where I is an $s \times s$ identity matrix, so that the transformation from λ to μ is a variance stabilizing transformation in the sense that $\widetilde{W}$ is not a function of μ.

Furthermore, the covariance between $\widetilde{V}_i$ and $\widetilde{V}_{jk}$ is zero because

$$E[\widetilde{V}_i \widetilde{V}_{jk}] = \frac{\partial^2}{\partial \mu^j \partial \mu^k} E_\mu[\widetilde{V}_i(\mu_0)] = W^{-\frac{1}{2}} \frac{\partial^2}{\partial \mu^j \partial \mu^k} \mu = 0.$$

It immediately follows that $\widetilde{V}_{jk}^* = \widetilde{V}_{jk}$, so that in this coordinate system, $\widetilde{T}_{ij}^* = \widetilde{T}_{ij} = E[U_0 \widetilde{V}_{ij}]$.

The hyper-surface $H(\lambda)$, can be treated as a function of μ rather than λ; it is still the same surface, but the alternative parametric representation facilitates a simpler analysis. The hyper-surface is parameterized as

$$H(\mu) = \begin{pmatrix} E_\mu[U_1(\mu_0)] \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_s \end{pmatrix}.$$

The important consequence of this re-parameterisation is that the elements of the matrix $\widetilde{W}$ constitute the tensor corresponding to the *first fundamental form* of $H(\mu)$. The elements of the first fundamental form, $g_{\alpha\beta}$, $\alpha = 1, 2, \cdots, s$, $\beta = 1, 2, \cdots, s$ are obtained by taking the scalar product of the vectors H_α and H_β, where H_α is the derivative of H with respect to μ_α. Now,

$$H_\alpha = \frac{\partial}{\partial \mu_\alpha} \begin{pmatrix} E_\mu[U_1(\mu_0)] \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_s \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \mu_\alpha} E_\mu[U_1(\mu_0)] \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

where the solitary 1 is in the position of the original μ_α. The term $\frac{\partial}{\partial \mu_\alpha} E_\mu[U_1(\mu_0)]$ is equal to zero on account of the fact that U_1 is orthogonal to $\widetilde{V}_\alpha$. Clearly the scalar product of H_α and H_β is equal to 1 if $\alpha = \beta$ and equal to zero otherwise. The benefit of using the new local Cartesian coordinate system is that the *metric tensor* $g_{\alpha\beta}$ in H, is the identity matrix.

Furthermore, in this parameterization, the elements of the *second fundamental form* of H, that is $b_{\alpha\beta}$, can be obtained from the formula

$$b_{\alpha\beta} = H_{\alpha\beta}.\mathbf{n},$$

where $H_{\alpha\beta}$ is the vector $\frac{\partial^2}{\partial \mu_\alpha \partial \mu_\beta} H$, and $\mathbf{n}$ is a unit normal vector to the surface H at μ_0. Here, $\mathbf{n} = (1, 0, 0, \cdots 0)^T$, so that $b_{\alpha\beta} = \frac{\partial^2}{\partial \mu_\alpha \partial \mu_\beta} E_\mu[U_1(\mu_0)]$. This is equal to $E[U_1 \widetilde{V}_{\alpha\beta}] = E[U_0 \widetilde{V}_{\alpha\beta}]$. Consequently the tensor $b_{\alpha\beta}$ is non other than the matrix $\widetilde{T}^*$!

These observations lead to the following theorem:

Theorem 6.1 *The quantity $tr(T^*W^{-1})$ is equal to the sum of the principle curvatures of H at μ_0.*

Proof

First transform to the μ coordinate system given in (6.1).

Following Theorem 115.2 in Kreyszig (1968), the principle curvatures of H at μ_0, denoted by $\kappa_{(k)}$, are the roots of the equation

$$\det(\widetilde{T}_{ij} - \kappa \widetilde{W}_{i,j}) = 0.$$

But $\widetilde{W}$ is the identity matrix, so that the $\kappa_{(k)}$ are simply the eigenvalues of the matrix $\widetilde{T}$. Since $\widetilde{W}$ is an identity, $\mathrm{tr}(\widetilde{T}\widetilde{W}^{-1}) = \mathrm{tr}(\widetilde{T})$, but the trace of a symmetric matrix $\widetilde{T}$ is equal to the sum of the eigenvalues. Consequently $\mathrm{tr}(T^*W^{-1}) = \mathrm{tr}(\widetilde{T}^*) = \sum_{k=1}^{s} \kappa_{(k)}$ as required. ■

The principle curvatures are associated with curves on the surface $H(\mu)$, in directions known as the *principle directions* of the covariant tensor T_{ij}. It is well known that these directions are mutually orthogonal. Furthermore, the sum of the curvatures in s mutually orthogonal directions of a hypersurface is constant, and referred to as the *mean curvature*, M^*, of the surface H. As a particular case $M^* = \sum \kappa_{(k)}$, where $\kappa_{(k)}$ are the principle curvatures. Therefore an alternative way of stating Theorem 6.1, is to say that the asymptotic bias in U_1 is equal to minus one half of the mean curvature of H, at μ_0. But M^* could equally well be represented as the sum of the curvatures in any s orthogonal directions; in particular it can be found from the directions $\mu_1, \mu_2 \cdots \mu_s$. To check this, consider the curvature evaluated at μ_0 of the curve lying in $\mathbb{R}^2$ with parametric form

$$\begin{pmatrix} E_\mu[U_1(\mu_0)] \\ \mu_i \end{pmatrix}.$$

The absolute value of the curvature at μ_0 can be calculated from the well known formula

$$|\kappa_i| = \frac{|E_\mu[U_1(\mu_0)]''\mu_i' - E_\mu[U_1(\mu_0)]'\mu_i''|}{(E_\mu[U_1(\mu_0)]'^2 + \mu_i'^2)^{\frac{3}{2}}}, \tag{6.2}$$

where $'$ denotes differentiation with respect to μ_i. Now, $\mu_i' = 1$, $\mu_i'' = 0$,$E_\mu[U_1(\mu_0)]' = 0$ as U_1 is orthogonal to $\widetilde{V}_i$, and finally $E_\mu[U_1(\mu_0)]'' = E[U_0\widetilde{V}_{ii}]$. Putting all this together establishes that (6.2) is equal to $|E[U_0\widetilde{V}_{ii}]| = |\widetilde{T}_{ii}|$. The sign of the curvature has been lost because (6.2) pays no attention to it, but the sign can be retrieved by noting that the sign of the curvature is the same as the sign of $b_{ii} = \widetilde{T}_{ii}$, so that the curvature is simply $\widetilde{T}_{ii}$. Summing the $\widetilde{T}_{ii}$ over i establishes that the mean curvature can be obtained by summing the curvature in the directions $\mu_1, \mu_2, \cdots, \mu_s$.

The next curvature result concerns the change in informations between the U_1 and U_2 projective scores. It is useful to look at the explicit form of the extended information matrix M_2 in the μ parameterization. For clarity the superscript $\tilde{}$ will be dropped, and the case where $s = 2$ presented. The M_2 matrix possesses the remarkably simple form:

$$\begin{pmatrix} E[U_0^2] & E[U_0V_1] & E[U_0V_2] & nT_{11} & nT_{12} & nT_{21} & nT_{22} \\ E[V_1U_0] & n & 0 & 0 & 0 & 0 & 0 \\ E[V_2U_0] & 0 & n & 0 & 0 & 0 & 0 \\ nT_{11} & 0 & 0 & 2n^2 + O(n) & O(n) & O(n) & O(n) \\ nT_{21} & 0 & 0 & O(n) & n^2 + O(n) & n^2 + O(n) & O(n) \\ nT_{12} & 0 & 0 & O(n) & n^2 + O(n) & n^2 + O(n) & O(n) \\ nT_{22} & 0 & 0 & O(n) & O(n) & O(n) & 2n^2 + O(n) \end{pmatrix}.$$

The matrix has been constructed using the fact that W is equal to the identity matrix, and (3.10) has been used to construct the lower right square matrix corresponding to $E[\mathbf{V}_2\mathbf{V}_2^T]$.

The difference between the U_1 information and the U_2 information is simply

$$\begin{pmatrix} nT_{11} \\ nT_{21} \\ nT_{12} \\ nT_{22} \end{pmatrix}^T \begin{pmatrix} 2n^2+O(n) & O(n) & O(n) & O(n) \\ O(n) & n^2+O(n) & n^2+O(n) & O(n) \\ O(n) & n^2+O(n) & n^2+O(n) & O(n) \\ O(n) & O(n) & O(n) & 2n^2+O(n) \end{pmatrix}^+ \begin{pmatrix} nT_{11} \\ nT_{21} \\ nT_{12} \\ nT_{22} \end{pmatrix}$$

which again is a contraction of a tensor to a scalar and hence invariant under transformations of λ. Using Theorem 3.7 for the Moore-Penrose inverse of

$$\begin{pmatrix} 2n^2+O(n) & O(n) & O(n) & O(n) \\ O(n) & n^2+O(n) & n^2+O(n) & O(n) \\ O(n) & n^2+O(n) & n^2+O(n) & O(n) \\ O(n) & O(n) & O(n) & 2n^2+O(n) \end{pmatrix},$$

establishes that asymptotically, that is, as $n \to \infty$, $I_{U_1} - I_{U_2} = \frac{1}{2}\sum_i \sum_j T_{ij}^2$. This naturally leads to:

Theorem 6.2 *The asymptotic difference in information between the U_1 and U_2 estimating equations,*

$$I_{U_1} - I_{U_2},$$

is equal to one half of the sum of the squares of the principle curvatures of H at μ_0.

Proof

$$I_{U_1} - I_{U_2} = \tfrac{1}{2}\textstyle\sum_i \sum_j \widetilde{T}_{ij}^2 = \tfrac{1}{2}\sum_i \sum_j \widetilde{T}_{ij}\widetilde{T}_{ji}$$

from the symmetry of $\widetilde{T}_{ij}$. But, $\sum_i \sum_j \widetilde{T}_{ij}\widetilde{T}_{ji} = \operatorname{tr}(\widetilde{T}\widetilde{T})$. But, $\operatorname{tr}(\widetilde{T}\widetilde{T})$ is equal to the sum of the squares of the eigenvalues of $\widetilde{T}$, which by Theorem 6.1, are simply the principle curvatures of H, which gives the result. ∎

6.2 Example

To illustrate the curvature results, the Circular Normal problem, Example 4, is ideal. Some insight is gained from slightly changing the notation: consider the bivariate normal random variable, (Y_1, Y_2), where the covariance matrix of (Y_1, Y_2) is an identity, so that Y_1 and Y_2 are independent. Denote the means of Y_1 and Y_2 by η_1 and η_2 respectively. Define the parameter of interest ρ and the nuisance parameter λ by the following transformations of η_1 and η_2:

$$\begin{aligned} \rho &= \rho(\eta_1, \eta_2) = \sqrt{\eta_1^2 + \eta_2^2} \\ \lambda &= \lambda(\eta_1, \eta_2) = \mathrm{ArcTan}[\eta_1, \eta_2] \end{aligned}$$

Of course, this is simply the transformation to polar coordinates.

The projective scores and their information are obtained according to the following calculations that were implemented in `Mathematica`, (the code is included as Appendix A).

The log-likelihood for the bivariate density is

$$-\log(2\pi) - \tfrac{1}{2}\left((y_1 - \rho\cos\lambda)^2 + (y_2 - \rho\sin\lambda)^2\right),$$

and the ρ and λ scores, U_0 and V_1 are given by

$$U_0 = y_1\, \cos\lambda + y_2\, \sin\lambda - \rho,$$

and

$$V_1 = -\rho y_1\, \sin\lambda + \rho y_2\, \cos\lambda,$$

respectively. Consider the joint distribution of U_0 and V_1. They are linear combinations of y_1 and y_2:

$$\begin{pmatrix} U_0 \\ V_1 \end{pmatrix} = \begin{pmatrix} \cos\lambda & \sin\lambda \\ -\rho\sin\lambda & \rho\cos\lambda \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} \rho \\ 0 \end{pmatrix},$$

and it follows from standard theory that

$$\begin{pmatrix} U_0 \\ V_1 \end{pmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \rho^2 \end{pmatrix}\right].$$

In order to calculate the extended information matrix, higher order Bhattacharyya basis functions must be obtained. Using the following relationships between the score functions, U_0 and V_1, and their derivatives with respect to λ facilitates the computations:

$$\begin{aligned} U_0' = &\quad -y_1 \sin\lambda + y_2 \cos\lambda = \tfrac{1}{\rho} V_1 \\ V_1' = &\quad -\rho y_1 \cos\lambda - \rho Y_2 \sin\lambda = -\rho\,(U_0 + \rho) \\ U_0'' = &\quad \tfrac{1}{\rho} V_1' = -\,(U_0 + \rho) \\ V_1'' = &\quad -\rho\,(U_0') = -V_1 \end{aligned}.$$

Since there is only one nuisance parameter in this model, the notation for the Bhattacharyya basis functions can be simplified by writing V_{11} as V_2, V_{111} as V_3 *e.t.c.* Using this notation, the recurrence relation given in Theorem 3.1 becomes

$$V_{t+1} = V_t' + V_t V_1,$$

and a repeated application of the theorem can be used to express V_{t+2} in terms of V_t and its derivatives:

$$V_{t+2} = V_{t+1}' + V_{t+1} V_1 = V_t'' + 2V_t' V_1 + V_t V_1' + V_t V_1^2.$$

From this representation certain orthogonality properties of the V_t can be established by induction. In particular,

Theorem 6.3 *If V_t is of the form*

$$V_t = \sum_{i=1}^{s} c_i V_1^{a_i} U_0^{b_i}$$

for some constants a_i,b_i, c_i, where the a_i are all odd or even then V_{t+2} can be expressed as

$$V_{t+2} = \sum_{i=1}^{t} f_i V_1^{d_i} U_0^{e_i},$$

for some constants d_i,e_i, f_i, where the d_i are all odd or even accordingly .

Proof

Trivially V_1 is of this form with a_i odd, and $V_2 = V_1^2 - \rho(U_0 + \rho)$ is of this form with a_i even. Assume the result is true up to V_r, then

$$V_{t+1} = V_t' + V_t V_1 = \sum c_k a_k V_1^{a_k-1} V_1' U_0^{b_k} + \sum c_k b_k V_1^{a_k} U_0^{b_k-1} U_0' + \sum c_k V_1^{a_k+1} U_0^{b_k}$$

. It follows that V_{t+2} equals

$$\begin{aligned}
&\sum c_k a_k(a_k-1) V_1^{a_k-2} V_1'^2 U_0^{b_k} + c_k a_k V_1^{a_k-1} V_1'' U_0^{b_k} + c_k a_k b_k V_1^{a_k-1} V_1' U_0^{b_k-1} U_0' \\
&\quad + \sum c_k b_k a_k V_1^{a_k-1} V_1' + c_k b_k (b_k - 1) V_1^{a_k} U_0^{b_k-2} U_0'^2 + c_k b_k V_1^{a_k} U_0^{b_k-1} U_0'' \\
&\quad + \sum c_k (a_k+1) V_1^{a_k} V_1' U_o^{b_k} + c_k V_1^{a_k+1} U_0^{b_k-1} U_0' \\
&\quad + \sum c_k a_k V_1^{a_k} V_1' U_0^{b_k} + c_k b_k V_1^{a_k+1} U_0^{b_k-1} U_0' + c_k V_1^{a_k+2} U_0^{b_k}.
\end{aligned}$$

Using the identities for the derivatives of U_0 and V_1 it is now a matter of simple verification to establish the theorem. ■

It follows that for V_t with t odd, has only odd powers of V_1 in it, and since V_1 follows a normal distribution with mean 0, the odd moments all have expectation 0. Furthermore U_0 and V_1 are independent so that $E[V_t U_0]$ for t odd, equals zero. In addition the product of V_t and V_{t+1} has only odd powers of V_1 in it so that these products are also mean zero. To summarize we have:

$$\begin{aligned}
E[U_0 V_t] &= 0 \text{ for } t \text{ odd.} \\
E[V_t V_{t+1+2i}] &= 0 \text{ for all } t \text{ and } i.
\end{aligned}$$

The practical interpretation of these results is that V_t, for t odd, can be left out of the M_t extended information matrix. The non-zero elements in the matrix are easily obtained by using the fact that U_0 and V_1 are independent and the formula for the even moments of a mean zero normal distribution:

$$E[Z^{2k}] = \frac{2k!}{k!2^k}.$$

Because U_0 and V_1 are independent and consequently uncorrelated the U_1 score is simply equal to U_0, so that the surface $H(\lambda)$ is given by the equation

$$E_\lambda \begin{pmatrix} y_1 \cos\lambda_0 + y_2 \sin\lambda_0 - \rho_0 \\ \frac{-\rho_0 y_1 \sin\lambda_0 + \rho_0 y_2 \cos\lambda_0}{\sqrt{E_{\lambda_0}[(-\rho_0 y_1 \sin\lambda_0 + \rho_0 y_2 \cos\lambda_0)^2]}} \end{pmatrix}$$

A straight forward calculation shows that this is equal to

$$\rho_0 \begin{pmatrix} cos(\lambda - \lambda_0) - 1 \\ \sin(\lambda - \lambda_0) \end{pmatrix}, \tag{6.3}$$

which is recognizable as the parametric form of a circle with radius ρ_0.

The value of the Theorems 6.1 and 6.2, is that they enable the bias and information correction terms to be interpreted in the context of curvatures of the surface $H(\lambda)$.

The extended information matrix for the circular normal problem is

$$M_2 = \begin{pmatrix} 1 & 0 & -\rho \\ 0 & \rho^2 & 0 \\ -\rho & 0 & \rho^2 + 2\rho^4 \end{pmatrix},$$

from which the bias $-\frac{1}{2}\mathrm{tr}(T^* W^{-1})$ and information correction $T^*(E[V_2^* V_2^*])^{-1} T^*$ can be directly obtained. These calculations are simplified by noting that V_1 and V_2 are uncorrelated so that $V_2^* = V_2$. Consequently

$$-\frac{1}{2}\mathrm{tr}(T^*W^{-1}) = -\tfrac{1}{2} \quad (-\rho)(\rho^2)^{-1} = \frac{1}{2\rho}. \tag{6.4}$$

$$T^*(E[V_2^*V_2^*])^{-1}T^* = (-\rho)(\rho^2+2\rho^4)^{-1}(-\rho) = \frac{\rho^2}{\rho^2+2\rho^4}. \tag{6.5}$$

Theorems 6.1 and 6.2 state that (6.4) and (6.5) also have interpretations as curvatures of the surface $H(\lambda)$ given in (6.3). $H(\lambda)$ is the equation for a circle of radius ρ_0, and therefore has a constant curvature of $\frac{1}{\rho_0}$. The sign of the curvature can be obtained by noting that the sign of the second fundamental form of $H(\lambda)$, b_{11}, is -1, so that the curvature is $-\frac{1}{\rho_0}$. Theorem 6.1 states that the asymptotic bias in U_1 is minus one half of the sum of the principle curvatures of $H(\lambda)$. But $H(\lambda)$ is a function of only one variable, λ, so that there is single principle curvature, namely $-\frac{1}{\rho_0}$, and the bias is given by $\frac{1}{2\rho_0}$ which agrees with (6.4).

Theorem 6.2, states that the *asymptotic* information correction is one half of the sum of the squares of the principle curvatures of $H(\lambda)$, so that the correction is equal to $\frac{1}{2\rho_0^2}$. This is easily reconciled with (6.5), by noting that the asymptotics of the problem can be approached by letting ρ go to infinity, so that the leading term of (6.5) is obtained as

$$\frac{1}{2\rho^2} = \frac{\rho^2}{\rho^2+2\rho^4} - \frac{1}{2\rho^2+4\rho^4}. \tag{6.6}$$

The fundamental usefulness of the curvature interpretations is to be able to say that bias in the U_1 score equation is attributable to a model curvature , and that no curvature means no bias. Similarly, the greater the curvature, the greater the bias, and the greater the information correction needs to be.

6.3 Comparison with Efron's curvature

In his 1985 paper, Efron considers the curvature of the "level surfaces", which for this circular normal problem will be denoted by $\mathcal{C}_\rho$, where $\mathcal{C}_\rho = \{\eta_1, \eta_2 : \sqrt{\eta_1^2 + \eta_2^2} = \rho\}$. These are the curves corresponding to sets of η_1 and η_2 that map onto the same value of the parameter of interest. Clearly the level curves are circles of radius ρ_0, so that they correspond to the curve $H(\lambda)$. But is this in general the case?

To answer this, consider a slightly more general problem, in which the kernel of the log-likelihood is

$$-\frac{1}{2}\left[(y_1 - \eta_1(\psi,\lambda))^2 + (y_2 - \eta_2(\psi,\lambda)^2)\right].$$

The circular normal problem is obtained by making the identifications $\eta_1 = \psi\cos\lambda$ and $\eta_2 = \psi\sin\lambda$, and by convention ψ is written as ρ to emphasize the fact that it is a radius. The score functions for ψ and λ are

$$\begin{aligned} U_0(\psi,\lambda) &= (y_1 - \eta_1)\eta_1^\psi + (y_2 - \eta_2)\eta_2^\psi, \\ V_1(\psi,\lambda) &= (y_1 - \eta_1)\eta_1^\lambda + (y_2 - \eta_2)\eta_2^\lambda, \end{aligned}$$

where the superscripts indicate partial derivatives. Denote the true values of η_1 and η_2 by η_{10} and η_{20} respectively. Then the curve $H(\lambda)$ is given by

$$E_\lambda = \begin{pmatrix} (y_1 - \eta_{10})\eta_{10}^\psi + (y_2 - \eta_{20})\eta_{20}^\psi - \frac{\eta_{10}^\psi\eta_{10}^\lambda + \eta_{20}^\psi\eta_{20}^\lambda}{{\eta_{10}^\lambda}^2 + {\eta_{20}^\lambda}^2}((y_1 - \eta_{10})\eta_{10}^\lambda + (y_2 - \eta_{20})\eta_{20}^\lambda) \\ \frac{(y_1 - \eta_{10})\eta_{10}^\lambda + (y_2 - \eta_{20})\eta_{20}^\lambda}{\sqrt{{\eta_{10}^\lambda}^2 + {\eta_{20}^\lambda}^2}} \end{pmatrix}. \tag{6.7}$$

The fact that the variance of (Y_1, Y_2) is an identity matrix has been used to obtain (6.7). Since ψ has been fixed, the expectation with respect to λ can equally well be thought of as the expectation with respect to η_1 and η_2, under the constraint that

η_1 and η_2 lie on one of the "level curves". It follows that after taking expectations (6.7) can be written as

$$H(\eta) = \begin{pmatrix} \eta_{10}^{\psi} - \eta_{10}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} & \eta_{20}^{\psi} - \eta_{20}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} \\ \frac{\eta_{10}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} & \frac{\eta_{20}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} \end{pmatrix} \begin{pmatrix} \eta_1 - \eta_{10} \\ \eta_2 - \eta_{20} \end{pmatrix}, \qquad (6.8)$$

which is recognizable as a transformation of η_1 and η_2. If this transformation is "rigid", that is a location or rotation/reflection (more precisely, a direct congruent transformation), then the curvature of the transformed curve, will be equal to that of the original curve, so that Efron's curvature will be equal to the curvature of $H(\lambda)$. In the case of the circular normal problem the matrix

$$\begin{pmatrix} \eta_{10}^{\psi} - \eta_{10}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} & \eta_{20}^{\psi} - \eta_{20}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} \\ \frac{\eta_{10}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} & \frac{\eta_{20}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} \end{pmatrix} = \begin{pmatrix} \cos\lambda_0 & \sin\lambda_0 \\ -\sin\lambda_0 & \cos\lambda_0 \end{pmatrix},$$

which is a rotation matrix, so that the transformation is indeed rigid and the curvatures are equal. In general the matrix will not be a rotation matrix, but by the use of a simple scale transformation it is possible to achieve one! Note first that

$$\det \begin{pmatrix} \eta_{10}^{\psi} - \eta_{10}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} & \eta_{20}^{\psi} - \eta_{20}^{\lambda}\frac{\eta_{10}^{\psi}\eta_{10}^{\lambda}+\eta_{20}^{\psi}\eta_{20}^{\lambda}}{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}} \\ \frac{\eta_{10}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} & \frac{\eta_{20}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} \end{pmatrix} = \frac{\eta_{10}^{\psi}\eta_{20}^{\lambda} - \eta_{10}^{\lambda}\eta_{20}^{\psi}}{\sqrt{\eta_{10}^{\lambda\,2} + \eta_{20}^{\lambda\,2}}},$$

which can be shown to be equal to

$$\sqrt{E_{\lambda_0}[U_1(\lambda_0)^2]},$$

and consider the curvature of the surface

$$G(\lambda) = \begin{pmatrix} E_{\lambda}\left[\frac{U_1(\lambda_0)}{\sqrt{E_{\lambda_0}[U_1(\lambda_0)^2]}}\right] \\ W(\lambda_0)^{-\frac{1}{2}} E_{\lambda}[\mathbf{V}_1(\lambda_0)] \end{pmatrix}. \qquad (6.9)$$

After taking the expectations, one can show that (6.9) is equal to

$$\begin{pmatrix} \frac{\eta_{20}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} & -\frac{\eta_{10}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} \\ \frac{\eta_{10}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} & \frac{\eta_{20}^{\lambda}}{\sqrt{\eta_{10}^{\lambda\,2}+\eta_{20}^{\lambda\,2}}} \end{pmatrix} \begin{pmatrix} \eta_1 - \eta_{10} \\ \eta_2 - \eta_{20} \end{pmatrix}$$

which is clearly a location and then rotation transformation of η_1, η_2, so that the curve $G(\lambda)$ has precisely the same curvature as that of Efron's "level curves". In a sense the two curvatures can be thought of as equivalent, at least up to a scale factor. In Efron's 1985 paper the normalized signed distance plays a fundamental role. Denoting this distance as W_ψ, then

$$W_\psi \equiv \frac{X_\psi - \mathrm{tr}(d_{\hat{\eta}(\psi)})}{1 - \mathrm{tr}(d^2_{\hat{\eta}(\psi)})}.$$

Here, $\mathrm{tr}(d_{\hat{\eta}(\psi)})$ is a bias correction, and $1 - \mathrm{tr}(d^2_{\hat{\eta}(\psi)})$ is the square root of the variance correction to the signed distance X_ψ, the signed distance being defined as distance between the data and the restricted *m.l.e.* for the parameter of interest. The elements of the matrix $d_{\hat{\eta}(\psi)}$ are one half of the "usual curvatures", so $\mathrm{tr}(d_{\hat{\eta}(\psi)})$ is precisely the same as the bias correction given in Theorem 6.1. The variance correction is

$$1 - \left(1 - \mathrm{tr}\left(d^2_{\hat{\eta}(\psi)}\right)\right)^2, \tag{6.10}$$

which does not look at first sight like the information/variance type correction suggested in Theorem 6.2. But, all the expressions are *asymptotic*, in the sense that they become more accurate as the curvatures go to zero, so that (6.10) can be re-expressed as

$$1 - (1 - \mathrm{tr}(d^2_{\hat{\eta}(\psi)}))^2 \approx 1 - (1 - 2\,\mathrm{tr}(d^2_{\hat{\eta}(\psi)})) = 2\,\mathrm{tr}(d^2_{\hat{\eta}(\psi)}) = \tfrac{1}{2}\sum \kappa_i^2,$$

in other words, one half of the sum of the squares of the curvatures. This provides some degree of reconciliation between Efron's variance correction and the one suggested in this thesis. Their difference will naturally vary from problem to problem, but to take an example from Efron's paper consider observing a normal random variable with mean on a 6-dimensional hyper-sphere of radius 5. Then the principle curvatures are all equal to $\frac{1}{5}$. The variance correction (6.10) is equal to

$$1-(1-\sum_{k=1}^{5}(\tfrac{1}{2\times 5})^2)^2 = 1-(1-0.05)^2 = 0.0975, \tag{6.11}$$

and the variance correction suggested by Theorem 6.2 is

$$\tfrac{1}{2}\textstyle\sum_{k=1}^{5}(\tfrac{1}{5})^2 = 0.1,$$

which is rather close to (6.11).

It is of some interest to compare the U_2 substitution score function, and Efron's normalized signed distance, which could be used as a score function. We do so in terms of expectation, the characteristic of interest being their degree of unbiasedness. For the circular normal problem

$$\hat{\lambda}_\rho = \mathrm{ArcTan}[y_1, y_2],$$

so that

$$U_0(\hat{\lambda}_\rho) = r-\rho,$$

and

$$V_1(\hat{\lambda}_\rho) = 0,$$

where r is the observed distance of (y_1, y_2) from the origin. The U_2 projective score can be calculated as (see the appendix for the program that was used to do this),

$$\frac{-\rho^2 + 2\,\rho^3\,U_0(\psi,\lambda) + V_1(\psi,\lambda)^2}{\rho + 2\,\rho^3}$$

so that

$$U_2(\hat{\lambda}_\psi) = r-\rho-\frac{r}{1+2\rho^2}. \tag{6.12}$$

The normalized signed distance is

$$W_\rho = \frac{r-\rho-\frac{1}{2\rho}}{1-\frac{1}{(2\rho)^2}}.$$

The expectations of these functions was taken over a range of ρ values between 0.05 and 3.0, where the expectation of r was found by a numerical integration. It is clear that the normalized signed distance runs into a major problem when $\rho = \frac{1}{2}$, where it is not defined. Figure 6.1 shows the expectation comparisons.

For reasonably large values of ρ there is very little difference between the two functions, but the U_2 substitution score does perform better for small ρ. The fact that the correction term $\frac{r}{1+2\rho^2}$, is a function of the data through r seems to aid in the improved behavior for ρ small. In this case the correction term is about equal to $-r$, which balances the first r in (6.12). It is the *empirical* nature of the correction term that distinguishes the projective score correction from many of the other current approaches. It might at first seem simpler to explicitly calculate the bias in the profile score and then just construct an adjusted score with this bias removed. However it is conceivable, that by allowing this correction term to be *stochastic* in nature, but of course, with limiting behavior equal to the exact bias, improvements can be obtained in problems with high curvature/small sample size. This is an interesting area of ongoing research.

The question of how to reconcile the fact that Efron's curvature, and the curvature of the level surfaces $H(\lambda)$ differ by a multiplicative constant $\sqrt{E_{\lambda_0}[U_1^2]}$ still remains. A heuristic explanation is that Efron is essentially working with the likelihood, and his signed distance *squared* is a likelihood ratio test statistic. This is well known to be invariant under 1 to 1 transformations of the parameter of interest, so that the signed distance itself will be invariant under 1 to 1 transformation of the parameter of interest. However, the projective scores are *not* invariant under one to one transformations of ψ, the Jacobian of the transformation appears as a multiplicative constant in the score function for the parameter of interest in the

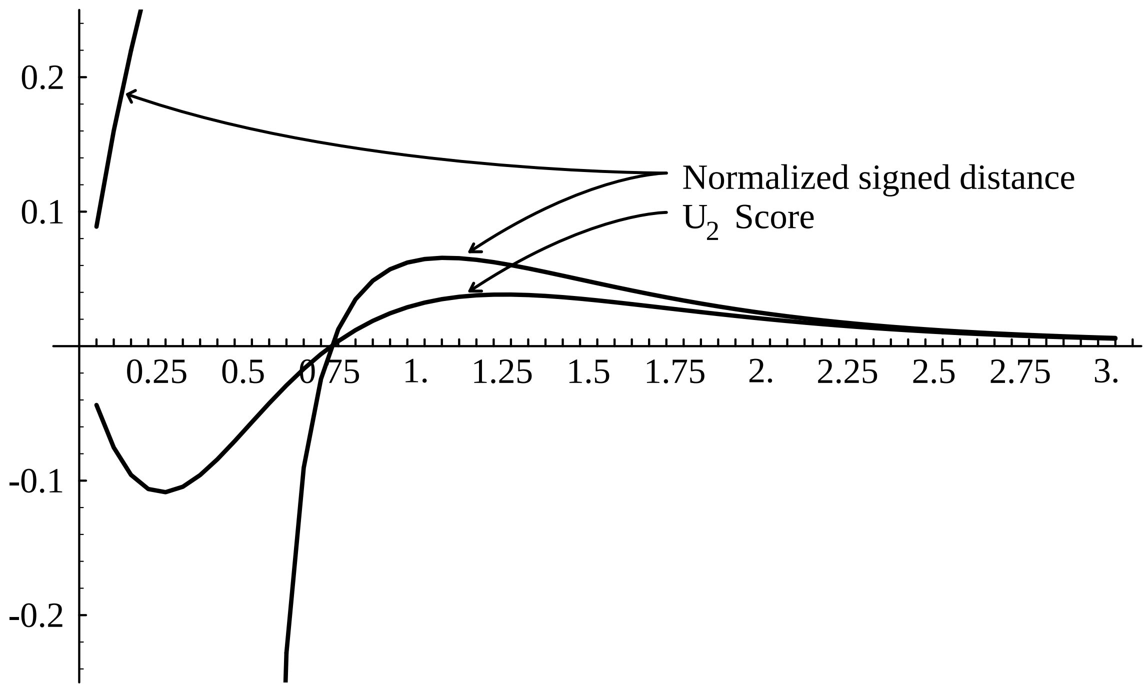

Figure 6.1: Bias comparison between the substitution U_2 score and the normalized signed distance.

transformed version. It is this non-invariance which accounts for the scale factor difference between the two curvatures.

6.4 A comparison of marginal and projective inference in the Circular Normal model

The circular normal problem is especially interesting because it allows a marginal inference approach. To be precise the distribution of $Y_1^2 + Y_2^2$ does not depend on the unknown angle λ, so that the marginal distribution of $R^2 = Y_1^2 + Y_2^2$ provides an alternative basis for inference.

The distribution of $R^2 = Y_1^2 + Y_2^2$ is a non-central χ^2 distribution on 2 degrees of freedom, with non-centrality parameter ρ^2:

$$f_{R^2}(r^2;\rho) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\rho^2}{2}}\left(\frac{\rho^2}{2}\right)^k}{k!}\,\frac{(r^2)^k e^{-\frac{r^2}{2}}}{k!\;2^{k+1}}.$$

The log-likelihood is given by

$$-\frac{\rho^2}{2} - \frac{r^2}{2} + ln\left\{\sum_{k=0}^{\infty} \frac{\left(\frac{\rho^2}{2}\right)^k (r^2)^k}{k!\;k!\;2^{k+1}}\right\}.$$

The ρ score, $U(r^2,\rho)$ is

$$-\rho + \frac{2}{\rho}\left\{\sum_{k=0}^{\infty} \frac{k(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}\Bigg/\left\{\sum_{k=0}^{\infty} \frac{(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}.$$

The information in ρ is $E[U(r^2,\rho)^2]$, and $U(r^2,\rho)^2$ is given by

$$\rho^2 - 4\left\{\sum_{k=0}^{\infty} \frac{k(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}\Bigg/\left\{\sum_{k=0}^{\infty} \frac{(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}$$
$$+\frac{4}{\rho^2}\left\{\sum_{k=0}^{\infty} \frac{k(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}^2\Bigg/\left\{\sum_{k=0}^{\infty} \frac{(\frac{\rho^2}{2})^k (r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}^2.$$

Using the fact that the non-central χ^2 distribution can be considered as an infinite sum of central χ^2, with a random number of degrees of freedom coming from a Poisson distribution with parameter $\frac{\rho^2}{2}$, leads to the following simplification for the information:

$$E[U(r^2,\rho)^2] = -\rho^2 + E\left[\frac{4}{\rho^2}\left\{\sum_{k=0}^{\infty}\frac{k(\frac{\rho^2}{2})^k(r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}^2 \Big/ \left\{\sum_{k=0}^{\infty}\frac{(\frac{\rho^2}{2})^k(r^2)^k e^{-\frac{1}{2}r^2}}{k!\;k!\;2^{k+1}}\right\}^2\right]$$

which can be approximated by a numerical integration.

The information in the projective scores for $U_0 = U_1$, and U_2, U_4, U_6 and U_8 was compared to the information in the marginal score, and a plot of the comparison follows (note that the higher order information curves are all close to one another, and are not individually labelled):

These informations were calculated under the true model parameters, that is, no estimates of λ were substituted into the projective scores, the true λ was used. The plot shows the successive decrease in the informations in the projective scores, with the information in the marginal score lying below that of the projective scores. The theory developed in Chapter 3 states that the projective informations should be decreasing to a limiting information, associated with that of the optimal kernel U_∞. The question of whether or not this limiting information lies above the marginal information is of interest. Lindsay (1992) has an argument that demonstrates that the space that is orthogonal to Λ_∞ is strictly larger than the space of mean zero estimating functions that depend on ρ alone. It is therefore quite feasible that the function U_∞ is not equal to the marginal score function. The important question of whether or not an increase in information, above that offered by the marginal score, is achievable when estimates are substituted for the true λ remains unanswered.

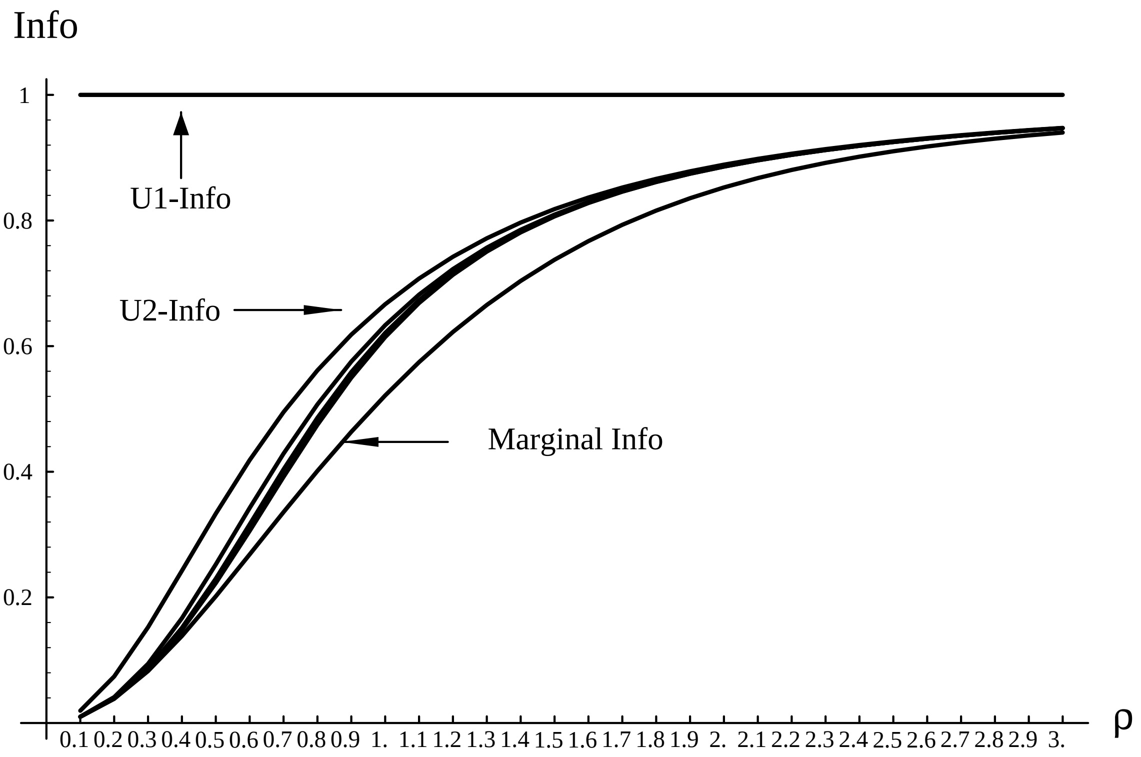

Figure 6.2: Information comparison between the projective and marginal scores.