ON THE COMPARISON OF STATISTICAL CURVATURE WITH GAUSSIAN CURVATURE

William W.S. Chen
Department of Statistics
The George Washington University
Washington D.C. 20013

ABSTRACT: Efron [1] introduced the new statistical curvature to measure the shape of a one-parameter exponential family. This family, through the sample space of all possible probability distributions, can be considered a “straight line”. In this paper, using the student t-distribution, we compare this new statistical curvature with the classical Gaussian curvature. Efron’s defined curvature has greatly reduced the quantities of curvature. In order to compare Efron’s statistical curvature with the Gaussian curvature this paper will look at the degree of freedom.

KEYWORDS: Comparison, Covariance Matrix, Curvature Reduction, Degree of Freedom, First Fundamental Form

INTRODUCTION

In the one parameter exponential families we usually have good properties for estimation, testing, and other inference results. In an exponential family with the correct choice of \( \theta \) function, we can show that the locally most powerful test is uniformly most powerful test. The maximum likelihood estimator for parameter \( \theta \) is a sufficient statistics and achieves the Cramer-Rao lower bound. However, if we consider an arbitrary one-parameter family then they may not enjoy the above exponential family properties. In this paper we define a quantity called the statistical curvature which is different from the classical Gaussian curvature. It is well known that if the family is an exponential family, then statistical curvature at \( \theta \) is zero and positive, for at least some \( \theta \) values, otherwise. In general, for nonexponential families, the maximum likelihood estimation is not a sufficient statistics.
We may interest to know how much information does it lose if we compare with all the data set. The answer can be expressed in terms of a statistical curvature. This theory goes back to Fisher, R.A.[2] and Rao, C.R.[3][4][5]. Rao has coined the term “second order efficiency” for this property of the maximum likelihood estimation that gives it a preferred place in the class of “first order efficient”. The detailed computation formula of this loss function can be found in Efron’s paper and we will not repeat it here. However, this paper is interested in finding the difference between Efron’s new statistical curvature and the classical Gaussian curvature. We find that Efron’s approach has greatly reduced the quantities of curvature thereby reducing the loss of information. In other words, with a smaller curvature then the estimator will be more efficient and can be more closing to the properties of “exponential families”. In section 2, we will define the Gaussian Curvature. In section 3, we will define the statistical curvature. In section 4, we give more detailed proof to demonstrate the results in previous sections are correct. The final comparison table will be available in the concluding remarks. More relevant information can be found in the Kass, R.E. and Vos, P.W.[6]. Chen, W.W.S. [7][8] [9] [10] contributed some relevant results about Gaussian curvature and Geodesic equations.

**Define Gaussian Curvature**

We start from normal curvature,

\[ \kappa = \frac{\mathcal{II}}{\mathcal{I}} = \frac{e_\lambda^2 + 2f e_\lambda d v + g v^2}{E_\lambda^2 + 2F e_\lambda d v + G v^2} = \frac{e + 2f \lambda + g \lambda^2}{E + 2F \lambda + G \lambda^2} = \kappa(\lambda) , \quad \text{where} \quad \lambda = \frac{dv}{du} \quad (2.1) \]

Where \( \mathcal{II} \): second fundamental form, and \( \mathcal{I} \): first fundamental form, adopted in reference [12], page 75 equation(5-8).

However, in statistics, we minor modified each of these quantity by including the expectation.(see equation (2.12), (2.13),(2.14)). The extreme values of \( \kappa \) can be characterized by \( \frac{d \kappa}{d \lambda} = 0 \)

\[ (E + 2F \lambda + G \lambda^2)(f + g \lambda) - (e + 2f \lambda + g \lambda^2)(F + G \lambda) = 0 \quad (2.2) \]

since \( E + 2F \lambda + G \lambda^2 = (E + F \lambda) + \lambda(F + G \lambda) \)

\( e + 2f \lambda + g \lambda^2 = (e + f \lambda) + \lambda(f + g \lambda) \).
We can cast equation (2.1) into the simpler form:
\[
\kappa = \frac{H}{I} = \frac{f + g\lambda}{F + G\lambda} = \frac{e + f\lambda}{E + F\lambda}
\]  
(2.3)

The equation (2.3) can be transformed into two equations that satisfy \( \kappa \) value.
\[
(F\kappa - e) + (F\kappa - f)\lambda = 0, \quad (F\kappa - f) + (G\kappa - g)\lambda = 0.
\]  
(2.4)

Equation (2.4) can be simultaneously satisfied if and only if
\[
\begin{vmatrix}
E\kappa - e & F\kappa - f \\
F\kappa - f & G\kappa - g
\end{vmatrix} = 0 
\]  
(2.5)

This quadratic equation (2.5) in \( \kappa \) has \( \kappa_1 \) and \( \kappa_2 \) as roots. From equation (2.5) we derive
\[
M = \frac{\kappa_1 + \kappa_2}{2} = \frac{EG - 2fF + eG}{2(EG - F^2)}, \quad \text{the mean curvature,}
\]
\[
K = \kappa_1 \kappa_2 = \frac{eg - f^2}{EG - F^2}, \quad \text{the Gaussian curvature}
\]
(sometimes called total curvature)  
(2.6)

In 1886, R. Baltzer used algebra to prove Gauss’ findings. Here are the results of Baltzer’s findings:
\[
K = \frac{eg - f^2}{EG - F^2}
\]

\[
= \frac{1}{(EG - F^2)^2} \left| \begin{array}{ccc}
\frac{1}{2}E_vv & F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u \\
\frac{1}{2}E_u & \frac{1}{2}E_v & 0 \\
\frac{1}{2}G_{uv} & F & G
\end{array} \right|
\]  
(2.7)

Where all the symbols E, F, G and their first, second derivatives adopted from geometry text such as references [12][13]. (for example, reference [12] page 112 equation (3-6)).

The above form of the Gaussian Curvature has useful points. Since it avoids computing the coefficient of second fundamental form, it is much easier to compute. In the reference [11][12][13] we can find more related information to (2.7). In this paper, we use it to compute the total curvature of the t-distribution.

Example : Let \( \Omega_3 \) be the location-scale manifold of density that has the student t distribution and has the form:
\[
\Omega_3 = \left\{ f(x) = \frac{1}{v} \frac{\Gamma \left( \frac{r+1}{2} \right)}{\sqrt{\pi r} \Gamma \left( \frac{r}{2} \right)} \left\{ 1 + \frac{1}{r} \left( \frac{x-u}{v} \right)^2 \right\}^{-\frac{r+1}{2}} \mid x \in \mathbb{R}, \ (u,v) \in \mathbb{R} \times \mathbb{R}_+ \right\}.
\]

where \( u \) is a location parameter, \( v \) is a scale parameter and \( r \) defined as the degree of freedom. Let us define the following variables to simplify the notation:

\[
a = \frac{1}{r}, \quad b = \frac{r+1}{2}, \quad c_r = \frac{\Gamma \left( \frac{r+1}{2} \right)}{\sqrt{\pi r} \Gamma \left( \frac{r}{2} \right)}.
\]

Then the logarithm of likelihood function of family \( t \) can be written as follows:

\[
\ln f(x) = \ln c_r - b \ln(1 + a \left( \frac{x-u}{v} \right)^2) - \ln v.
\] (2.8)

From equation (2.8), we can derive the first and second partial derivatives:

\[
\frac{\partial \ln f}{\partial u} = \frac{2ab(x-u)}{v^2 + a(x-u)^2},
\]

\[
\frac{\partial^2 \ln f}{\partial u^2} = \frac{2ab(x-u)^2 - v^2}{(a(x-u)^2 + v^2)^2},
\] (2.9)

\[
\frac{\partial \ln f}{\partial v} = \frac{2ab(x-u)^2 v^{-3}}{1 + a(x-u)^2 v^{-2}} - v^{-1},
\]

\[
\frac{\partial^2 \ln f}{\partial v^2} = v^{-2} \left[ 1 - \frac{6ab(x-u)^2 v^{-2}(1 + a(x-u)^2 v^{-2}) + 4a^2b(x-u)^4 v^{-4}}{(1 + a(x-u)^2 v^{-2})^2} \right],
\] (2.10)

\[
\frac{\partial^2 \ln f}{\partial v \partial u} = \frac{-4abv(x-u)}{(a(x-u)^2 + v^2)^2}.
\] (2.11)

We can now take the expected values of (2.9),(2.10)and (2.11)and derive the following equations (2.12),(2.13) and (2.14):

\[
E = -E\left( \frac{\partial^2 \ln f}{\partial u^2} \right) = -2ab\left( \frac{-r}{v^2(r+3)} \right) = \frac{r+1}{v^2(r+3)},
\] (2.12)

\[
F = -E\left( \frac{\partial^2 \ln f}{\partial v \partial u} \right) = 0,
\] (2.13)

\[
G = -E\left( \frac{\partial^2 \ln f}{\partial v^2} \right) = \frac{-1}{v^2(1 - 3 \frac{r+1}{r+3})} = \frac{1}{v^2(2r)}.
\] (2.14)
Based on above results of equations (2.12)(2.13) and (2.14), we can further derive their derivatives as follows:

\[
\begin{align*}
E_u &= 0; \quad E_v = \frac{-2(r+1)}{v^3(r+3)}; \quad G_u = 0; \quad G_v = \frac{-4r}{v^3(r+3)} \\
EG - F^2 &= \frac{2r(r+1)}{v^4(r+3)^2}; \quad F = 0; \quad F_u = 0; \quad F_v = 0 \\
G_u &= 0; \quad G_{uv} = 0; \quad F_{uv} = 0; \\
E_{vv} &= \frac{-2(r+1)}{r+3} \left( \frac{-3v^2}{v^6} \right) = \frac{6(r+1)}{v^4(r+3)} \\
-\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} &= \frac{-3(r+1)}{v^4(r+3)} + 0 + 0 = \frac{-3(r+1)}{v^3(r+3)}
\end{align*}
\]

Then substitute equations (2.15), (2.16) and (2.17) into equation (2.7):

\[
\begin{bmatrix}
\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\
F_v - \frac{1}{2}G_u & E & F \\
\frac{1}{2}G_v & F & G
\end{bmatrix}
= \begin{bmatrix}
-\frac{3(r+1)}{v^4(r+3)} & 0 & \frac{r+1}{v^3(r+3)} \\
0 & \frac{r+1}{v^3(r+3)} & 0 \\
-\frac{2r}{v^3(r+3)} & 0 & \frac{2r}{v^2(r+3)}
\end{bmatrix}
\]

\[
\begin{align*}
= -\frac{6r(r+1)^2}{v^8(r+3)^2} + \frac{2r(r+1)^2}{v^6(r+3)^3} = -\frac{4r(r+1)^2}{v^8(r+3)^3} \\
= \begin{bmatrix}
0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\
\frac{1}{2}E_v & E & F \\
\frac{1}{2}G_u & F & G
\end{bmatrix}
= \begin{bmatrix}
0 & -\frac{r+1}{v^3(r+3)} & 0 \\
-\frac{r+1}{v^3(r+3)} & \frac{r+1}{v^2(r+3)} & 0 \\
0 & 0 & \frac{2r}{v^2(r+3)}
\end{bmatrix}
\end{align*}
\]

\[
= -\left( \frac{r+1}{v^3(r+3)} \right)^2 \left[ \frac{2r}{v^2(r+3)} \right] = -\frac{2r(r+1)^2}{v^8(r+3)^3}
\]

Recall that \( EG - F^2 = \frac{2r(r+1)}{v^4(r+3)^2} \). Finally, we find that the Gaussian curvature is as follow:
More detailed proof on equation (2.12),(2.13) and (2.14) will be given in section 4.

**Define Statistical Curvature**

Let \( \mathcal{Z} = \{ f_\theta(x), \theta \in \Theta \} \) be an arbitrary family of density function indexed by the parameter \( \theta \in \Theta \), a possibly infinite interval of the real line. Let

\[
\begin{align*}
  l_\theta(x) &= \log f_\theta(x), \quad \hat{l}_\theta(x) = \frac{\partial}{\partial \theta} l_\theta(x), \quad \hat{\hat{l}}_\theta(x) = \frac{\partial^2}{\partial^2 \theta} l_\theta(x).
\end{align*}
\]

We assume the derivative exists continuously and can be uniformly dominated by integrable functions in a neighborhood of the given \( \theta \), so that

\[
E_\theta(l_{\hat{\theta}}) = 0, \quad E_\theta(l_{\hat{\hat{\theta}}}) = -E_\theta(\hat{l}_\theta) = i_\theta.
\]

Finally, we let \( M_\theta \) be the covariance matrix of \( (l_\theta, \hat{l}_\theta, \hat{\hat{l}}_\theta) \),

\[
M_\theta = \begin{pmatrix}
  v_{20}(\theta) & v_{11}(\theta) \\
v_{11}(\theta) & v_{02}(\theta)
\end{pmatrix} = \begin{pmatrix}
  E_\theta(l_{\hat{\theta}}^2) & E_\theta(\hat{l}_\theta, \hat{\hat{l}}_\theta) \\
E_\theta(\hat{l}_\theta, \hat{\hat{l}}_\theta) & E_\theta(l_{\hat{\theta}}^2)
\end{pmatrix}
\]

and define the statistical curvature of \( \mathcal{Z} \) at \( \theta \) to be

\[
\gamma_\theta = \frac{\mid M_\theta \mid^{1/2}}{i_\theta^3} = \frac{\mid v_{02}(\theta) \mid^{1/2}}{i_\theta^3} - \frac{\mid v_{11}(\theta) \mid^{1/2}}{i_\theta^3}.
\]

(3.1)

In making this definition we assume \( 0 < i_\theta < \infty \) and \( v_{02}(\theta) < \infty \).

Now we return to our translation family example,

t-distribution, and list the elements of the \( |M_\theta| \) matrix as follows:

\[
\begin{align*}
  v_{20}(\theta) &= i(\theta) = \frac{r+1}{r+3}, \quad v_{11}(\theta) = 0 (\text{due to symmetry}), \\
  v_{02}(\theta) &= \frac{r+1}{r+3} \left( \frac{r+2}{r(r+5)} \frac{r+10}{r+7} \right).
\end{align*}
\]

(3.2)

(3.3)

Finally, our computed statistical curvature of student t-distribution is given as:
\[
\gamma_0^2 = \frac{6(3r^2 + 18r + 19)}{r(r+1)(r+5)(r+7)}
\]

(3.4)

This is a monotone decreasing function of degree of freedom. A more detailed derivation of equation (3.2), (3.3) and (3.4) will be given in the next section.

**Derive Expectation Value**

In this section we give some more detailed computational procedure that demonstrates how we find the coefficient of the first fundamental form.

\[
E(\frac{\partial^2 \ln f}{\partial u^2}) = 2abE(\frac{a(x-u)^2-v^2}{(a(x-u)^2+v^2)^2}) = 2ab(\frac{-r}{v^2(r+3)}) = \frac{-(r+1)}{v^2(r+3)}
\]

let \( y = \frac{x-u}{v} \) and \( vdy = dx \)

\[
E(\frac{a(x-u)^2-v^2}{(a(x-u)^2+v^2)^2}) = C_\alpha \left( \frac{a(x-u)^2-v^2}{(a(x-u)^2+v^2)^2} \right) (1+a(\frac{x-u}{v})^2)^{-b} dx
\]

\[
= \frac{C_\alpha}{v^2} \left[ (ay^2-1)(1+ay^2)^{-(b+2)} dy \right]
\]

\[
= \frac{C_\alpha}{v^2} \left[ ay^2(1+ay^2)^{-(b+2)} dy - [(1+ay^2)^{-(b+2)} dy] \right]
\]

\[
= \frac{\alpha}{v^2} C_\alpha^{-1} (a\alpha^2\mu_{2,r+4}-1)
\]

\[
= \frac{1}{v^2} \left( \frac{r+4}{r} \right)^{-\frac{1}{2}} \left[ \frac{r+4}{r} \left( \frac{r+2}{r+1} \right) \left[ \frac{1}{r} \left( \frac{r+4}{r+2} \right) - 1 \right] \right]
\]

\[
= \frac{-r}{v^2(r+3)}
\]

\[
E = -E(\frac{\partial^2 \ln f}{\partial u^2}) = \frac{r+1}{v^2(r+3)}
\]

(4.1)

where \( \alpha = (a(2(b+2)-1))^{\frac{1}{2}} = (a(r+4))^{\frac{1}{2}} = (\frac{r+4}{r})^{\frac{1}{2}} \)

and \( a\alpha^2 = \frac{1}{r+4}; \mu_{2,r+4} = \frac{r+4}{r+2} \)
\[ C_r C_{r+4}^{-1} = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}} \frac{\sqrt{\pi} \left( r+4 \right) \Gamma\left(\frac{r+4}{2}\right)}{\Gamma\left(\frac{r+5}{2}\right)} = \sqrt{\frac{r+4}{r}} \frac{(r+2)r}{(r+3)(r+1)} \]

\[ E\left( \frac{\partial^2 \ln f}{\partial v \partial u} \right) = 4abE\left( \frac{(x-u)}{(a(x-u)^2 + v^2)^2} \right) = 0 \]

\[ F = -E\left( \frac{\partial^2 \ln f}{\partial v \partial u} \right) = -4ab*0 = 0 \] (4.2)

\[ G = -E\left( \frac{\partial^2 \ln f}{\partial v^2} \right) = -C_r \left[ \int v^{-2} (1 + ay^2)^{-b} dy + \right] \left[ ((1 + ay^2) - 6abyv^{-2} + 4a^2bv^4) \right] (1 + ay^2)^{(b+2)} dy \]

\[ = -C_r \int (1 + ay^2)^{-b} dy \cdot \int 2abv^2 (3 + ay^2)(1 + ay^2)^{(b+2)} dy \]

\[ = -C_r \int (C_r^{-1} - 2ab) y^2 (3 + ay^2)(1 + ay^2)^{(b+2)} dy \]

\[ = -C_r \frac{v}{v^2} \left( C_r^{-1} - 2ab \right) \left[ 3a^2 \mu_{2,r+4} C_{r+4}^{-1} + a\alpha^5 \mu_{4,r+4} C_{r+4}^{-1} \right] \]

\[ = -\frac{1}{v^2} \int (1 - 2abC_r C_{r+4}^{-1} \left[ 3a^2 \mu_{2,r+4} + a\alpha^5 \mu_{4,r+4} \right]) \]

\[ = -\frac{1}{v^2} \left( 1 - 2\left( \frac{1}{r} \right) \frac{r+1}{2} \right) \sqrt{\frac{r+4}{r}} \frac{(r+2)r}{(r+3)(r+1)} \left[ 3\left( \frac{r}{r+4} \right)^3 + \frac{1}{r} \left( \frac{r}{r+4} \right)^2 + \frac{5}{r+2} \left( \frac{r}{r+4} \right)^{\frac{5}{2}} \right] \]

\[ = -\frac{1}{v^2} \left( 1 - 2\left( \frac{1}{r} \right) \frac{r+1}{2} \right) \sqrt{\frac{r+4}{r}} \frac{(r+2)r}{(r+3)(r+1)} \left[ 3\left( \frac{r}{r+4} \right)^3 + \frac{3}{r} \left( \frac{r}{r+4} \right)^{\frac{5}{2}} \right] \]

\[ = -\frac{1}{v^2} \left( 1 - 3\left( \frac{r+1}{r+3} \right) \right) \sqrt{r+4}(r+2) \]

\[ G = -E\left( \frac{\partial^2 \ln f}{\partial v^2} \right) = -\frac{1}{v^2} \left( 1 - 3\left( \frac{r+1}{r+3} \right) \right) = \frac{2r}{v^2(r+3)} \] (4.3)

Equation (4.1), (4.2) and (4.3) proves the listed equations (2.12), (2.13) and (2.14) are correct.

Recall that
\[ i_\theta = -2abx \quad \text{if} \quad 1 + ax^2; \quad E(\dot{i}_\theta) = 0; \quad \dot{i}_\theta^2 = \frac{4a^2b^2x^2}{(1 + ax^2)^2}. \]

\[ \text{Var}(i_\theta) = E(\dot{i}_\theta^2) = (E(\dot{i}_\theta))^2 \]

\[ E(\dot{i}_\theta^2) = 4a^2b^2E(\frac{x^2}{(1 + ax^2)^2}) = 4a^2b^2C_r \int x^2(1 + ax^2)^{-(b+2)} \, dx \]

\[ = 4a^2b^2\alpha^3 \mu_{2+r+4}C_{r+5}^{-1} \]

\[ E(\ddot{i}_\theta^2) = 4\left(\frac{1}{r} \right)^2 \left( \frac{r+1}{2} \right)^2 \left( \frac{r}{r+4} \right)^2 \left( \frac{r+4}{r+2} \right)^2 \left( \frac{r+6}{r+5}(r+2)r \right) \]

\[ \left\{ \frac{1}{r} \right\}^2 \left( \frac{r+6}{r+8} \right)^2 \left( \frac{3(r+8)}{r+6} \right) \quad \text{or} \quad \left( \frac{r+1}{r} \right)^2 \frac{r^2 + 8r + 19}{r(r+7)(r+5)(r+3)(r+6)} \]

\[ \dot{i}_\theta^2 = \nu_2(\theta) = \frac{(r+1)^2}{r+3} \]

\[ \nu_2(\theta) = E(\ddot{i}_\theta^2) - \dot{i}_\theta^2 = \left( \frac{r+1}{r} \right) \left( \frac{r^2 + 8r + 19}{r(r+7)(r+5)(r+3)} \right) - \left( \frac{r+1}{r+3} \right)^2 \]

Then it is straightforward to find our statistical curvature as follow:

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)
\[ |M_\theta| = v_{20}^2 = \left( \frac{r+1}{r+3} \right)^2 \left( \frac{r+2}{r+3} \right)^2 \left( \frac{r+5}{r+3} \right)^2 \]
\[ i_\theta^2 = \left( \frac{r+1}{r+3} \right)^2 \]
\[ \frac{|M_\theta|}{i_\theta^2} = \gamma_\theta^2 = \frac{6(3r^2 + 18r + 19)}{r(r+1)(r+5)(r+7)} \]
\[ K^2 = \left( \frac{r+3}{4r^2} \right)^2 \]

Equations (4.4), (4.5), (4.6) and (4.7) demonstrate that the new statistical curvature is the desired result.

**CONCLUSION REMARKS**

In this section we summarize the Efron defined statistical curvature and classical Gaussian curvature together,

\[ \frac{|M_\theta|}{i_\theta^2} = \gamma_\theta^2 = \frac{6(3r^2 + 18r + 19)}{r(r+1)(r+5)(r+7)}; \quad K^2 = \left( \frac{r+3}{4r^2} \right)^2 \]

for the same selected degree of freedom, \( r \), to compare these two methods. We concluded that Efron’s defined curvature has greatly reduced the curvature thereby increasing the efficiency of the estimators.

<table>
<thead>
<tr>
<th>( r )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>( \rightarrow \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_\theta^2 )</td>
<td>2.5</td>
<td>1.063</td>
<td>0.306</td>
<td>0.107</td>
<td>0.0334</td>
<td>-18/( r^2 )</td>
</tr>
<tr>
<td>( K^2 )</td>
<td>4</td>
<td>1.5625</td>
<td>0.64</td>
<td>0.4225</td>
<td>0.3306</td>
<td>0.25</td>
</tr>
<tr>
<td>( K^2 / \gamma_\theta^2 )</td>
<td>1.6</td>
<td>1.47</td>
<td>2.09</td>
<td>3.95</td>
<td>9.9</td>
<td></td>
</tr>
</tbody>
</table>

**REFERENCES**


