Ann. Inst. Statist. Math.
Vol. 41, No. 3, 429-450 (1989)

FUNDAMENTAL EQUATIONS FOR STATISTICAL SUBMANIFOLDS WITH APPLICATIONS TO THE BARTLETT CORRECTION

PAUL W. VOS

Department of Mathematics, University of Oregon, Eugene, OR 97403, U.S.A.

(Received August 10, 1988; revised December 27, 1988)

Abstract. Many applications of Amari's dual geometries involve one or more submanifolds imbedded in a supermanifold. In the differential geometry literature, there is a set of equations that describe relationships between invariant quantities on the submanifold and supermanifold when the Riemannian connection is used. We extend these equations to statistical manifolds, manifolds on which a pair of dual connections is defined. The invariant quantities found in these equations include the mean curvature and the statistical curvature which are used in statistical calculations involving such topics as information loss and efficiency. As an application of one of these equations, the Bartlett correction is interpreted in terms of curvatures and other invariant quantities.

Key words and phrases: Dual geometries, statistical manifolds, submanifolds, exponential families, curvatures, imbedding curvature tensor, Bartlett correction, likelihood ratio statistic.

1. Introduction

In 1985, Amari introduced dual geometries and applied this theory to statistical inference in exponential families. The observed geometries of Barndorff-Nielsen (1986) are related to Amari's (1985) expected geometries and enjoy similar dual geometric structures. More recently, the dual geometries have been applied to generalized linear models and quasi-likelihood functions (Vos (1987)), statistical models outside of exponential families, Gaussian time series models (Amari (1987)), and linear systems (Amari (1986)). In many applications of these dual geometries, there is interest not only in a single manifold but also in one or more submanifolds and how these relate to the supermanifold. In particular, we are interested in the relationship between various tensors on the submanifolds and supermanifold. For the Riemannian or metric connection that is typically found in the differential geometry literature, these relationships are expressed by the Gauss formula, Weingarten formula, and the equations of Gauss, Codazzi and Ricci. These equations have such an important role for the Riemannian connection that they are called the fundamental equations for submanifolds (Spivak (1975)). We show here that the Weingarten formula and the equations of Ricci and of Gauss take on a slightly different form with the dual connections, while the Gauss formula and the equation of Codazzi are unchanged. Since these equations have a central role in the study of Riemannian connections, their importance may also be realized in the study of dual connections and their application to statistical manifolds.

One such application is given for the Bartlett correction, which has gained much attention in recent years (e.g., Barndorff-Nielsen and Cox (1984), Barndorff-Nielsen and Blæsild (1986), Møller (1986) and Barndorff-Nielsen and Hall (1988)). Lawley (1956) expresses the Bartlett correction using various moments of the score function and higher order derivatives of the log likelihood function. McCullagh and Cox (1986) express the Bartlett correction in terms of invariant quantities and interpret this correction in normal regression theory using the Riemannian curvature and the square of the mean curvature. In Section 5, we extend McCullagh and Cox's (1986) result for normal theory regression to exponential family regression. The curvatures and other geometric quantities used in this section will be defined using the dual geometries. We shall use the statistical curvature, the mixture curvature, and the second Riemannian scalar curvature, each of which appears in other statistical calculations. In Section 6, we show that Amari's (1987) approximating local exponential family allows us to extend the results of Section 5 to quite general statistical families.

2. Preliminary definitions

Since the dual geometries can be used in a diversity of applications, we give a formal definition of a Riemannian manifold with dual connections. It is easily verified that the expected geometries, observed geometries, and the geometric structure arising from the aforementioned applications satisfy our requirements of a manifold with dual connections. This formal development of dual geometries was first given by Lauritzen (1987).

We briefly introduce some notation and terminology. A more complete description of these terms can be found in Amari (1985) and Lauritzen (1987). We consider an n-dimensional Riemannian manifold S and an m-dimensional regular submanifold M. Since M is regular, there exists a coordinate system (U, ϕ) on S for each $p \in M$ such that

$$\phi(p) = (0,\ldots,0); \qquad \phi(U) = (-\varepsilon,\varepsilon)^n\,, \tag{2.1}$$
$$\phi(U \cap M) = \{(x^1,\ldots,x^m,0,\ldots,0)\colon |x^i| < \varepsilon\}\,.$$

Any coordinate system that satisfies (2.1) is called a *preferred coordinate neighborhood.* The *tangent space* of S at p, denoted by S_p, is an n-dimensional vector space with *canonical basis* for θ

$$\left\{ \partial_{ip}\colon \partial_{ip} = \frac{\partial}{\partial\theta^i}\bigg|_{\theta=\phi(p)} \right\}.$$

A *vector field* X is a family of smooth tangent vectors $\{X_p\colon p \in S\}$ where $X_p \in S_p$. For X to be smooth, we require that $X\colon C^\infty(S) \mapsto C^\infty(S)$.

The Riemannian metric on S is denoted by $\langle\!\langle X, Y\rangle\!\rangle$ for vector fields X and Y; the metric evaluated at p is written $\langle\!\langle X_p, Y_p\rangle\!\rangle$. Vectors in nearby tangent spaces are related by an *affine connection* $\tilde\nabla$, which is any function from $\underline{X}(S) \times \underline{X}(S)$ to $\underline{X}(S)$ that satisfies

$$\tilde\nabla_X(Y+Z) = \tilde\nabla_X Y + \tilde\nabla_X Z\,,$$
$$\tilde\nabla_X \alpha Y = (X\alpha) + \alpha\tilde\nabla_X Y\,,$$
$$\tilde\nabla_{\alpha X+\beta Y} Z = \alpha\tilde\nabla_X Z + \beta\tilde\nabla_Y Z\,.$$

The vector field $\tilde\nabla_X Y$ can be interpreted as the instantaneous change of the vector field Y in the direction of X. Amari (1985) and Lauritzen (1987) describe how $\tilde\nabla$ relates vectors in different tangent spaces.

Since M is a regular submanifold of S, the tangent space of M at p, M_p, is a linear subspace of S_p and the metric and connection on S can be used to define the corresponding quantities on M. We shall be concerned more with vector fields on M than on S, and so we change notation by letting $X, Y,\ldots$ represent vector fields on M and $\tilde X, \tilde Y,\ldots$ be extensions of $X, Y,\ldots$ to vector fields on S. The vector field $\tilde X$ is an extension of X if $\tilde X|_M = X$. A metric $\langle\cdot,\cdot\rangle$ on M can be defined by

$$\langle X_p, Y_p\rangle = \langle\!\langle X_p, Y_p\rangle\!\rangle\,, \tag{2.2}$$

because $M_p \subset S_p$. The metric defined by (2.2) is called the *induced metric* on M.

The *induced connection* ∇ cannot be defined quite so simply. Since ∇ is not a tensor, it depends on the vector fields X, Y and we must define $\nabla_X Y(p)$ rather than $\nabla_{X_p} Y_p$. In order to use $\tilde\nabla_{\tilde X}\tilde Y$ to define $\nabla_X Y$, we must check that $\tilde\nabla_{\tilde X}\tilde Y$ is the same on M for all extensions $\tilde X, \tilde Y$. Appendix 1 shows that this is indeed the case, so that we can write $\tilde\nabla_X Y$ as an abbreviation for $(\tilde\nabla_{\tilde X}\tilde Y)|_M$ without ambiguity. Since $(\tilde\nabla_X Y)(p) \in S_p$ but we need $(\nabla_X Y)(p) \in M_p$, we define

$$\nabla_X Y(p) = P_{M_p}(\tilde{\nabla}_X Y(p)) \,, \tag{2.3}$$

where P_{M_p} is the orthogonal projection onto M_p. It is easily shown that $\nabla_X Y$ defined by (2.3) is a connection on M. Although there are many connections that can be defined on a Riemannian manifold, there is only one connection that satisfies the following two conditions

(i) $\nabla_X Y - \nabla_Y X = XY - YX$,

(ii) $Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$.

Any connection satisfying (i) is called *torsion-free*, and the unique connection satisfying both (i) and (ii) is the Riemannian or *metric connection*. To distinguish the metric connection from arbitrary connections we use the symbol ∇^0. It can be shown that if $\tilde{\nabla}^0$ is the metric connection on S, then the induced connection ∇^0 is the metric connection on the submanifold M. An extension of this result is given in Proposition 3.1.

So far we have described a Riemannian manifold and submanifold with a single connection. A *statistical manifold* (Lauritzen (1987)) is a Riemannian manifold in which there exists a pair of torsion-free affine connections ∇ and ∇^* that satisfy

$$Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla^*_Z Y\rangle \,.$$

The connections ∇ and ∇^* are called *dual connections*, and it is easily shown that $(\nabla^*)^* = \nabla$. One can also show that any torsion-free affine connection ∇ has a dual connection given by $\nabla^* = 2\nabla^0 - \nabla$ (see, for example, Lauritzen (1987)). To emphasize the relationship between these two connections, we use the terminology *dual connection* rather than *torsion-free affine connection*. For a pair of dual connections, it is possible to generate an entire family of dual connections (e.g., α-connections of Amari (1985)). For our purposes, however, we shall just be interested in a single pair of dual connections ∇ and ∇^* and the self-dual connection ∇^0.

3. The Gauss and Weingarten formulas

It will be useful to write the dual connections on the supermanifold S as

$$\tilde{\nabla}_X Y = \top(\tilde{\nabla}_X Y) + \perp(\tilde{\nabla}_X Y) \,, \tag{3.1a}$$

$$\tilde{\nabla}^*_X Y = \top(\tilde{\nabla}^*_X Y) + \perp(\tilde{\nabla}^*_X Y) \,, \tag{3.1b}$$

where $\top(\cdot)$ and $\perp(\cdot)$ are smooth functions such that $\top(\cdot)|_{S_p} = P_{M_p}$ and $\perp(\cdot)|_{S_p} = (I - P_{M_p})$. For $v \in \underline{X}(S)$, $\top(v)$ and $\perp(v)$ are called the tangential component of v and the normal component of v, respectively. The following proposition shows that the induced connections

$$\nabla_X Y = \top(\tilde{\nabla}_X Y)\,, \tag{3.2a}$$

$$\nabla^*_X Y = \top(\tilde{\nabla}^*_X Y) \tag{3.2b}$$

are dual.

PROPOSITION 3.1. *On M, the connections ∇ and ∇^* defined by* (3.2) *are dual with respect to $\langle\cdot,\cdot\rangle$. Furthermore, the normal vector fields $h(X, Y) = \perp(\tilde{\nabla}_X Y)$ and $h^*(X, Y) = \perp(\tilde{\nabla}^*_X Y)$ are symmetric and bilinear.*

PROOF. Making two substitutions into (3.1a), we have

$$\tilde{\nabla}_X Y = \nabla_X Y + h(X, Y)\,. \tag{3.3}$$

Replacing X and Y by αX and βY, respectively, where $\alpha, \beta \in C^\infty(M)$, we obtain

$$\begin{aligned}\tilde{\nabla}_{\alpha X}(\beta Y) &= \alpha\{(X\beta)Y + \beta\tilde{\nabla}_X Y\} \\ &= \{\alpha(X\beta)Y + \alpha\beta\nabla_X Y\} + \alpha\beta h(X, Y)\,.\end{aligned} \tag{3.4}$$

Considering the tangent and normal component of (3.4), we have

$$\nabla_{\alpha X}(\beta Y) = \alpha(X\beta)Y + \alpha\beta\nabla_X Y\,, \tag{3.5}$$

$$h(\alpha X, \beta Y) = \alpha\beta h(X, Y)\,. \tag{3.6}$$

Since the affine connection $\tilde{\nabla}_X Y$ is additive in X and Y, i.e., $\tilde{\nabla}_{X+Z} Y = \tilde{\nabla}_X Y + \tilde{\nabla}_Z Y$ and $\tilde{\nabla}_X(Y + W) = \tilde{\nabla}_X Y + \tilde{\nabla}_X W$, $\nabla_X Y$ and $h(X, Y)$ are each additive in X and Y. Equation (3.5) together with the additivity of $\nabla_X Y$ make ∇ an affine connection on M. The additivity of $h(X, Y)$ and (3.6) show that $h(X, Y)$ is bilinear.

Since $\tilde{\nabla}$ has no torsion, we can write

$$\begin{aligned}0 &= \tilde{\nabla}_X Y - \tilde{\nabla}_Y X - [X, Y] \\ &= \nabla_X Y + h(X, Y) - \nabla_Y X - h(Y, X) - [X, Y]\,,\end{aligned} \tag{3.7}$$

where $[X, Y] = XY - YX$. Grouping the normal and tangential components of (3.7), we have

$$h(X, Y) = h(Y, X)$$

and

$$0 = \nabla_X Y - \nabla_Y X - [X, Y],$$

because $[X, Y](p) \in M_p$. These equations show that ∇ has no torsion and that $h(X, Y)$ is symmetric. The mappings

$$p \mapsto (\nabla_X Y)_p \quad \text{and} \quad p \mapsto h(X, Y)_p \tag{3.8}$$

are C^∞ because $\tilde{\nabla}_X Y$, $\top(\cdot)$ and $\perp(\cdot)$ are C^∞. In (3.8), we write $(\nabla^*_X Y)_p$ for $\nabla^*_X Y(p)$ and $h(X, Y)_p$ for $h(X, Y)(p)$. The proof for $\nabla^*_X Y$ and $h(X, Y)$ follows immediately by substituting $\tilde{\nabla}^*$ for $\tilde{\nabla}$ in the preceding argument. The connections ∇ and ∇^* are torsion-free, so we need only show they are dual. Duality of ∇ and ∇^* is proved from the duality of $\tilde{\nabla}$ and $\tilde{\nabla}^*$ as the following calculation shows

$$\begin{aligned} X\langle Y, Z\rangle &= \tilde{X}\langle\!\langle \tilde{Y}, \tilde{Z}\rangle\!\rangle|_M \\ &= \{\langle\!\langle \tilde{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z}\rangle\!\rangle + \langle\!\langle \tilde{Y}, \tilde{\nabla}^*_{\tilde{X}}\tilde{Z}\rangle\!\rangle\}|_M \\ &= \langle\!\langle \tilde{\nabla}_X Y, Z\rangle\!\rangle + \langle\!\langle Y, \tilde{\nabla}^*_X Z\rangle\!\rangle \\ &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla^*_X Z\rangle, \end{aligned}$$

where X, Y, Z are any vector fields in $T(M)$. □

When $\tilde{\nabla}$ is the Riemannian connection, equation (3.3) is called the *Gauss formula*. Proposition 3.1 allows us to extend the Gauss formula to any dual connection. Notice that equation (3.3) contains the dual equation

$$\tilde{\nabla}^*_X Y = \nabla^*_X Y + h^*(X, Y)$$

by taking $\tilde{\nabla}^*$ to be the primal and $\tilde{\nabla}$ its dual.

In the differential geometry literature (Chen (1984)) when $\tilde{\nabla} = \tilde{\nabla}^*$, $h(\cdot, \cdot)$ is called the second fundamental form of M (the first fundamental form is the metric $\langle\cdot, \cdot\rangle$). Following Amari (1985), however, we call $h(\cdot, \cdot)$ the *imbedding curvature tensor* of M in S for $\tilde{\nabla}$ and $h^*(\cdot, \cdot)$ is the imbedding curvature tensor of M in S for $\tilde{\nabla}^*$. In statistical applications, the dual connections used most often are the exponential connections ($\tilde{\nabla}$) and its dual, the mixture connection ($\tilde{\nabla}^*$), so that $h(\cdot, \cdot)$ is the exponential imbedding curvature and $h^*(\cdot, \cdot)$ is the mixture imbedding curvature. Since $h(X, Y)$ is linear in X and $Y, h(X, Y) = h(X_p, Y_p)$ depends only on $X_p, Y_p \in M_p$ and not on their extensions. This implies that for each ξ in

$\underline{X}^{\perp}(M) = \{\tilde{X} \in \underline{X}(S)|_M: \tilde{X}_p \perp M_p\}$ there exists linear transformations $(A_\xi)_p$: $M_p \mapsto M_p$ at each p, such that

$$\langle A_\xi X, Y\rangle = \langle\!\langle h(X, Y), \xi\rangle\!\rangle , \tag{3.9a}$$

for all $X, Y \in T(M)$. Of course $h^*(X, Y)$ has an A_ξ^* with the properties A_ξ has for $h(X, Y)$,

$$\langle A_\xi^* X, Y\rangle = \langle\!\langle h^*(X, Y), \xi\rangle\!\rangle . \tag{3.9b}$$

We can write $\tilde{\nabla}_X \xi$ in its tangential and normal components just as we did for $\tilde{\nabla}_X Y$,

$$\tilde{\nabla}_X \xi = \top (\tilde{\nabla}_X \xi) + D_X \xi , \tag{3.10a}$$

$$\tilde{\nabla}_X^* \xi = \top (\tilde{\nabla}_X^* \xi) + D_X^* \xi . \tag{3.10b}$$

We do not introduce a new symbol for the tangential components, $\top (\tilde{\nabla}_X \xi)$ and $\top (\tilde{\nabla}_X^* \xi)$, because they are related to $A_\xi^* X$ and $A_\xi X$ as Proposition 3.2 states. Proposition 3.3 justifies the symbolism chosen for the normal components, $D_X \xi$ and $D_X^* \xi$.

PROPOSITION 3.2. *$A_\xi X$ and $A_\xi^* X$ are each bilinear in X and ξ. Furthermore, for all $X_p \in M_p$ and $\xi \in \underline{X}^{\perp}(M)$*

$$\top (\tilde{\nabla}_{X_p} \xi) = - A_{\xi_p}^* X_p , \tag{3.11a}$$

$$\top (\tilde{\nabla}_{X_p}^* \xi) = - A_{\xi_p} X_p . \tag{3.11b}$$

PROOF. Bilinearity of $A_\xi X$ follows from bilinearity of $h(\cdot, \cdot)$ and $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ in (3.9a). This implies $A_\xi X$ is only a function of ξ_p and X_p. Similar remarks hold true for $A_\xi^* X$. Equation (3.11b) is proven as follows

$$\begin{aligned} 0 &= \langle\!\langle \tilde{\nabla}_X Y, \xi\rangle\!\rangle + \langle\!\langle Y, \tilde{\nabla}_X^* \xi\rangle\!\rangle \\ &= \langle\!\langle h(X, Y), \xi\rangle\!\rangle + \langle Y, \top (\tilde{\nabla}_X^* \xi)\rangle \\ &= \langle A_\xi X, Y\rangle + \langle Y, \top (\tilde{\nabla}_X^* \xi)\rangle . \end{aligned}$$

This proves $\top (\tilde{\nabla}_{X_p}^* \xi) = - A_{\xi_p} X_p$. Equation (3.11a) is proven similarly. Smoothness (C^∞) of $\top$ and $\tilde{\nabla}$ imply $- A_\xi^* X$ is C^∞. Likewise, $- A_\xi X$ is C^∞. □

Before stating Proposition 3.3, we give an informal definition of the normal bundle of M. At each $p \in M$, there exists an $(n - m)$-dimensional linear subspace $M_p^{\perp} \subset S_p$. If we collect these vector spaces in a smooth manner, the resulting object becomes a manifold called the normal bundle on M and is denoted by $T^{\perp}(M)$. The manifold formed by using the tangent spaces M_p is $T(M)$, the tangent bundle of M.

PROPOSITION 3.3. *D and D^* are Riemannian dual connections with respect to the induced metric on $T^{\perp}(M)$.*

PROOF. That $D(D^*)$ is a torsion-free connection follows from the properties of the connection $\tilde{\nabla}(\tilde{\nabla}^*)$. Also, $D(D^*)$ is C^∞ because $\tilde{\nabla}(\tilde{\nabla}^*)$ is C^∞. The induced metric on $T^{\perp}(M)$ is just $\langle\langle \xi, \eta \rangle\rangle$ for all $\xi, \eta \in T^{\perp}(M)$. Hence, the duality of D and D^* follows from

$$\begin{aligned} X\langle\langle \xi, \eta \rangle\rangle &= \langle\langle \tilde{\nabla}_X \xi, \eta \rangle\rangle + \langle\langle \xi, \tilde{\nabla}^*_X \eta \rangle\rangle \\ &= \langle\langle D_X \xi, \eta \rangle\rangle + \langle\langle \xi, D^*_X \eta \rangle\rangle . \qquad \square \end{aligned}$$

For the Riemannian connection, the eigenvalues and eigenvectors of A_ξ are called the *principal curvatures* and *principal directions* for ξ, respectively. The trace of A_ξ is the component of the *mean curvature vector* in the direction of ξ. Proposition 3.2 shows that we may extend these definitions to any dual connection.

Substituting (3.11a) and (3.11b) into (3.10a) and (3.10b), respectively, we obtain

$$\tilde{\nabla}_X \xi = -A^*_\xi X + D_X \xi , \tag{3.12a}$$

$$\tilde{\nabla}^*_X \xi = -A_\xi X + D^*_X \xi . \tag{3.12b}$$

We shall call (3.12a) and (3.12b) the *Weingarten formulas* for dual connections since they reduce to the Weingarten formula (3.13) when $\tilde{\nabla} = \tilde{\nabla}^*$

$$\tilde{\nabla}_X \xi = -A_\xi X + D_X \xi . \tag{3.13}$$

4. The fundamental equations

The fundamental equations for submanifolds with dual connections are statements about the ***Riemannian curvature tensor*** $R(X, Y)$ on M and the Riemannian curvature tensor $\tilde{R}(X, Y)$ on S restricted to M. The tensor $R(X, Y)$ is defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z .$$

The definition for $\tilde{R}(X, Y)$ follows by replacing ∇ with $\tilde{\nabla}$ in the above equation. It should be noted that for fixed $X_p, Y_p \in M_p$, $R(X_p, Y_p)$: $M_p \mapsto M_p$ and $\tilde{R}(X_p, Y_p)$: $S_p \mapsto S_p$ are linear transformations. For a geometric interpretation of this transformation, see Amari (1985). Proposition 4.1 gives the fundamental equations for submanifolds.

PROPOSITION 4.1. *Let $\tilde{\nabla}$ be a dual connection on S and ∇ the induced connection on M. Let $\tilde{R}(\cdot,\cdot)$ and $R(\cdot,\cdot)$ be the Riemannian curvature tensors for $\tilde{\nabla}$ and ∇, respectively. Then,*

$$\langle\!\langle \tilde{R}(X,Y)Z, W\rangle\!\rangle = \langle R(X,Y)Z, W\rangle + \langle\!\langle h(X,Z), h^*(Y,W)\rangle\!\rangle - \langle\!\langle h^*(X,W), h(Y,Z)\rangle\!\rangle\,, \tag{4.1}$$

$$(\tilde{R}(X,Y)Z)^{\perp} = D_X h(Y,Z) - h(\nabla_X Y, Z) - h(Y, \nabla_X Z) - \{D_Y h(X,Z) - h(\nabla_Y X, Z) - h(X, \nabla_Y Z)\}\,, \tag{4.2}$$

$$\langle\!\langle R_D(X,Y)\xi, \eta\rangle\!\rangle = \langle\!\langle \tilde{R}(X,Y)\xi, \eta\rangle\!\rangle + \langle [A^*_\xi, A_\eta](X), Y\rangle\,, \tag{4.3}$$

where $R_D(\cdot,\cdot)$ is the Riemannian curvature tensor on $T^{\perp}(M), \xi, \eta \in \underline{X}^{\perp}(M)$ and $[A^*_\xi, A_\eta] = A^*_\xi A_\eta - A_\eta A^*_\xi$.

The proof of Proposition 4.1 is a laborious calculation which we postpone to Appendix 2. Notice that equations (4.1)–(4.3) are each a statement about $\tilde{R}(X,Y)$. The *Gauss equation* (4.1) describes the tangential component of the image of a vector $Z_p \in M_p$ under this transformation. The normal component of $\tilde{R}(X,Y)Z$ is given in the *equation of Codazzi* (4.2). Finally, the normal component of the image of a vector field $\xi \in T^{\perp}(M)$ is described by the *equation of Ricci* (4.3). In the next section we show that the Gauss equation is useful in statistical calculations because it relates the imbedding curvature tensor and its dual. The role of (4.2) and (4.3) is less clear; we present them here for completeness.

In many applications $\langle\!\langle \tilde{R}(\tilde{X},\tilde{Y})\tilde{Z}, \tilde{W}\rangle\!\rangle = 0$ for all $\tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{W} \in \underline{X}(S)$ in which case S is said to be flat in the connection $\tilde{\nabla}$. When S is flat, equations (4.1)–(4.3) simplify to

$$\langle R(X,Y)Z, W\rangle = \langle\!\langle h^*(X,W), h(Y,Z)\rangle\!\rangle - \langle\!\langle h(X,Z), h^*(Y,W)\rangle\!\rangle\,, \tag{4.4}$$

$$D_X h(Y,Z) - h(\nabla_X Y, Z) - h(Y, \nabla_X Z) = D_Y h(X,Z) - h(\nabla_Y X, Z) - h(X, \nabla_Y Z)\,, \tag{4.5}$$

$$\langle\!\langle R_D(X,Y)\xi, \eta\rangle\!\rangle = \langle [A^*_\xi, A_\eta](X), Y\rangle\,, \tag{4.6}$$

respectively.

It is often useful to have these equations in coordinate form. We use $a, b, c, d = 1, \ldots, m$ to index vector fields whose vectors at each $p \in M$ form a basis for M_p, $\kappa, \lambda = m + 1, \ldots, n$ to index vector fields whose vectors at p form a basis for $M_p^{\perp}$ (the orthogonal complement of M_p in S_p), and

$\alpha, \beta, \gamma, \delta = 1, \ldots, m, m+1, \ldots, n$ to index vector fields that form a basis for $S_p = M_p + M_p^{\perp}$. We denote these vector fields by $Z_a, \ldots, Z_d \in \underline{X}(M)$, $Z_\kappa, Z_\lambda \in \underline{X}^{\perp}(M)$ and $Z_\alpha, \ldots, Z_\delta \in \underline{X}(S)$ where $Z_a = Z_\alpha$ when $a = \alpha$ and $Z_\kappa = Z_\alpha$ when $\kappa = \alpha$. If we use the preferred coordinate neighborhood (U, ϕ) at p and let $\{\partial_\alpha\}$ be the canonical basis for ϕ, then $Z_a = \partial_a$ while $Z_\kappa = \perp(\partial_\kappa)$. The components of the metric on S are given by

$$g_{\alpha\beta} = \langle\!\langle Z_\alpha, Z_\beta \rangle\!\rangle .$$

The components of the connection $\tilde{\nabla}$ are

$$\tilde{\Gamma}_{ab\gamma} = \langle\!\langle \tilde{\nabla}_{Z_a} Z_b, Z_\gamma \rangle\!\rangle ,$$

and the components of the imbedding curvature $h(Z_a, Z_b)$ are

$$H_{ab\kappa} = \langle\!\langle h(Z_a, Z_b), Z_\kappa \rangle\!\rangle .$$

We can raise or lower indices of these components by multiplying by $g^{\alpha\beta} = (g_{\alpha\beta})^{-1}$, the inverse of the inner product matrix. In particular,

$$\tilde{\Gamma}^{\gamma}_{ab} = \sum_{\gamma'=1}^{n} \tilde{\Gamma}_{ab\gamma'} g^{\gamma'\gamma} = \tilde{\Gamma}_{ab\gamma'} g^{\gamma'\gamma} .$$

In the second equality of the preceding display, we have used the Einstein summation convention, where we sum over the range of an index if it appears both as a subscript and as a superscript. To raise the subscript κ on $H_{ab\kappa}$ we need only multiply by $g^{\kappa'\kappa}$ since $g^{\kappa a} = 0$. Notice that

$$h(Z_a, Z_b) = H^{\kappa}_{ab} Z_\kappa ,$$
$$\tilde{\nabla}_{Z_a} Z_b = \tilde{\Gamma}^{\gamma}_{ab} Z_\gamma = \Gamma^{c}_{ab} Z_c + H^{\kappa}_{ab} Z_\kappa ,$$

where $\Gamma^{c}_{ab} = \langle \nabla_{Z_a} Z_b, Z_{c'} \rangle g^{c'c}$. If we let

$$R_{abcd} = \langle R(Z_a, Z_b) Z_c, Z_d \rangle ,$$
$$(R_D)_{ab\kappa\lambda} = \langle\!\langle R_D(Z_a, Z_b) Z_\kappa, Z_\lambda \rangle\!\rangle$$

be the components of the Riemannian curvature tensors on $T(M)$ and $T^{\perp}(M)$, respectively, then equations (4.4)–(4.6) become

$$\begin{aligned} R_{abcd} &= (H^{*}_{ad\kappa} H_{bc\lambda} - H_{ac\kappa} H^{*}_{bd\lambda}) g^{\kappa\lambda} , \\ (\partial_a H^{\kappa}_{bc}) - (\partial_b H^{\kappa}_{ac}) &= \Gamma^{\delta}_{ac} \Gamma^{\kappa}_{b\delta} - \Gamma^{\delta}_{bc} \Gamma^{\kappa}_{a\delta} , \qquad (4.7) \\ (R_D)_{ab\kappa\lambda} &= (H_{ac\lambda} H^{*}_{bd\kappa} - H^{*}_{ac\kappa} H_{bd\lambda}) g^{cd} , \end{aligned}$$

respectively. Notice that $-(R_D)_{ab\kappa\lambda} = (R_D^*)_{ab\lambda\kappa}$ and $-R_{abcd} = R^*_{abdc}$. Since R_{abcd} is skew-symmetric in ab, we have

$$R_{bacd} = R^*_{abdc}\ . \tag{4.8}$$

Equation (4.8) shows that the Riemannian *scalar curvature* $R = g^{bd}g^{ac}R_{bacd}$ is the same in the primal and the dual connection.

Many of the tensors in this and the previous section occur in statistical calculations. The imbedding curvature tensor $H_{ab\kappa}$ is particularly important as are its contractions

$$\gamma^2 = H_{ac\kappa}H^{\kappa}_{bd}g^{ad}g^{bc}\ ,$$

$$M^2 = H_{ad\kappa}H^{\kappa}_{bc}g^{ad}g^{bc}\ .$$

The first contraction γ^2 is called the statistical curvature, while M^2 is called the square of the mean curvature. The difference between these two quantities, $\bar{S} = M^2 - \gamma^2$, has been used by Amari (1985) to describe information loss in exponential families and will be used in the next section.

5. Bartlett correction in exponential family regression

McCullagh and Cox (1986) have given geometric interpretations for the Bartlett correction in normal theory regression and the results in this section are an extension of their work. We show the Bartlett correction for exponential family regression can be written in terms of geometric quantities that are used in many statistical calculations. In order to accomplish this task, we show how McCullagh's (1987) theory of tensor notation relates to the dual geometries.

First, we show how exponential family regression can be viewed as a statistical manifold. Let $Y = (Y^1, \ldots, Y^n)'$ be a random vector with density from a given exponential family of density functions S,

$$S = \left\{ p\colon p(y;\theta) = \exp\left(\sum_{i=1}^{n} \theta^i \eta_i(y) - \psi(\theta) \right) \right\}.$$

Usually, the components Y^i are independent and each marginal density belongs to the same exponential family so that any density $p \in S$ can be written as

$$p = p(y;\theta) = \prod_{i=1}^{n} f(y^i;\theta^i)\ ,$$

where $f(y^i;\theta^i)$ is a univariate exponential family with natural parameter

$\theta^i \in \mathbb{R}^1$. For the exponential families commonly used in practice (e.g., generalized linear models), we find that η_i is the identity function and that the parameterization $\theta = (\theta^1,\ldots,\theta^n)'$ typically defines a homeomorphism ϕ_θ between S with the weak topology and a space homeomorphic to $\mathbb{R}^n$. When S is a regular minimal standard exponential family, ϕ_θ is homeomorphic on its image. A weaker set of sufficient conditions to ensure that ϕ_θ is a homeomorphism are given in Theorem 8.3 of Barndorff-Nielsen (1978). Using ϕ_θ, the topological space S becomes a manifold that we also call S.

For each realization y^i of Y^i, we observe k covariates $x_1^i,\ldots,x_k^i$ and denote the $n \times k$ covariate matrix by X. (In this section and the next, vector fields will not be denoted by X or Y as they were in Sections 2, 3 and 4, so there will be no conflict of notation.) The exponential regression hypothesis states that the random vector Y has density with natural parameter

$$\theta = L(\beta, X)\,, \tag{5.1}$$

where $L(\cdot\,,\cdot)$ is a known function and $\beta = (\beta^1,\ldots,\beta^m)'$ is a column of unknown parameters. For a generalized linear model, $k = m$ and $L(\beta, X) = \underline{L}(X\beta)$ for another known function $\underline{L}(\cdot)$. In many applications, $L(\cdot\,, X)$ or $\underline{L}(\cdot)$ is a 1-1 immersion so that the set of all densities satisfying (5.1) can be made into a regular submanifold

$$M = \{p \in S\colon \phi_\theta(p) = L(\beta, X)\}\,.$$

To define a metric on S_p it is sufficient to consider the canonical basis $\{\partial_{ip} = \partial/\partial\theta^i|_{\theta(p)}\}_1^n$ for θ. Corresponding to each basis vector ∂_{ip} there exists a random variable $\partial_{ip}l = \partial_i l(\theta; Y)|_{\theta=\phi_\theta(p)}$ where $l(\theta; y) = \log f(y;\theta)$ is the log likelihood function. The random variable $\partial_{ip}l$ is called the 1-representation for ∂_{ip} (Amari (1985)). The metric can now be defined as follows

$$\langle\!\langle \partial_{ip}, \partial_{jp}\rangle\!\rangle = E(\partial_{ip}l\,\partial_{jp}l)\,. \tag{5.2}$$

The components of the metric with respect to the basis $\{\partial_{ip}\}$ are given by

$$g_{ij} = g_{ij}(p) = E(\partial_{ip}l\,\partial_{jp}l)\,,$$

the components of the expected Fisher information matrix for the parameter θ. When $l(\theta; Y)$ is a smooth function of θ^i, then $\partial_i l$ becomes a smooth vector field and the metric defined pointwise by (5.2) is also smooth.

We also use the 1-representation to define a pair of dual connections on S. If $(\tilde{\nabla}_{\partial_i}\partial_j)^{(1)}$ is the 1-representation for $\tilde{\nabla}_{\partial_i}\partial_j$, then

$$(\tilde{\nabla}_{\partial_i}\partial_j)^{(1)} = \partial_i\partial_j l(\theta^i; Y^i) - E(\partial_i\partial_j l)\,.$$

The components of this connection are given by

$$\tilde{\Gamma}_{ijk} = \langle\!\langle \tilde{\nabla}_{\partial_i}\partial_j, \partial_k \rangle\!\rangle = E(\partial_i\partial_j l \partial_k l)\,. \tag{5.3}$$

Notice that $\tilde{\Gamma}_{ijk}$ is a function on S and that the expectation and derivatives in (5.3) are evaluated at the same point $p \in S$. The components of the dual connection are given by

$$\tilde{\Gamma}^*_{ijk} = \tilde{\Gamma}_{ijk} + T_{ijk}\,, \tag{5.4}$$

where $T_{ijk} = E(\partial_i l \partial_j l \partial_k l)$ is known as the skewness tensor for S. The first connection $\tilde{\Gamma}_{ijk}$ is called the exponential connection and its dual $\tilde{\Gamma}^*_{ijk}$ is the mixture connection. It is a simple calculation to show that $\tilde{\Gamma}^*_{ijk}$ defined by (5.4) is dual to $\tilde{\Gamma}_{ijk}$ (Vos (1987)). Lauritzen (1987) discusses further properties of the tensor with components T_{ijk} and its characterization of statistical manifolds.

The metric $\langle \cdot, \cdot \rangle$ on M_p and the connection ∇ on $T(M)$ are the metric and connection induced from the corresponding quantities defined on the supermanifold S. The β parameterization on M defines a coordinate chart (ϕ_β, M) so that

$$\partial_r = \frac{\partial}{\partial \beta^r}$$

is the canonical basis field for β. Derivatives with respect to β will be represented with subscripts $r, s, t, u, \ldots$. For a fixed $p_0 \in M$, we can define a coordinate chart with a canonical basis vector field that extends $\{\partial_r\}$ to a basis vector field $\{\partial_\rho\}$ on all $T(S)$, such that $\{\partial_\rho\} = \{\partial_r, \partial_\kappa\}$ for $\rho = 1, \ldots, n$, $r = 1, \ldots, m$ and $\kappa = m+1, \ldots, n$ and $\langle\!\langle \partial_r, \partial_\kappa \rangle\!\rangle_{p_0} = 0$. In other words, $\{\partial_\kappa\}$ is a basis for $M_{p_0}^{\perp}$, the orthogonal complement of M_{p_0} in S_{p_0}. The components of the metric and dual connections on M with respect to $\{\partial_r\}$ will be denoted by g_{rs}, Γ_{rst}, Γ^*_{rst}, respectively. For notational simplicity we have not written ∂_{rp_0} and $\partial_{\kappa p_0}$ for the natural basis vectors in S_{p_0}. The context will determine whether ∂_r is vector in M_{p_0} or a vector field on M.

Having placed a pair of dual geometries on exponential family regression, we are now ready to consider the Bartlett correction. We consider the log of the likelihood ratio statistic

$$w(\beta_0) = 2\{l(\hat{\beta}; Y) - l(\beta_0; Y)\}$$

to test H_0: $\beta = \beta_0$ where $\beta_0 = \phi_\beta(p_0)$ and we have written $l(\beta; Y)$ for $l(\phi_\theta(\phi_\beta^{-1}(\beta)); Y)$. Under appropriate regularity conditions, $w(\beta_0)$ has asymptotically the chi-squared distribution with m degrees of freedom when $\beta = \beta_0$. The expectation of $w(\beta_0)$ can often be written as

$$E_{\beta_0}(w(\beta_0)) = m\left(1 + \frac{b(\beta_0)}{nm} + O(n^{-2})\right).$$

By dividing $w(\beta_0)$ by the *Bartlett correction*, $(1 + b(\beta_0)/n)$, the approximation to the chi-squared distribution is improved, often to $O(n^{-2})$ (Barndorff-Nielsen and Hall (1988)).

McCullagh and Cox (1986) give the following invariant expression for $b(\beta_0)$ in the Bartlett correction

$$(5.5)\qquad b(\beta_0) = \frac{1}{12}(3\bar{\rho}_{13}^2 + 2\bar{\rho}_{23}^2 - 3\bar{\rho}_4) + \frac{1}{4}m^{-1}v^{r,s}v^{t,u}(2v_{rt,su} - v_{rs,tu} - 2v_{r,s,tu}).$$

The first three quantities, $\bar{\rho}_{13}^2$, $\bar{\rho}_{23}^2$ and $\bar{\rho}_4$, are multivariate generalizations of the univariate measures of skewness (ρ_3^2) and kurtosis (ρ_4). The v's in the last term, call it $c(\beta_0)$, are defined in terms of the cumulants of the derivatives of the log likelihood and are tensors. For non-linear normal theory regression, McCullagh and Cox (1986) show that $c(\beta_0)$ reduces to a simple expression involving the Riemannian scalar curvature and the mean curvature. We extend their regression example to allow for error structures from an exponential family.

We begin by relating the notation of McCullagh and Cox (1986) to the tangent vectors in S_{p_0}. By equation (24) in McCullagh and Cox (1986),

$$(5.6)\qquad U_r = \partial_r l = \frac{\partial l(\beta; Y)}{\partial \beta^r}, \qquad U_{rs} = \partial_r \partial_s l = \frac{\partial^2 l(\beta; Y)}{\partial \beta^r \partial \beta^s},$$

so that U_r is just the 1-representation for the basis vector ∂_r and $U_{rs} - E(U_{rs})$ is the 1-representation for $\tilde{\nabla}_{\partial_r}\partial_s$. Substituting from equation (5.6) into the expression following (24) in McCullagh and Cox (1986), we have

$$n\kappa_r = E(U_r) = E(\partial_r l) = 0,$$

$$n\kappa_{rs} = E(U_{rs}) = E(\partial_r \partial_s l) = -g_{rs},$$

$$(5.7)\qquad n\kappa_{r,s} = \operatorname{cov}(U_r U_s) = E(\partial_r l \partial_s l) = g_{rs},$$

$$n\kappa_{r,st} = \operatorname{cov}(U_r, U_{st}) = E(\partial_r l \partial_s \partial_t l) = \Gamma_{str},$$

$$n\kappa_{r,s,t} = \operatorname{cum}(U_r, U_s, U_t) = E(\partial_r l \partial_s l \partial_t l) = T_{rst},$$

where the first equality in each line of (5.7) is the definition given in McCullagh and Cox (1986). These authors also use the following tensors

$$V_r = U_r \quad \text{and} \quad V_{rs} = U_{rs} - \beta_{rs}^a U_a ,$$

where

$$\beta_{rs}^a = \kappa^{a,b}\kappa_{b,rs} = g^{ab}\Gamma_{rsb}^{(1)} = \Gamma_{rs}^a ,$$

and $a = 1,\ldots,m$. Notice that $V_{rs} = \perp(U_{rs})$ so that $\bar{V}_{rs} = V_{rs} - E(U_{rs})$ is the 1-representation for $h(\partial_r, \partial_s)$ and $\beta_{rs}^a U_a = \Gamma_{rs}^a U_a$ is the 1-representation for $\nabla_{\partial_r}\partial_s$.

We are now ready to interpret the invariant quantities in the Bartlett correction. The ν's in the last term of the Bartlett correction are n^{-1} times the cumulants of V_r and V_{rs}; in particular,

$$\begin{aligned} n\nu_{r,s} &= \operatorname{cov}(V_r, V_s) , \\ n\nu_{rs,tu} &= \operatorname{cov}(V_{rs}, V_{tu}) , \qquad (5.8) \\ n\nu_{r,s,tu} &= \operatorname{cum}(V_r, V_s, V_{tu}) , \end{aligned}$$

and the matrix inverse of $\nu_{r,s}$ is $\nu^{r,s}$. From the second equation in (5.8), we see that

$$\begin{aligned} n\nu_{rs,tu} &= E(\bar{V}_{rs}, \bar{V}_{tu}) \qquad (5.9) \\ &= \langle\!\langle (h(\partial_r, \partial_s), h(\partial_t, \partial_u) \rangle\!\rangle \\ &= H_{rs}^\kappa g_{\kappa\lambda} H_{tu}^\lambda , \end{aligned}$$

where $H_{rs}^\kappa \partial_\kappa = h(\partial_r, \partial_s)$ so that H_{rs}^κ are the components of $h(\partial_r, \partial_s)$ with respect to $\{\partial_\kappa\}$. The trace of the exponential imbedding curvature $H^\kappa = g^{rs}H_{rs}^\kappa$ gives the components of the exponential mean curvature vector $H = H^\kappa\partial_\kappa \in M_p$. There is not complete agreement on terminology in this case; some authors (Spivak (1975), p. 96) call $m^{-1}H$ the mean curvature vector. The quantity

$$\begin{aligned} n^{-1}\nu^{r,s}\nu^{t,u}\nu_{rs,tu} &= g^{rs}g^{tu}(H_{rs}^\kappa H_{tu}^\lambda)g_{\kappa\lambda} \qquad (5.10) \\ &= H^\kappa H^\lambda g_{\kappa\lambda} \\ &= \|H\|^2 = M^2 \end{aligned}$$

can now be interpreted as the squared length of the exponential mean curvature vector. While

$$n^{-1}v^{r,s}v^{t,u}v_{rt,su} = g^{rs}g^{tu}H_{rt}^{\kappa}H_{su}^{\lambda}g_{\kappa\lambda} = \gamma^2 , \tag{5.11}$$

where γ^2 is the statistical curvature.

We have one more term to consider, namely

$$\begin{aligned} n v_{r,s,tu} &= \operatorname{cum}(V_r, V_s, V_{tu}) \\ &= E(V_r V_s H_{tu}^{\kappa} V_\kappa) \\ &= T_{rt\kappa} H_{tu}^{\kappa} . \end{aligned} \tag{5.12}$$

Using equation (5.4) we can express $T_{\rho\sigma\tau}$ as

$$T_{\rho\sigma\tau} = \tilde{\Gamma}^{*}_{\rho\sigma\tau} - \tilde{\Gamma}_{\rho\sigma\tau} ,$$

where the components are with respect to $\{\partial_\rho\} = \{\partial_r, \partial_\kappa\}$ with $\rho = 1,\dots,n$, $r = 1,\dots,m$ and $\kappa = m+1,\dots,n$. Recall that we chose $\{\partial_\rho\}$ such that $\langle \partial_r, \partial_\kappa \rangle_{p_0} = 0$ so that

$$T_{rs\kappa} = H_{rs\kappa} - H^{*}_{rs\kappa} ,$$

where $H^{*}_{rs\kappa}$ is the imbedding curvature tensor in the mixture connection which is dual to $H_{rs\kappa}$.

Making this substitution into equation (5.12) we obtain

$$n v_{r,s,tu} = H^{*\kappa}_{rs} H^{\lambda}_{tu} g_{\kappa\lambda} - H^{\kappa}_{rs} H^{\lambda}_{tu} g_{\kappa\lambda} . \tag{5.13}$$

We can make the same definitions for the mixture connection as for the exponential connection so that $H^{*\kappa} = g^{rs}H^{*\kappa}_{rs}$ are the components of the mean curvature vector for the mixture connection. Equation (5.13) can be contracted to give

$$n^{-1}v^{r,s}v^{t,u}v_{r,s,tu} = M^{11} - M^{2} , \tag{5.14}$$

where

$$M^{11} = H^{*\kappa}H^{\lambda}g_{\kappa\lambda} = \langle H^{*}, H \rangle . \tag{5.15}$$

Equation (5.15) shows that M^{11} can be interpreted as the inner product of the mean curvature vector in the exponential connection and the mean curvature vector in the mixture connection. The first superscript of M^{11} indicates the number of terms in the inner product that come from the exponential connection; the second superscript indicates the number from the mixture connection. Although $n^{-1}v^{r,t}v^{s,u}v_{r,s,tu}$ does not appear in (5.5), this contraction of $v_{r,s,tu}$ can be written in terms of geometric quantities,

$$n^{-1}v^{r,t}v^{s,u}v_{r,s,tu} = \gamma^{11} - \gamma^2 ,$$

where $\gamma^{11} = H_{rt\kappa}H^{*}_{su\lambda}g^{rs}g^{tu}g^{\kappa\lambda}$. In (5.18) we see that $\gamma^{11} - \gamma^2$ appears in an invariant expression for the Bartlett correction.

Making substitutions from (5.10), (5.11) and (5.14) into (5.5) we have

$$(5.16) \qquad c(\beta_0) = \frac{n}{4m}(2\gamma^2 + M^2 - 2M^{11}) .$$

When the error distribution is normal, (5.16) must reduce to $n(M^2 - 2R)/4m$, the expression for the Bartlett correction given by McCullagh and Cox (1986). To see this, we can use the Gauss equation (4.7)

$$R_{rstu} = (H_{rs\kappa}H^{*}_{tu\lambda} - H_{rt\kappa}H^{*}_{su\lambda})g^{\kappa\lambda} ,$$

since $\tilde{R}_{rstu} = 0$ for exponential families. Multiplying both sides by $g^{rs}g^{tu}$ we find

$$(5.17) \qquad R = g^{rs}g^{tu}R_{rstu} = M^{11} - \gamma^{11} .$$

Substituting (5.17) into (5.16) gives

$$(5.18) \qquad c(\beta_0) = \frac{n}{4m}\{(M^2 - 2R) + 2(\gamma^2 - \gamma^{11})\} .$$

Equations (5.16) and (5.18) are not the only invariant expressions for $c(\beta_0)$ in terms of statistical geometric quantities. Amari (1985) defines a quantity $\bar{S}_{rstu} = H_{rs\kappa}H_{tu\lambda} - H_{rt\kappa}H_{su\lambda}$ whose contraction

$$(5.19) \qquad \bar{S} = g^{rs}g^{tu}\bar{S}_{rstu} = M^2 - \gamma^2 ,$$

we call the second Riemannian scalar curvature since $\bar{S} = R$ for the self-dual connection. Substituting (5.19) into (5.16), we obtain

$$(5.20) \qquad c(\beta_0) = \frac{n}{4m}(-2\bar{S} + 3M^2 - 2M^{11}) .$$

For normal error distributions $\bar{S} = R$, so the second Riemannian scalar curvature is an extension of the scalar Riemannian curvature. Another interpretation for $\bar{S}$ follows from (5.19). Efron (1975) originally defined γ^2 only for 1-dimensional submanifolds. In this special case $\gamma^2 = M^2$, so that both γ^2 and M^2 are multi-dimensional generalizations for the original statistical curvature that Efron defined for one dimension. The second

Riemannian scalar curvature is simply the difference between these two extensions. Vos (1987) considers $\bar{S}$ and its relationship to R and $R^{(0)}$, the scalar Riemannian curvature in the metric connection.

6. Bartlett correction in statistical manifolds

Having given a rather detailed discussion of the Bartlett correction in exponential family regression, it is now a fairly simple matter to extend this interpretation to statistical manifolds outside the exponential family. We consider a set of n-dimensional distributions

$$M = \{p(y;\beta) \colon \beta \in \mathscr{B}\} ,$$

where $y = (y^1,\ldots,y^n)' \in \mathbb{R}^n$ and $\mathscr{B}$ is homeomorphic to $\mathbb{R}^m$. We also assume that the parameterization is such that M becomes a smooth manifold. We shall not require that $p(y;\beta)$ come from an exponential family, nor shall we require that M be a submanifold of some larger manifold of distributions. Rather, we approximate M at a point $p_0 \in M$ with a curved exponential family $\tilde{M}(p_0)$, called the local exponential family (Amari (1987)). Since we only consider the local exponential family at p_0, we can write $\tilde{M}$ for $\tilde{M}(p_0)$ without possibility of confusion.

In order to define $\tilde{M}$, we shall require the random variables

$$U_r = U_r(p_0) = \partial_r l|_{\beta_0}$$

and

$$U_{rs} = U_{rs}(p_0) = (\partial_r \partial_s l + g_{rs})|_{\beta_0} ,$$

where $\beta_0 = \phi_\beta(p_0)$, l is the log likelihood function, and $g_{rs} = -E(\partial_{rs} l)$ is the Fisher information matrix for β_0. We assume that span $\{U_{rs}\}$ has dimension $m_2 = m(m+1)/2$ and that the dimension of span $\{U_r, U_{st}\}$ is $m + m_2$. Following Amari (1987), we define an exponential family at p_0 that depends on an $(m + m_2)$-dimensional natural parameter θ with components θ^i

$$\tilde{S} = \{q \colon q(y;\theta) = p(y;\beta_0) \exp\{\theta^i U_i - \psi(\theta)\}\} ,$$

where $U_i = U_r$ if $i = r \le m$ and $U_i = U_{rs}$ if i is in the range $m+1,\ldots,m+m_2$. For values of i larger than m, we assign a value to each ordered pair rs where $r \le s$. Finally, we define the approximating local exponential family at p_0

$$\tilde{M} = \{\tilde{p} \in \tilde{S} \colon \tilde{p}(y;\beta') = q(y;\theta(\beta'))\} ,$$

which is indexed by the parameter β' so that $\theta^i(\beta') = (\beta' - \beta_0)^r$ when $i = r \leq m$ and $\theta^i(\beta') = (\beta' - \beta_0)^r(\beta' - \beta_0)^s$ when $i \geq m + 1$. Notice that $\tilde{p}(y;\beta') = p(y;\beta_0)$ when $\beta' = \beta_0$ and that $\tilde{U}_r = U_r$ and $\tilde{U}_{rs} = U_{rs}$ where $\tilde{U}_r = \partial_r' \log \tilde{p} = \partial / \partial \beta'^r \log \tilde{p}$ and $\tilde{U}_{rs} = \partial_r' \partial_s' \log \tilde{p}$ at $\beta' = \beta_0$.

Since each term in the Bartlett correction depends only on the first and second order derivatives of the log likelihood, we can interpret each term using the approximating local exponential family $\tilde{M}$. In particular, the scalars $\bar{\rho}_{13}^2$, $\bar{\rho}_{23}^2$ and $\bar{\rho}_4$ are the multivariate measures of skewness and kurtosis for the curved exponential family $\tilde{M}$. Furthermore, since V_{rs} is defined using only U_r, U_{rs}, and the expectations of these random variables at $\beta' = \beta$, V_{rs} is the 1-representation for the exponential imbedding curvature of $\tilde{M}$ in $\tilde{S}$. If we assign a value κ from $m + 1, \ldots, m + m_2$ to each rs where $r \leq s$, then V_κ spans the 1-representation for $\tilde{M}_{p_0}^{\perp}$, the orthogonal complement of $\tilde{M}_{p_0}$ in $\tilde{S}_{p_0}$. We can write $V_{rs} = H_{rs}^{\kappa} V_\kappa$ so that H_{rs}^{κ} are the components of the exponential imbedding curvature of $\tilde{M}$ in $\tilde{S}$. Hence,

$$nv_{rs,tu} = \operatorname{cov}(V_{rs}, V_{tu}) = H_{rs}^{\kappa} H_{tu}^{\lambda} g_{\kappa\lambda}\,, \tag{6.1}$$

where $g_{\mu\lambda} = \operatorname{cov}(V_\kappa, V_\lambda)$. Furthermore, if we let $T_{r,s,\kappa} = \operatorname{cum}(U_r, U_s, V_\kappa)$, then

$$\begin{aligned} nv_{r,s,tu} &= \operatorname{cum}(U_r, U_s, V_{tu}) \\ &= H_{tu}^{\kappa} T_{r,s,\kappa} \\ &= H_{tu}^{\kappa}(H_{rs\kappa}^{*} - H_{rs\kappa})\,. \end{aligned} \tag{6.2}$$

The last equality follows from the definition of $H_{rs\kappa}^{*}$, the components of the mixture imbedding curvature. Notice that equations (6.1) and (6.2) are identical to (5.9) and (5.13), except that H_{rs}^{κ} and $H_{rs}^{*\kappa}$ are now the imbedding curvatures of $\tilde{M}$ in $\tilde{S}$. Since the Bartlett correction involves $\bar{\rho}_{13}^2, \bar{\rho}_{23}^2, \bar{\rho}_4$ and contractions of (6.1) and (6.2), we see that for a general statistical manifold the Bartlett correction takes the same form as for exponential family regression (equations (5.16), (5.18) and (5.20)). The only difference is that the scalar curvatures $M^2, \gamma^2, \bar{S}$ and R and the invariants M^{11} and γ^{11} describe how the approximating exponential family $\tilde{M}$ is imbedded in $\tilde{S}$.

7. Conclusion

The dual geometries can be applied to a variety of situations and when these geometries are used in statistical inference a number of curvature tensors and other invariant quantities arise. In order to relate these quantities we have derived the fundamental equations for statistical submanifolds. To illustrate how these equations can be utilized, we have considered the Bartlett correction for which we have given an invariant expression using curvatures that describe how a statistical submanifold is curving in a supermanifold. Each term in this expression is some contraction of the exponential imbedding curvature and/or its dual. The exponential imbedding curvature as well as its contractions, the statistical curvature and the square of the mean curvature, appear in many statistical calculations. The Gauss equation for statistical submanifolds is useful because it relates these imbedding curvatures to the Riemannian scalar curvature. The extension of this result beyond exponential family regression is a simple matter if one uses Amari's (1987) local exponential family.

Acknowledgements

This work originated while the author was at the University of Chicago under support from the McCormick Fellowship and National Science Foundation grants MCS 83-01459, DMS 84-12233 and DMS 86-01732. Additional support was provided by National Science Foundation grant DMS 88-03584.

Appendix 1

PROPOSITION A1.1. *If $\tilde{X}$ and $\tilde{Y}$ are extensions to S of the vector fields X and Y on M, respectively, then $\tilde{\nabla}_{\tilde{X}}\tilde{Y}|_M$ is independent of the extension.*

PROOF. In the following, we will use a canonical basis $\{\partial_i\}_1^n$ on S and a canonical basis $\{\partial_a\}_1^m$ on M, so that $\tilde{X} = \tilde{X}^i\partial_i$ and $X = X^a\partial_a$ for smooth component functions $\tilde{X}^i$ and X^a. Notice that $\partial_i \neq \partial_a$ even when $i = a$. Since $M_p \subset S_p$, there is a linear mapping $B(p)$ with components $B_a^i(p)$ at each p such that $\partial_{ap} = B_a^i(p)\partial_{ip}$. As a function on M, $B_a^i \in C^\infty(M)$ for all i, a and

$$\tilde{X}^i|_M = B_a^i X^a, \tag{A1.1}$$

since $\tilde{X}^i\partial_i|_M = X^a\partial_a|_M = X^a B_a^i \partial_i|_M$. Writing $\tilde{X}$ and $\tilde{Y}$ in terms of the canonical basis on S, we obtain

$$\tilde{\nabla}_{\tilde{X}}\tilde{Y} = (\tilde{X}^i\partial_i\tilde{Y}^j)\partial_j + \tilde{X}^i\tilde{Y}^j\tilde{\nabla}_{\partial_i}\partial_j. \tag{A1.2}$$

From (A1.1) and (A1.2), we see that

$$\tilde{\nabla}_{\tilde{X}}\tilde{Y}|_M = X^a\partial_a(B_b^j Y^b)\partial_j + B_a^i X^a B_b^j Y^b \tilde{\nabla}_{\partial_i}\partial_j. \qquad \square$$

Appendix 2

PROOF OF PROPOSITION 4.1. Recall that $X, Y, Z \in T(M)$ and $\tilde{R}(\cdot,\cdot)$

is the curvature tensor of S defined by

$$\tilde{R}(X, Y)Z = \tilde{\nabla}_X\tilde{\nabla}_Y Z - \tilde{\nabla}_Y\tilde{\nabla}_X Z - \tilde{\nabla}_{[X,Y]}Z .$$

Substituting from the Gauss formula (3.3) we find

$$\begin{aligned}\tilde{R}(X, Y)Z = {} & \tilde{\nabla}_X(\nabla_Y Z + h(Y, Z)) - \tilde{\nabla}_Y(\nabla_X Z + h(X, Z)) \\ & - (\nabla_{[X,Y]}Z + h([X, Y], Z)) \\ = {} & R(X, Y)Z + h(X, \nabla_Y Z) - h(Y, \nabla_X Z) - h([X, Y], Z) \\ & + \tilde{\nabla}_X h(Y, Z) - \tilde{\nabla}_Y h(X, Z) ,\end{aligned}$$

where $R(\cdot, \cdot)$ is the curvature tensor of M. Now, substituting from the Weingarten formula (3.12a) we obtain

$$\begin{aligned}\tilde{R}(X, Y)Z = {} & R(X, Y)Z - A^*_{h(Y,Z)}X + A^*_{h(X,Z)}Y \\ & + h(X, \nabla_Y Z) - h(Y, \nabla_X Z) - h([X, Y], Z) \\ & + D_X h(Y, Z) - D_Y h(X, Z) .\end{aligned} \tag{A2.1}$$

Equation (A2.1) immediately gives us the equation of Codazzi (4.2) since ∇ is torsion-free; the equation of Gauss (4.1) follows by taking $W \in T(M)$ and applying equation (3.9b).

Finally, we use the following calculation

$$\begin{aligned}\langle\!\langle \tilde{R}(X, Y)\xi, \eta\rangle\!\rangle = {} & \langle\!\langle \tilde{\nabla}_X\tilde{\nabla}_Y\xi, \eta\rangle\!\rangle - \langle\!\langle \tilde{\nabla}_Y\tilde{\nabla}_X\xi, \eta\rangle\!\rangle \\ & - \langle\!\langle \tilde{\nabla}_{[X,Y]}\xi, \eta\rangle\!\rangle \\ = {} & - \langle\!\langle \tilde{\nabla}_X(A^*_\xi Y), \eta\rangle\!\rangle + \langle\!\langle \tilde{\nabla}_X D_Y\xi, \eta\rangle\!\rangle \\ & + \langle\!\langle \tilde{\nabla}_Y(A^*_\xi X), \eta\rangle\!\rangle - \langle\!\langle \tilde{\nabla}_Y D_X\xi, \eta\rangle\!\rangle \\ & - \langle\!\langle D_{[X,Y]}\xi, \eta\rangle\!\rangle \\ = {} & - \langle\!\langle h(X, A^*_\xi Y), \eta\rangle\!\rangle + \langle\!\langle h(Y, A^*_\xi X), \eta\rangle\!\rangle \\ & + \langle\!\langle D_X D_Y\xi, \eta\rangle\!\rangle - \langle\!\langle D_Y D_X\xi, \eta\rangle\!\rangle \\ & - \langle\!\langle D_{[X,Y]}\xi, \eta\rangle\!\rangle \\ = {} & - \langle A_\eta A^*_\xi Y, X\rangle + \langle A_\eta A^*_\xi X, Y\rangle \\ & + \langle\!\langle R_D(X, Y)\xi, \eta\rangle\!\rangle .\end{aligned}$$

The equation of Ricci (4.3) follows from the last equality upon noting that the symmetry of $h(X, Y)$ in (3.9a) and (3.9b) implies

$$\langle A_\eta A_\xi^* Y, X\rangle = \langle A_\xi^* Y, A_\eta X\rangle = \langle Y, A_\xi^* A_\eta X\rangle . \qquad \square$$

REFERENCES

Amari, S. (1985). *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, **28**, Springer, New York.

Amari, S. (1986). Geometrical theory on manifolds of linear systems, *METR*, **86-1**, University of Tokyo.

Amari, S. (1987). Differential geometrical theory of statistics, referred in Amari *et al.* (1987).

Amari, S., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L. and Rao, C. R. (1987). *Differential Geometry in Statistical Inference*, Institute of Mathematical Statistics, Monograph Series, **10**, Hayward, California.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Barndorff-Nielsen, O. E. (1986). Likelihood and observed geometries, *Ann. Statist.*, **14**, 856–873.

Barndorff-Nielsen, O. E. and Blæsild, P. (1986). A note on the calculation of Bartlett adjustments, *J. Roy. Statist. Soc. Ser. B*, **48**, 353–358.

Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator, *J. Roy. Statist. Soc. Ser. B*, **46**, 483–495.

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika*, **75**, 374–378.

Chen, B.-Y. (1984). *Total Mean Curvature and Manifolds of Finite Type*, World Scientific, Singapore.

Efron, B. (1975). Defining the curvature of a statistical problem, *Ann. Statist.*, **6**, 1189–1242.

Lauritzen, S. L. (1987). Statistical manifolds, referred in Amari *et al.* (1987).

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood-ratio criteria, *Biometrika*, **43**, 295–303.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman & Hall, New York.

McCullagh, P. and Cox, D. R. (1986). Invariants and likelihood ratio statistics, *Ann. Statist.*, **14**, 1419–1430.

Møller, J. (1986). Bartlett adjustments for structured covariances, *Scand. J. Statist.*, **13**, 1–15.

Spivak, M. (1975). *A Comprehensive Introduction to Differential Geometry*, Volume IV, Publish or Perish, Boston.

Vos, P. W. (1987). Dual geometries and their applications to generalized linear models, Ph. D. Dissertation, University of Chicago.