Ann. Inst. Statist. Math.
Vol. 43, No. 1, 95–113 (1991)

YOKES AND TENSORS DERIVED FROM YOKES

P. BLÆSILD

Department of Theoretical Statistics, Institute of Mathematics, Aarhus University, DK-8000, Aarhus C, Denmark

(Received June 19, 1989; revised November 30, 1989)

Abstract. A yoke on a differentiable manifold M gives rise to a whole family of derivative strings. Various elemental properties of a yoke are discussed in terms of these strings. In particular, using the concept of intertwining from the theory of derivative strings it is shown that a yoke induces a family of tensors on M. Finally, the expected and observed α-geometries of a statistical model and related tensors are shown to be derivable from particular yokes.

Key words and phrases: Bartlett adjustment factor, contrast functions, derivative strings, expected geometries, exponential family, intertwining, observed geometries, statistical manifold, tensorial components, yoke.

1. Introduction

The concept of a *yoke* was introduced in Barndorff-Nielsen (1988*a*) and has been further discussed by Barndorff-Nielsen (1987*a*, 1988*b*) and Blæsild (1987*a*). A yoke on a differentiable manifold M induces a whole family of connections on M. In Section 4, where the definition of a yoke is reviewed, it is shown that the *expected* and *observed α-geometries* of a statistical model introduced by Chentsov and Amari (cf. Amari (1985, 1987)) and Barndorff-Nielsen (1986*a*, 1987*b*), respectively, are particular instances of geometries derived from yokes. Consequently, these statistical geometries may be studied within a unified framework.

Some quantities derived from a statistical model, for instance the Bartlett adjustment factor of the log likelihood ratio statistic, are known to be parametrization invariant or, equivalently, if the statistical model is considered as a differentiable manifold these quantities are geometrical quantities. Typically, the quantities are expressed as a sum of invariant terms each such term being a contraction (product) of *tensors*. At the end of Section 5 it is illustrated that the most commonly used tensors in statistical theory are related to the *expected yoke* or to the *observed yoke*, introduced in Section 4. More generally, a yoke on M induces via the concept of *derivative strings* a whole family of tensors. The relevant part of the theory of derivative strings is reviewed in Subsection 2.2. After the definition of a yoke in Section 4, we show that a yoke gives rise to a whole family of derivative strings including what we refer to as the *α-connection string*, which is a generalization of the α-connection. Furthermore, we give in Section 4 various elemental properties of yokes. Examples 2 and 3 are concerned with, respectively, the expected and observed yokes of a statistical model, which produce the above-mentioned α-geometries of the model. Moreover, in Example 2 we discuss the relation between the expected geometries and the geometries introduced by Eguchi (1983, 1985) by means of contrast functions (or functionals), and in Example 3 we discuss observed geometries similar to those of Eguchi.

Tensors are obtained from derivative strings by means of the concept of *intertwining*. After giving the relevant results concerning intertwining in the beginning of Section 5, we derive various properties of the tensors corresponding to some of the derivative strings induced by a yoke. Furthermore, we show in Section 5 that tensors of statistical interest, such as the skewness tensor, the α-curvature tensor and the tensors entering the invariant expression for the Bartlett adjustment factor, can be expressed in terms of the tensors derived from a yoke. Finally, we illustrate the well-known fact that the concept of a statistical manifold, introduced by Lauritzen (1987), is insufficient for the discussion of asymptotical statistical theory and we comment on a proposal of extending the definition of Lauritzen.

Subsection 2.1 is concerned with notation and local coordinates of the manifold M. The definition and properties of a yoke given in Section 4 are formulated in terms of general, non-standard operations on functions defined on $M \times M$. These operations are reviewed in Section 3 in terms of local coordinates. However, using the theory of derivative strings it is shown that the operations are parametrization invariant. As mentioned above Subsection 2.2 gives a short review of those concepts from this theory which are relevant in the present context.

2. Notation, local coordinates and derivative strings

2.1 *Local coordinates*

Throughout the paper M denotes a d-dimensional differentiable manifold. A chart around $p \in M$ is a pair (U, ω) consisting of an open neighbourhood U around p and a homeomorphism ω from U onto an open subset of R^d. We speak of $\omega = (\omega^1, \ldots, \omega^d)$ as a set of *local coordinates* and use the letters $i, j, k, l, \ldots$ to denote arbitary components of ω. Since we shall be concerned with local properties only we often implicitly assume that M can be covered by one chart. The set of realvalued smooth functions whose domain of definition includes some open neighbourhood of p is denoted by $C_p^\infty M$. We write ∂_{kp} or just ∂_k for the coordinate frames, corresponding to ω, of the tangent space at p.

We let K_t signify a set $k_1 \cdots k_t$ of t indices related to the local coordinates ω with the convention that K_0 is the empty set. For $f \in C_p^\infty M$ and $t, \tau = 0, 1, 2, \ldots$ we let $f_{/K_t}(p)$ and $f_{K_t/\tau}(p)$ or just $f_{/K_t}$ and $f_{K_t/\tau}$ be defined by

$$f_{/K_t}(p) = f_{/K_t} = \partial_{K_t} f = \partial_{k_1} \cdots \partial_{k_t} f \tag{2.1}$$

and

$$f_{K_t/\tau}(p) = f_{K_t/\tau} = \sum_{K_t/\tau} f_{/K_{t1}} \cdots f_{/K_{t\tau}}, \tag{2.2}$$

where the right-hand sides are to be understood as $f(p)$ and 1, respectively, if $t = 0$. In (2.2) and in the rest of the paper $\sum_{K_t/\tau}$ signifies that the summation is over all *ordered* partitions of $K_t = k_1 \cdots k_t$ into τ (non-empty) subsets $K_{t1}, \ldots, K_{t\tau}$ such that the order of the indices in each of these subsets is the same as their order in K_t and such that for $\mu = 1, \ldots, \tau - 1$ the first index in $K_{t\mu}$ comes before the first index in $K_{t,\mu+1}$ as compared with the ordering within K_t. For $\tau > t$ the sum is to be interpreted as 0. The number of indices in the subset $K_{t\mu}$ is denoted $|K_{t\mu}|$.

Let $\psi = (\psi^1, \ldots, \psi^d)$ be an alternative set of local coordinates for which arbitrary components are denoted by the letters $a, b, c, d, \ldots$. For $t, \tau = 0, 1, 2, \ldots$ and for two sets of indices C_t and K_τ related to the local coordinates ψ and ω, respectively, we set

$$\omega^{K_\tau}_{/C_t} = \begin{cases} 1 & \text{if } \tau = t = 0, \\ \sum_{C_t/\tau} \omega^{k_1}_{/C_{t1}} \cdots \omega^{k_\tau}_{/C_{t\tau}} & \text{if } 1 \le \tau \le t, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Throughout the paper the Einstein summation convention is adopted and with this convention one has

$$\partial_{C_t} = \sum_{\tau=1}^{t} \omega^{K_\tau}_{/C_t} \partial_{K_\tau}. \tag{2.4}$$

Finally, we use the generalized Kronecker delta

$$\delta^{J_s}_{K_t} = \begin{cases} 1 & \text{if } J_s = K_t, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

2.2 *Derivative strings*

The discussion of derivative strings in this subsection has been extracted from Barndorff-Nielsen *et al.* (1988) which reviews and extends the papers on that subject by Barndorff-Nielsen (1986*b*) and Barndorff-Nielsen and Blæsild (1987*a*, 1987*b*, 1988).

A *derivative string* on a manifold M is defined relative to a set of local coordinates ω as a collection $\bar{H}$ of arrays

$$H^{I_r L_u}_{J_s K_t}, \quad t = 1, \ldots, T, \quad u = 1, \ldots, U, \tag{2.6}$$

satisfying the transformation law

$$H^{A_r D_u}_{B_s C_t} = \sum_{\tau=1}^{t} \sum_{v=u}^{U} H^{I_r L_v}_{J_s K_\tau} \omega^{K_\tau}_{/C_t} \psi^{D_u}_{/L_v} \psi^{A_r}_{/I_r} \omega^{J_s}_{/B_s} \tag{2.7}$$

for $t = 1, \ldots, T$ and $u = 1, \ldots, U$. The string $\bar{H}$ is said to be of *tensorial degree* (r, s) and of *length* (T, U), and we denote the class of all such strings on M by $\mathcal{S}^{rU}_{sT}$. We do, in fact, allow that some of the numbers r, s, T and U are 0, in which case the relevant groups of indices do not occur in (2.6) and (2.7). In particular, if

$U = 0$ and $T > 0$ we speak of (2.6) as a *costring* (or a (r, s) costring), while (2.6) is called a *contrastring* provided $U > 0$ and $T = 0$. We refer to the indices $i_1 \cdots i_r$ and $j_1 \cdots j_s$ of (2.6) as *tensorial indices* and to $k_1 \cdots k_t$ and $l_1 \cdots l_u$ as *structural indices.* Furthermore, the indices $i_1 \cdots i_r$ and $l_1 \cdots l_u$ are called *contravariant indices* in contrast to the indices $j_1 \cdots j_s$ and $k_1 \cdots k_t$ which are referred to as *covariant indices.*

A derivative string $\bar{H}$ is *structurally symmetric* if the arrays in (2.6) are invariant under arbitrary permutations of structural indices within each group of such indices.

The transformation law (2.7) generalizes those for tensors, affine connections, and derivatives of scalars. Accordingly, members of the particular classes $\mathcal{S}_{0\infty}^{10}$ and $\mathcal{S}_{0\infty}^{00}$ are referred to, respectively, as *connection strings* and *scalar strings.*

Important classes of derivative strings are the so-called co- and contrastrings generated by a connection string. If $\bar{\Gamma}$ denotes any connection string of infinite length and with $\Gamma^i_{k_1} = \delta^i_{k_1}$, then for r fixed $r = 2, 3, \ldots$ the set of arrays $\{\Gamma^{I_r}_{K_t}: t = 1, 2, \ldots\}$, where

$$(2.8) \qquad \Gamma^{I_r}_{K_t} = \sum_{K_t/r} \Gamma^{i_1}_{K_{t1}} \cdots \Gamma^{i_r}_{K_{tr}},$$

constitutes a $(r, 0)$ derivative costring of infinite length, which is referred to as the $(r, 0)$ *derivative costring generated by* $\bar{\Gamma}$. The elements of $\{G^{L_u}_{J_s}: u = 1, \ldots, s\}$, *the* $(0, s)$ *derivative contrastring generated by* $\bar{\Gamma}$, are determined recursively from the equations

$$(2.9) \qquad \sum_{s=u}^{t} G^{L_u}_{J_s} \Gamma^{J_s}_{K_t} = \delta^{L_u}_{K_t}.$$

If ∇ denotes a connection on the tangent bundle of M, one has that $\bar{\Gamma} = \{\Gamma^i_{K_t}: t = 1, 2, \ldots\}$, where $\Gamma^i_{K_t}$ is determined by

$$\nabla_{\partial_{k_t}}(\cdots(\nabla_{\partial_{k_2}}\partial_{k_1})\cdots) = \Gamma^i_{K_t}\partial_i,$$

is an element of $\mathcal{S}_{0\infty}^{10}$ with $\Gamma^i_{k_1} = \delta^i_{k_1}$ and we refer to $\bar{\Gamma}$ as *the canonical connection string* corresponding to ∇. Similarly, the *canonical derivative co- and contrastrings* corresponding to ∇ are the co- and contrastrings generated by $\bar{\Gamma}$.

Multiple derivative strings are strings with more than one set of structural covariant or structural contravariant indices. In this paper we consider only *double derivative costrings of degree* $(0, 0)$ *and of length* (T, U), i.e. collections $\bar{H}$ of arrays of the form

$$(2.10) \qquad H_{K_t M_u}, \quad t = 1, \ldots, T, \quad u = 1, \ldots, U$$

satisfying the transformation law

$$(2.11) \qquad H_{C_t D_u} = \sum_{\tau=1}^{t} \sum_{\upsilon=1}^{u} H_{K_\tau M_\upsilon} \omega^{K_\tau}_{/C_t} \omega^{M_\upsilon}_{/D_u}.$$

The set of such strings is denoted $\mathcal{S}^{00}_{0(T,U)}$ and a double costring is structurally symmetric if the arrays in (2.10) are invariant under arbitrary permutations of the indices within each of the sets K_t and M_u.

Here we have given coordinate-bound definitions of derivative strings. In Barndorff-Nielsen and Blæsild (1988) it is shown that (multiple) structurally symmetric derivative strings are smoothly varying multilinear forms on spaces which are suitable products of the tangent bundle of M, the zero-truncated jet bundles (of various orders) of M and the duals of such bundles.

3. Functions on $M \times M$ and related strings

The product space $M \times M = \{(p, \tilde{p})\colon\ p \in M,\ \tilde{p} \in M\}$ is a differentiable manifold of dimension $2d$. If $\omega = (\omega^1, \dots, \omega^d)$ and $\tilde{\omega} = (\tilde{\omega}^1, \dots, \tilde{\omega}^d)$ denote local coordinates in neighbourhoods U and $\tilde{U}$ around $p \in M$ and $\tilde{p} \in M$, respectively, then

$$(\omega, \tilde{\omega}) = (\omega^1, \dots, \omega^d, \tilde{\omega}^1, \dots, \tilde{\omega}^d)$$

is a set of local coordinates around $(p, \tilde{p}) \in M \times M$, which we refer to as a set of *product coordinates*. Taking $p = \tilde{p}$, $\omega = \tilde{\omega}$ and $U = \tilde{U}$ one obtains, that

$$(\omega, \omega)(q, \tilde{q}) = (\omega(q), \omega(\tilde{q})), \qquad (q, \tilde{q}) \in U \times U,$$

defines a set of local coordinates around (p, p). This point belongs to the differentiable manifold $\operatorname{diag} M \times M = \{(p, p)\colon\ p \in M\}$. A set of local coordinates on this d-dimensional manifold is of the form

$$(\omega, \omega)(q, q) = (\omega(q), \omega(q)), \qquad (q, q) \in \operatorname{diag} U \times U.$$

For $f \in C^\infty(M \times M)$ and for a set of product coordinates $(\omega, \tilde{\omega})$ we let for $t = 0, 1, 2, \dots$ and $u = 0, 1, 2, \dots$

$$f_{K_t;M_u}(\omega, \tilde{\omega}) = \partial_{K_t} \tilde{\partial}_{M_u} f(\omega, \tilde{\omega}) = \partial_{k_t} \cdots \partial_{k_1} \tilde{\partial}_{m_u} \cdots \tilde{\partial}_{m_1} f(\omega, \tilde{\omega}) \tag{3.1}$$

and

$$\not{f}_{K_t;M_u}(\omega) = f_{K_t;M_u}(\omega, \omega). \tag{3.2}$$

Note that the functions $\not{f}_{;}$ are defined on M by considering the restrictions of $f_{;}$ defined on $M \times M$, to $\operatorname{diag} M \times M$.

Letting ψ, $\tilde{\psi}$ denote an alternative set of product coordinates such that $\psi = \psi(\omega)$ and $\tilde{\psi} = \psi(\tilde{\omega})$ one has, using (2.4), the following transformation formula

$$f_{C_t;D_u} = \partial_{C_t} \tilde{\partial}_{D_u} f = \left(\sum_{\tau=1}^{t} \sum_{\upsilon=1}^{u} \omega^{K_\tau}_{/C_t} \tilde{\omega}^{M_\upsilon}_{/D_u} \partial_{K_\tau} \tilde{\partial}_{M_\upsilon} \right) f = \sum_{\tau=1}^{t} \sum_{\upsilon=1}^{u} f_{K_\tau;M_\upsilon} \omega^{K_\tau}_{/C_t} \tilde{\omega}^{M_\upsilon}_{/D_u}$$

and consequently

$$\not{f}_{C_t;D_u} = \sum_{\tau=1}^{t} \sum_{\upsilon=1}^{u} \not{f}_{K_\tau;M_\upsilon} \omega^{K_\tau}_{/C_t} \omega^{M_\upsilon}_{/D_u}. \tag{3.3}$$

It follows from (3.3) that $f_{;} = \{f_{K_t;M_u}\colon\ t,\, u = 1, 2, \ldots\} \in \mathcal{S}^{00}_{0(\infty,\infty)}$, i.e. $f_{;}$ is a double costring of degree (0, 0) and of length (∞, ∞), and that $\{f_{K_t;}\colon\ t = 1, 2, \ldots\}$ and $\{f_{;M_u}\colon\ u = 1, 2, \ldots\}$ are elements of $\mathcal{S}^{00}_{0\infty}$. Furthermore, by restricting the length of one of the set of structural indices of $f_{;}$ to 1, i.e. by considering

$$\{f_{K_t;m}\ :\ t = 1, 2, \ldots\} \quad \text{or} \quad \{f_{k;M_u}\ :\ u = 1, 2, \ldots\}$$

one obtains a costring of degree (0, 1) and length ∞, as seen from (3.3).

For later use note that

$$\partial_n f_{K_t;M_u} = f_{K_t \cup n;M_u} + f_{K_t;M_u \cup n}. \tag{3.4}$$

4. Yokes

With the notation introduced in Section 3 a yoke is defined in the following way. A function $g \in C^\infty(M \times M)$ is called a *yoke* if for every ω we have that

$$g_{k;}(\omega) = 0, \qquad k = 1, \ldots, d \tag{4.1}$$

and that

$$\text{the matrix} \quad \{-g_{k_1 k_2;}(\omega)\} \quad \text{is positive definite.} \tag{4.2}$$

Example 1. In the terminology of Barndorff-Nielsen (1978), let M be a regular exponential family of order d and with minimal representation

$$\exp\{\omega^i t_i(x) - \kappa(\omega) - \phi(x)\}.$$

Then it is easily seen that the function $g \in C^\infty(M \times M)$ given by

$$g(\omega, \tilde{\omega}) = (\omega - \tilde{\omega})^i \kappa_{/i}(\tilde{\omega}) - \kappa(\omega) + \kappa(\tilde{\omega}) \tag{4.3}$$

is a yoke.

The concept of a yoke was introduced in Barndorff-Nielsen (1988*a*). A yoke gives rise to a whole family of connections on M as shown below. The *expected geometries* and the *observed geometries* of a statistical model may be derived from particular yokes as demonstrated in Examples 2 and 3 below. In these examples we also comment on the (expected) geometries of Eguchi (1983, 1985) defined by means of contrast functions. However, first we derive some basic properties of general yokes.

For a yoke g one has for $t = 1, 2, \ldots$, that

$$g_{K_t;} + \sum_{K_t/2} g_{K_{t1};K_{t2}} = 0 \tag{4.4}$$

which was proved in Blæsild (1988) by an induction argument based on (4.1) and using (3.4).

From (4.4) with $t = 2$ we obtain

$$\not{g}_{k_1;k_2} = -\not{g}_{k_1 k_2;} \tag{4.5}$$

and formulas (4.2) and (4.5) imply that the matrix $\{\not{g}_{k_1;k_2}\}$ is a (symmetric) positive definite matrix whose inverse will be denoted by $\{\not{g}^{k_1;k_2}\}$.

Furthermore, from Section 3 one finds that

$$\not{g}_; = \{\not{g}_{K_t;M_u} : t, u = 1, 2, \ldots\} \in \mathcal{S}^{00}_{0(\infty,\infty)}, \tag{4.6}$$

$$\overset{1}{\not{g}}_1 = \left\{\overset{1}{\not{g}}_{jK_t} = \not{g}_{K_t;j} : t = 1, 2, \ldots\right\} \in \mathcal{S}^{00}_{1\infty}, \tag{4.7}$$

$$\overset{-1}{\not{g}}_1 = \left\{\overset{-1}{\not{g}}_{jK_t} = \not{g}_{j;K_t} : t = 1, 2, \ldots\right\} \in \mathcal{S}^{00}_{1\infty}, \tag{4.8}$$

$$\overset{1}{\not{g}}_0 = \left\{\overset{1}{\not{g}}_{K_t} = \not{g}_{K_t;} : t = 1, 2, \ldots\right\} \in \mathcal{S}^{00}_{0\infty} \tag{4.9}$$

and

$$\overset{-1}{\not{g}}_0 = \left\{\overset{-1}{\not{g}}_{K_t} = \not{g}_{;K_t} : t = 1, 2, \ldots\right\} \in \mathcal{S}^{00}_{0\infty}. \tag{4.10}$$

Since $\mathcal{S}^{00}_{1\infty}$ is a vector space, it follows for $\alpha \in R$ that the set $\overset{\alpha}{\not{g}}_1$ of arrays

$$\overset{\alpha}{\not{g}}_{jK_t} = \frac{1+\alpha}{2}\overset{1}{\not{g}}_{jK_t} + \frac{1-\alpha}{2}\overset{-1}{\not{g}}_{jK_t} = \frac{1+\alpha}{2}\not{g}_{K_t;j} + \frac{1-\alpha}{2}\not{g}_{j;K_t} \qquad t = 1, 2, \ldots \tag{4.11}$$

constitutes a (0, 1) costring of infinite length.

Formula (4.6) implies that the matrix $\{\not{g}_{k;m}\}$ is a covariant tensor of degree 2 and, consequently, a Riemannian metric on M. Together with formulas (4.1), (4.2) and (4.5) this fact implies that the concept of a yoke is a geometrical concept, i.e. a yoke is invariant under transformations of the local coordinates in terms of which it is defined. Moreover, raising the tensorial index of the string $\overset{\alpha}{\not{g}}_1$ by means of the contravariant tensor $\{\not{g}^{i;j}\}$ one obtains a connection string $\overset{\alpha}{\not{g}}{}^1$ of infinite length whose elements are

$$\overset{\alpha}{\not{g}}{}^i_{K_t} = \overset{\alpha}{\not{g}}_{jK_t}\not{g}^{i;j} = \frac{1+\alpha}{2}\not{g}_{K_t;j}\not{g}^{i;j} + \frac{1-\alpha}{2}\not{g}_{j;K_t}\not{g}^{i;j}. \tag{4.12}$$

Since $\overset{\alpha}{\not{g}}{}^1 \in \mathcal{S}^{10}_{0\infty}$ and $\overset{\alpha}{\not{g}}{}^i_{k_1} = \delta^i_{k_1}$ one has, in particular, that

$$\overset{\alpha}{\not{g}}{}^a_{c_1c_2} = \left\{\overset{\alpha}{\not{g}}{}^i_{k_1k_2}\omega^{k_1}_{/c_1}\omega^{k_2}_{/c_2} + \omega^i_{/c_1c_2}\right\}\psi^a_{/i} \tag{4.13}$$

which is recognized as the transformation law for the Christoffel symbols of an affine connection $\overset{\alpha}{\nabla}$, called the *α-connection corresponding to g.*

From (4.11) and (4.12) it is seen that the costrings $\overset{\alpha}{\not g}_1$ and $\overset{\alpha}{\not g}^1$ are structurally symmetric, i.e. invariant under permutations of the structural indices; consequently the connection $\overset{\alpha}{\nabla}$ is torsion free.

For $g \in C^\infty(M \times M)$ let $\bar{g}$ be defined as

$$\bar{g}(\omega, \tilde{\omega}) = g(\omega, \tilde{\omega}) - g(\tilde{\omega}, \tilde{\omega}). \tag{4.14}$$

Clearly, one has that $\bar{g}$ is a yoke if and only if g is a yoke and in that case we refer to $\bar{g}$ as a *normalization* of g. If $g = \bar{g}$ or, equivalently, if g is identically equal to 0 on diag $M \times M$ we call g a *normalized yoke*. Furthermore, if g is a yoke one finds, using (4.1), that

$$\bar{g}_{;m_1}(\omega, \tilde{\omega}) = g_{;m_1}(\omega, \tilde{\omega}) - g_{m_1;}(\tilde{\omega}, \tilde{\omega}) - g_{;m_1}(\tilde{\omega}, \tilde{\omega}) = g_{;m_1}(\omega, \tilde{\omega}) - g_{;m_1}(\tilde{\omega}, \tilde{\omega})$$

and

$$\bar{g}_{;m_1 m_2}(\omega, \tilde{\omega}) = g_{;m_1 m_2}(\omega, \tilde{\omega}) - g_{m_2;m_1}(\tilde{\omega}, \tilde{\omega}) - g_{;m_1 m_2}(\tilde{\omega}, \tilde{\omega}).$$

Consequently, it follows from (4.5) and (4.14) that ${}^*g \in C(M \times M)$ given by

$${}^*g(\omega, \tilde{\omega}) = \bar{g}(\tilde{\omega}, \omega)$$

is a normalized yoke such that for $t = 0, 1, 2, \ldots$ and $u = 1, 2, \ldots$ one has

$${}^*\not g_{K_t;M_u} = \bar{\not g}_{M_u;K_t} = \not g_{M_u;K_t}. \tag{4.15}$$

Applying the formulas (4.7)–(4.15) we obtain that

$${}^*\overset{\alpha}{\not g}_1 = \overset{-\alpha}{\not g}_1, \tag{4.16}$$

$${}^*\overset{\alpha}{\not g}{}^1 = \overset{-\alpha}{\not g}{}^1 \tag{4.17}$$

and

$${}^*\overset{-1}{\not g}_0 = \overset{1}{\not g}_0. \tag{4.18}$$

As the last general result which will be mentioned here we have that a yoke on M induces a yoke on any (regular) submanifold N of M (for details, cf. Blæsild (1988)).

In Examples 2 and 3 below we consider two yokes of particular statistical interest. In these examples we let $p(x; \omega)$ denote the model function with respect to a dominating measure μ of a statistical model with sample space $\mathcal{X}$ and parameter space Ω. We assume that the model function is positive and we denote the log likelihood function by l, i.e.

$$l = l(\omega) = l(\omega; x) = \log p(x; \omega). \tag{4.19}$$

Note that $\{l_{K_t}\colon t = 1, 2, \ldots\}$ is a costring of degree (0, 0) and of infinite length which is referred to as the *string of log likelihood derivatives*.

Example 2. Let I denote the Kullback-Leibler information function

$$I(\tilde{\omega}, \omega) = E_{\tilde{\omega}}\{l(\tilde{\omega}) - l(\omega)\}, \tag{4.20}$$

where $E_{\tilde{\omega}}$ indicates the mean value under the probability measure corresponding to $\tilde{\omega}$. Assuming that the order of differentiation and integration may be interchanged and, furthermore, that the *expected* (or Fisher) *information* matrix, i.e.

$$i(\omega) = \{E_\omega(-l_{k_1 k_2})\},$$

is positive definite, it is easily seen that

$$g(\omega, \tilde{\omega}) = -I(\tilde{\omega}, \omega) = E_{\tilde{\omega}}\{l(\omega) - l(\tilde{\omega})\} \tag{4.21}$$

is a normalized yoke, which is called the *expected yoke*.

As proved in Blæsild (1988), one has for $t = 1, 2, \ldots$ and $u = 0, 1, 2, \ldots$ that

$$g_{K_t;M_u}(\omega, \tilde{\omega}) = \sum_{v=1}^{u} E_{\tilde{\omega}}\{l_{K_t}(\omega) l_{M_u/v}(\tilde{\omega})\} \tag{4.22}$$

and for $u = 1, 2, \ldots$ that

$$g_{;M_u}(\omega, \tilde{\omega}) = \sum_{v=1}^{u-1}(u - v)E_{\tilde{\omega}}\{l_{M_u/v}(\tilde{\omega})\} + \sum_{v=1}^{u} E_{\tilde{\omega}}\{(l(\omega) - l(\tilde{\omega}))l_{M_u/v}(\tilde{\omega})\}. \tag{4.23}$$

Notice that the quantities $g_{K_t;}(\omega, \tilde{\omega})$ are the moments evaluated at the parameter value $\tilde{\omega}$ of the log likelihood derivatives $l_{K_t}(\omega)$. These quantities are often referred to as the non-null moments of the log likelihood derivatives and they are valuable when assessing the robustness of test statistics depending on these derivatives (cf. McCullagh (1987)).

Replacement of $\tilde{\omega}$ by ω in (4.22) yields that the elements of the double string $\not{g}_{;}$ are given by

$$\not{g}_{K_t;M_u} = \nu_{K_t;M_u} = \sum_{v=1}^{u} \nu_{K_t,M_u/v} = \sum_{v=1}^{u} \sum_{M_u/v} \nu_{K_t,M_{u1},\ldots,M_{uv}}, \tag{4.24}$$

where ν denotes the *mixed moments of log likelihood derivatives*, i.e.

$$\nu_{K_t,M_{u1},\ldots,M_{uv}}(\omega) = E_\omega\{l_{K_t}(\omega) l_{M_{u1}}(\omega) \cdots l_{M_{uv}}(\omega)\}. \tag{4.25}$$

The double string $\nu_;$ was introduced in Blæsild (1987*b*). Letting λ denote the *joint cumulants of log likelihood derivatives*, i.e.

$$\lambda_{K_t,M_{u1},\ldots,M_{uv}}(\omega) = C_\omega\{l_{K_t}(\omega), l_{M_{u1}}(\omega), \ldots, l_{M_{uv}}(\omega)\}$$

one has the rather surprising result (cf. Blæsild (1987*b*)) that

$$\nu_{K_t;M_u} = \lambda_{K_t;M_u} = \sum_{v=1}^{u} \lambda_{K_t,M_u/v} = \sum_{v=1}^{u} \sum_{M_u/v} \lambda_{K_t,M_{u1},\ldots,M_{uv}}. \tag{4.26}$$

Consequently, the double strings $\not{g}_{;}$, $\nu_{;}$ and $\lambda_{;}$ are identical. Furthermore, one has that

$$\not{g}_{K_t;} = E_\omega\{l_{K_t}(\omega)\} = \nu_{K_t;} = \lambda_{K_t;} \tag{4.27}$$

and from (4.23) it follows that

$$\not{g}_{;M_u} = \sum_{v=1}^{u-1}(u-v)\nu_{M_u/v}. \tag{4.28}$$

Using (4.27), (4.4) and (4.24) one has the identity

$$\begin{aligned} \nu_{M_u} = \not{g}_{M_u;} &= -\sum_{M_u/2} \not{g}_{M_{u1};M_{u2}} \\ &= -\sum_{M_u/2} \sum_{\sigma=1}^{|M_{u2}|} \sum_{M_{u2}/\sigma} \nu_{M_{u1},M_{u2,1},\ldots,M_{u2,\sigma}} = -\sum_{v=2}^{u} \nu_{M_u/v} \end{aligned}$$

which gives the following alternative expression for $\not{g}_{;M_u}$

$$\not{g}_{;M_u} = -\sum_{v=2}^{u}(v-1)\nu_{M_u/v}. \tag{4.29}$$

The elements of the costrings $\overset{1}{\not{g}}{}^1$ and $\overset{-1}{\not{g}}{}^1$, expressed in terms of the mixed moments ν are, respectively,

$$\overset{1}{\not{g}}{}^i_{K_t} = \nu^{i;j}\nu_{K_t,j}, \tag{4.30}$$

$$\overset{-1}{\not{g}}{}^i_{K_t} = \nu^{i;j}\sum_{\tau=1}^{t}\nu_{j,K_t/\tau} \tag{4.31}$$

and it follows that the α-connections corresponding to g are the (*expected*) *α-connections* introduced by Chentsov and Amari (see Amari (1985, 1987)).

We conclude this example by showing that the geometries introduced by Eguchi (1983, 1985) by means of contrast functions are examples of geometries derived from yokes. Following Eguchi we say that for an arbitrary model parameterized by Ω, the function

$$\begin{aligned} \rho:\ & \Omega \times \Omega \to [0,\ \infty) \\ & (\omega,\ \tilde{\omega}) \to \rho(\omega,\ \tilde{\omega}) \end{aligned}$$

is a *contrast function* provided $\rho(\omega, \tilde{\omega}) = 0$ if and only if $\omega = \tilde{\omega}$. As noted by Eguchi (1983) any strictly convex function $w\colon R_+ \to R$ with $w(1) = 0$ generates, due to Jensen's inequality, a contrast function ρ_w given by

$$\rho_w(\omega, \tilde{\omega}) = E_\omega\{w(p(x;\, \tilde{\omega})/p(x;\, \omega))\}.$$

For each such function w we set

$$z = -w \circ \exp \tag{4.32}$$

and define the function ${}^z g\colon M \times M \to R$ by

$$\begin{aligned} &{}^z g : \Omega \times \Omega \to R \\ &(\omega, \tilde{\omega}) \to -\rho_w(\tilde{\omega}, \omega) = E_{\tilde{\omega}}\{z(l(\omega) - l(\tilde{\omega}))\}. \end{aligned} \tag{4.33}$$

Provided ${}^z g \in C^\infty(M \times M)$ one has, as shown in Blæsild (1988), that ${}^z g$ is a normalized yoke for which

$${}^z\not{g}_{K_t;} = \sum_{\tau=1}^{t} z^{(\tau)}(0)\nu_{K_t/\tau}, \tag{4.34}$$

$${}^z\not{g}_{;M_u} = \sum_{v=1}^{u}\sum_{\sigma=1}^{v}(-1)^\sigma \binom{v}{\sigma} z^{(\sigma)}(0)\nu_{M_u/v} \tag{4.35}$$

and

$${}^z\not{g}_{K_t;M_u} = \sum_{\tau=1}^{t}\sum_{v=1}^{u}\sum_{\sigma=0}^{v}(-1)^\sigma \binom{v}{\sigma} z^{(\tau+\sigma)}(0)\nu_{K_t/\tau, M_u/v} \tag{4.36}$$

for $t = 1, 2, \dots$ and $u = 1, 2, \dots$. Here $z^{(\tau)}(0)$ denotes the τ-fold derivative of z evaluated at 0.

Example 3. Let $\hat{\omega}$ denote the maximum likelihood estimator of ω. Assuming the existence of an auxiliary statistic a such that the transformation $x \to (\hat{\omega}, a)$ is one-to-one, the log likelihood function may be written as

$$l = l(\omega;\, \hat{\omega}, a). \tag{4.37}$$

In particular applications the statistic a is often ancillary, either exact or approximately, and in such cases inference on ω may be drawn by considering the *conditional normalized log likelihood function given* a, i.e.

$$g(\omega, \tilde{\omega}) = \bar{l}(\omega;\, \tilde{\omega}) = l(\omega;\, \tilde{\omega}, a) - l(\tilde{\omega};\, \tilde{\omega}, a). \tag{4.38}$$

It is easily seen that g is a normalized yoke provided that the observed information matrix evaluated at $\hat{\omega} = \omega$, i.e.

$$j(\omega) = \{-\not{l}_{k_1 k_2}(\omega)\}$$

is a positive definite matrix. This yoke, which is seen to depend on the value of the auxiliary statistic, is called the *observed yoke.*

Note that the analogy between the expected and observed yoke given by (4.21) and (4.38), respectively.

Writing

$$l(\omega,\, \tilde{\omega}) = l(\omega;\ \tilde{\omega},\, a)$$

it is easily seen that for u, $t = 1, 2, \ldots$ one has

$$\not{g}_{K_t;} = \not{l}_{K_t;}, \tag{4.39}$$

$$\not{g}_{;M_u} = -\sum_{M_u/2} \not{l}_{M_{u2};M_{u1}} \tag{4.40}$$

and

$$\not{g}_{K_t;M_u} = \not{l}_{K_t;M_u}. \tag{4.41}$$

The double string $\not{l}_{;}$, which was introduced in Barndorff-Nielsen (1986*b*), is called the *string of mixed log model derivatives.* The costrings of tensorial degree (1, 0) corresponding to $\alpha = 1$ and $\alpha = -1$ are given by, respectively,

$$\overset{1}{\not{g}}{}^{i}_{K_t} = \overset{1}{\not{l}}{}^{i}_{K_t} = \not{l}_{K_t;j}\not{l}^{i;j} \tag{4.42}$$

and

$$\overset{-1}{\not{g}}{}^{i}_{K_t} = \overset{-1}{\not{l}}{}^{i}_{K_t} = \not{l}_{j;K_t}\not{l}^{i;j}. \tag{4.43}$$

The corresponding α-connections were referred to as *the observed α-connections* in Barndorff-Nielsen (1986*a*, 1987*b*).

Finally, letting w: $R_+ \to R$ denote a strictly convex function with $w(1) = 0$ and setting $z = -w \circ \exp$ one finds that ${}^{z}g$ given by

$${}^{z}g(\omega,\, \tilde{\omega}) = z(l(\omega,\, \tilde{\omega}) - l(\tilde{\omega},\, \tilde{\omega}))$$

is a normalized yoke provided ${}^{z}g \in C^{\infty}(M \times M)$. It is proved in Blæsild (1988) that for t, $u = 1, 2, \ldots$ and with a notation similar to that in (2.2) one has

$${}^{z}\not{g}_{K_t;} = \sum_{\tau=1}^{t} z^{(\tau)}(0)\not{g}_{K_t/\tau;}, \tag{4.44}$$

$${}^{z}\not{g}_{;M_u} = \sum_{v=1}^{u} z^{(v)}(0)\not{g}_{;M_u/v}, \tag{4.45}$$

$${}^{z}\not{g}_{K_t;m_1} = \sum_{\tau=1}^{t} z^{(\tau)}(0) \sum_{K_t/\tau} \sum_{\mu=1}^{\tau} \not{g}_{K_{t1};} \cdots \not{g}_{K_{t\mu};m_1} \cdots \not{g}_{K_{t\tau};}, \tag{4.46}$$

and

$${}^{z}\not{g}_{k_1;M_u} = \sum_{v=1}^{u} z^{(v)}(0) \sum_{M_u/v} \sum_{\mu=1}^{v} \not{g}_{;M_{u1}} \cdots \not{g}_{k_1;M_{u\mu}} \cdots \not{g}_{;M_{uv}}. \tag{4.47}$$

Example 1. (continued) Since $E_{\tilde{\omega}} t_i = \kappa_{/i}(\tilde{\omega})$, it follows from (4.21) that (4.3) is the expected yoke for a regular exponential family.

The probability density function of the minimal sufficient statistic $t = (t_1, \ldots, t_d)$ is of the form

$$p(t;\ \omega) = \exp\{\omega^i t_i - \kappa(\omega)\} \tag{4.48}$$

and the equations $\kappa_{/i}(\hat{\omega}) = t_i$ establish a one-to-one correspondence between $\hat{\omega}$ and t. Consequently, no auxiliary statistic a is needed in this situation and from (4.19) and (4.48) one finds that the observed yoke is also given by (4.3).

From (4.3) it is easily seen that for $t = 2, 3, \ldots$ and $u = 1, 2, \ldots$ one has

$$\not{g}_{K_t;} = -\kappa_{/K_t}, \tag{4.49}$$

$$\not{g}_{k_1;M_u} = \kappa_{/k_1 \cup M_u}, \tag{4.50}$$

$$\not{g}_{K_t;M_u} = 0, \tag{4.51}$$

$$\not{g}_{;M_u} = -(u-1)\kappa_{/M_u}. \tag{4.52}$$

5. Tensors derived from yokes

Inspired by an idea in McCullagh and Cox (1986), Barndorff-Nielsen (1986*b*) introduced the concept of *intertwining* which has been further developed in Barndorff-Nielsen and Blæsild (1987*a*, 1987*b*). Given a connection string $\bar{\Gamma}$ with $\Gamma^i_{k_1} = \delta^i_{k_1}$ this concept establishes a one-to-one correspondence between a derivative string and a sequence of tensors. More specifically, if, for instance, $\bar{H}$ is a double derivative costring of degree (0, 0) and length (∞, ∞), it follows from Barndorff-Nielsen and Blæsild (1987*a*) that there exists a sequence of tensors $\bar{T} = \{T_{I_\tau J_\upsilon}\colon \tau, \upsilon = 1, 2, \ldots\}$, where $T_{I_\tau J_\upsilon}$ is a covariant tensor of degree $\tau + \upsilon$, such that

$$H_{K_t M_u} = \sum_{\tau=1}^{t} \sum_{\upsilon=1}^{u} T_{I_\tau J_\upsilon} \Gamma^{I_\tau}_{K_t} \Gamma^{J_\upsilon}_{M_u} \tag{5.1}$$

or equivalently

$$T_{I_\tau J_\upsilon} = \sum_{t=1}^{\tau} \sum_{u=1}^{\upsilon} H_{K_t M_u} G^{K_t}_{I_\tau} G^{M_u}_{J_\upsilon}. \tag{5.2}$$

In (5.1) and (5.2) the Γ- and G-quantities refer to, respectively, the co- and contrastrings generated by the connection string $\bar{\Gamma}$ (cf. Subsection 2.2). Furthermore, the relations (5.1) and (5.2) specialize in an obvious way to elements of $\mathcal{S}^{00}_{0\infty}$. The elements of the sequence $\bar{T}$ are referred to as the *tensorial components* of $\bar{H}$ with respect to $\bar{\Gamma}$ and the formulas (5.1) and (5.2) as the *intertwining* formulas.

Letting $\not{g}_{;}$ and $\overset{1}{\not{g}}{}^1$ denote, respectively, the double string and the 1-connection string corresponding to a yoke g we now derive and give some properties of the (covariant) tensorial components, which will be denoted by $\mathcal{T}_{;}$, of $\mathscr{g}_{;}$ with respect to $\overset{1}{\mathscr{g}}{}^{1}$. At the end of this section, we comment on the role of these tensors in statistical theory.

The tensorial components $\mathcal{T}_{;}$ of $\mathscr{g}_{;}$ with respect to $\overset{1}{\mathscr{g}}{}^{1}$ may be calculated from (4.12) with $\alpha = 1$ and (5.2) and the tensors up to degree four are:

$$
\begin{aligned}
&\mathcal{T}_{i_1;j_1} = \mathscr{g}_{i_1;j_1}, \quad \mathcal{T}_{i_1i_2;j_1} = 0, \quad \mathcal{T}_{i_1;j_1j_2} = \mathscr{g}_{i_1;j_1j_2} - \mathscr{g}_{j_1j_2;i_1}, \\
(5.3)\qquad &\mathcal{T}_{i_1i_2i_3;j_1} = 0, \quad \mathcal{T}_{i_1i_2;j_1j_2} = \mathscr{g}_{i_1i_2;j_1j_2} - \mathscr{g}_{i_1i_2;k}\mathscr{g}_{k';j_1j_2}\mathscr{g}^{k;k'}, \\
&\mathcal{T}_{i_1;j_1j_2j_3} = \mathscr{g}_{i_1;j_1j_2j_3} - \mathscr{g}_{j_1j_2j_3;i_1} - (\mathscr{g}_{i_1;j_1k} - \mathscr{g}_{j_1k;i_1})\mathscr{g}_{j_2j_3;k'}\mathscr{g}^{k;k'}[3]
\end{aligned}
$$

where [ ] refers to the number of similar terms obtained by suitable permutations of the indices.

In general one has that

$$
(5.4)\qquad \mathcal{T}_{I_\tau;j} = 0 \quad \text{if} \quad \tau \geq 2.
$$

To see this, we denote by $\overset{1}{\mathscr{G}}$ the contrastrings generated by $\overset{1}{\mathscr{g}}{}^{1}$. From (5.2), (2.9) and (4.12) it follows that

$$
\begin{aligned}
\mathcal{T}_{I_\tau;j} &= \sum_{t=1}^{\tau} \mathscr{g}_{K_t;m} \overset{1}{\mathscr{G}}{}^{K_t}_{I_\tau} \overset{1}{\mathscr{G}}{}^{m}_{j} = \sum_{t=1}^{\tau} \mathscr{g}_{K_t;m} \overset{1}{\mathscr{G}}{}^{K_t}_{I_\tau} \delta^m_j \\
&= \sum_{t=1}^{\tau} \mathscr{g}_{K_t;j} \overset{1}{\mathscr{G}}{}^{K_t}_{I_\tau} = \mathscr{g}_{j;j'} \sum_{t=1}^{\tau} \overset{1}{\mathscr{g}}{}^{j'}_{K_t} \overset{1}{\mathscr{G}}{}^{K_t}_{I_\tau} = \mathscr{g}_{j;j'} \delta^{j'}_{I_\tau} = 0.
\end{aligned}
$$

The tensorial components of $\overset{1}{\mathscr{g}}_0 = \{\mathscr{g}_{K_t;}\colon t = 1, 2, \ldots\} \in \mathcal{S}^{00}_{0\infty}$ will be denoted by $\mathcal{T}_{I_\tau;}$, $\tau = 1, 2, \ldots$. These tensors may be described in terms of the tensors $\mathcal{T}_{;}$ since we have the following formula analogous to (4.4)

$$
(5.5)\qquad \mathcal{T}_{I_\tau;} + \sum_{I_\tau/2} \mathcal{T}_{I_{\tau 1};I_{\tau 2}} = 0.
$$

Conversely, it is not possible to express the tensors $\mathcal{T}_{I_\tau;J_v}$ in terms of the tensors $\mathcal{T}_{I_\tau;}$.

In the light of these facts and in line with the concept of yokes, the tensors $\mathcal{T}_{I_\tau;J_v}$ may be considered as more fundamental than the tensors $\mathcal{T}_{I_\tau;}$.

The identity (5.5) may be proved by an induction argument, using the formulas (2.9), (5.1) and (5.2) and brute force. However, here we give a proof of (5.5) which may add to the understanding of the concept of tensorial components.

For an arbitrary *structural symmetric* connection string $\bar{\Gamma}$ we consider, in line with Murray and Rice (1987) and Mora (1988), the *extended normal coordinates at p with respect to* $\bar{\Gamma}$. These coordinates will be denoted by γ and we use the letters

$x, y, z, \ldots$ for arbitrary indices of γ. The coordinates γ, defined in a neighbourhood of p, satisfy the conditions

$$\gamma^x(p) = 0, \qquad \Gamma^x_{Z_t}(p) = 0 \quad \text{for} \quad t = 2, 3, \ldots \tag{5.6}$$

and so the concept of extended normal coordinates is indeed a generalization of the concept of normal coordinates of a torsion free connection. Formulas (2.8), (2.9) and (5.6) imply that

$$\Gamma^{X_r}_{Z_t}(p) = 0 \quad \text{for} \quad r > 2$$

and

$$G^{W_u}_{Y_s}(p) = \delta^{W_u}_{Y_s}. \tag{5.7}$$

For the double string $\mathscr{g}_{;}$, one obtains, using (5.2) and (5.7), that

$$\mathscr{T}_{X_\tau;Y_v}(p) = \sum_{t=1}^{\tau}\sum_{u=1}^{v} \mathscr{g}_{Z_t;W_u}(p)\overset{1}{\mathscr{G}}{}^{Z_t}_{X_\tau}(p)\overset{1}{\mathscr{G}}{}^{W_u}_{Y_v}(p) = \mathscr{g}_{X_\tau;Y_v}(p) \tag{5.8}$$

i.e. the tensorial components of $\mathscr{g}_{;}$ at p are simply the elements of $\mathscr{g}_{;}$ expressed in the system of extended normal coordinates at p. Similarly, one has that

$$\mathscr{T}_{X_\tau;}(p) = \mathscr{g}_{X_\tau;}(p) \tag{5.9}$$

and consequently, by (4.4), (5.8) and (5.9), that

$$\mathscr{T}_{X_\tau;}(p) + \sum_{X_\tau/2} \mathscr{T}_{X_{\tau 1};X_{\tau 2}}(p) = 0. \tag{5.10}$$

Since all the terms in (5.10) are covariant tensors of degree τ the proof of (5.5) is complete.

From (5.3)–(5.5), one finds the following expressions for the tensorial components up to degree four of $\overset{1}{\mathscr{g}}_0$ with respect to $\overset{1}{\mathscr{g}}{}^1$:

$$\begin{aligned}
\mathscr{T}_{i_1i_2;} &= -\mathscr{T}_{i_1;i_2} = -\mathscr{g}_{i_1;i_2}, \\
\mathscr{T}_{i_1i_2i_3;} &= -\mathscr{T}_{i_1;i_2i_3} = \mathscr{g}_{i_2i_3;i_1} - \mathscr{g}_{i_1;i_2i_3}, \\
\mathscr{T}_{i_1i_2i_3i_4;} &= -\mathscr{T}_{i_1;i_2i_3i_4} - \mathscr{T}_{i_1i_2;i_3i_4}[3] \\
&= \mathscr{g}_{i_2i_3i_4;i_1} - \mathscr{g}_{i_1;i_2i_3i_4} - \mathscr{g}_{i_1i_2;i_3i_4}[3] \\
&\quad + \mathscr{g}_{i_1i_2;k}\mathscr{g}_{k';i_3i_4}\mathscr{g}^{k;k'}[6] - \mathscr{g}_{i_1i_2;k}\mathscr{g}_{i_3i_4;k'}g^{k;k'}[3].
\end{aligned} \tag{5.11}$$

Since the strings $\overset{1}{\mathscr{g}}_0$ and $\overset{1}{\mathscr{g}}{}^1$ are both structurally symmetric one has generally that the tensors $\mathscr{T}_{I_\tau;}$ are symmetric. In particular, one has that the tensors

$$\mathscr{T}_{i_1;i_2} = -\mathscr{T}_{i_1i_2;} = \mathscr{g}_{i_1;i_2} \tag{5.12}$$

and

(5.13) $$\mathcal{T}_{i_1;i_2i_3} = -\mathcal{T}_{i_1i_2i_3;} = \not{g}_{i_1;i_2i_3} - \not{g}_{i_2i_3;i_1}$$

are symmetric.

Example 1. (continued) For a regular exponential family one obtains, using (4.49)–(4.51), (5.3) and (5.11), the following expressions for the non-vanishing tensors $\mathcal{T}_;$ up to degree four expressed in terms of the canonical parameter ω:

$$\begin{aligned}\mathcal{T}_{i_1;i_2} &= -\mathcal{T}_{i_1i_2;} = \kappa_{/i_1i_2},\\ \mathcal{T}_{i_1;i_2i_3} &= -\mathcal{T}_{i_1i_2i_3;} = \kappa_{/i_1i_2i_3},\\ \mathcal{T}_{i_1;i_2i_3i_4} &= -\mathcal{T}_{i_1i_2i_3i_4;} = \kappa_{/i_1i_2i_3i_4}.\end{aligned}$$

It is well-known that the difference between the lower Christoffel symbols for two connections is a (0, 3) tensor. For the α-connections corresponding to a yoke g, one has according to (4.12) and (5.3) that

(5.14) $$\begin{aligned}2\left(\overset{0}{\not{g}}_{jk_1k_2} - \overset{\alpha}{\not{g}}_{jk_1k_2}\right) &= \alpha\left(\overset{-1}{\not{g}}_{jk_1k_2} - \overset{1}{\not{g}}_{jk_1k_2}\right)\\ &= \alpha(\not{g}_{j;k_1k_2} - \not{g}_{k_1k_2;j})\\ &= \alpha\mathcal{T}_{j;k_1k_2}\end{aligned}$$

In particular, for the expected yoke considered in Example 2 formula (4.24) implies that

$$\mathcal{T}_{j;k_1k_2} = \nu_{j,k_1k_2} + \nu_{j,k_1,k_2} - \nu_{k_1k_2,j} = \nu_{j,k_1,k_2}.$$

In statistical theory this tensor is known as the *skewness tensor* (cf. Lauritzen (1987)).

Most tensors in classical differential geometry, for instance the curvature tensor, are expressed in terms of the canonical connection string corresponding to the connection considered. We now show that the connection strings $\overset{\alpha}{\not{g}}{}^1$ corresponding to a yoke g are not in general canonical. Letting ${}_c\overset{\alpha}{\not{g}}{}^1$ denote the canonical connection string corresponding to $\overset{\alpha}{\nabla}$, the α-connection induced by g, it follows from Subsection 2.2 that

$$\overset{\alpha}{\nabla}_{\partial_{k_t}}\left(\cdots\left(\overset{\alpha}{\nabla}_{\partial_{k_2}}\partial_{k_1}\right)\cdots\right) = {}_c\overset{\alpha}{\not{g}}{}^i_{K_t}\partial_i.$$

In particular, one has that

(5.15) $${}_c\overset{\alpha}{\not{g}}{}^i_{k_1k_2k_3} = \partial_{k_3}\overset{\alpha}{\not{g}}{}^i_{k_1k_2} + \overset{\alpha}{\not{g}}{}^{i'}_{k_1k_2}\overset{\alpha}{\not{g}}{}^i_{i'k_3}.$$

It is proved in Blæsild (1988) that

(5.16) $$\begin{aligned}{}_c\overset{\alpha}{\not{g}}{}^i_{k_1k_2k_3} - \overset{\alpha}{\not{g}}{}^i_{k_1k_2k_3} = \Bigg(&\frac{1+\alpha}{2}\mathcal{T}_{k_1k_2;k_3j} + \frac{1-\alpha}{2}\mathcal{T}_{k_3j;k_1k_2}\\ &- \frac{1-\alpha^2}{4}\mathcal{T}_{i';k_1k_2}\mathcal{T}_{j';jk_3}\not{g}^{i';j'}\Bigg)\not{g}^{i;j}\end{aligned}$$

which shows the assertion above. Note that for $\alpha = 1$ and $\alpha = -1$ formula (5.16) turns into

$$ {}_c\overset{1}{\not{g}}{}^i_{k_1k_2k_3} - \overset{1}{\not{g}}{}^i_{k_1k_2k_3} = \mathscr{T}_{k_1k_2;k_3j}\not{g}^{i;j} \tag{5.17}$$

and

$$ {}_c\overset{-1}{\not{g}}{}^i_{k_1k_2k_3} - \overset{1}{\not{g}}{}^i_{k_1k_2k_3} = \mathscr{T}_{k_3j;k_1k_2}\not{g}^{i;j}. \tag{5.18}$$

The *curvature tensor* $\overset{\alpha}{R}$ of $\overset{\alpha}{\nabla}$ may be expressed in terms of the canonical connection string as

$$\overset{\alpha}{R}_{k_3k_2k_1j} = \left({}_c\overset{\alpha}{\not{g}}{}^i_{k_1k_2k_3} - {}_c\overset{\alpha}{\not{g}}{}^i_{k_1k_3k_2} \right) \not{g}_{i;j},$$

so, using (5.16), we obtain

$$\begin{aligned} \overset{\alpha}{R}_{k_3k_2k_1j} =& \frac{1+\alpha}{2}(\mathscr{T}_{k_1k_2;k_3j} - \mathscr{T}_{k_1k_3;k_2j}) + \frac{1-\alpha}{2}(\mathscr{T}_{k_3j;k_1k_2} - \mathscr{T}_{k_2j;k_1k_3}) \\ &- \frac{1-\alpha^2}{4}(\mathscr{T}_{i';k_1k_2}\mathscr{T}_{j';k_3j} - \mathscr{T}_{i';k_1k_3}\mathscr{T}_{j';k_2j})\not{g}^{i';j'}, \end{aligned} \tag{5.19}$$

which for $\alpha = 1$ and $\alpha = -1$ reduces to, respectively,

$$\overset{1}{R}_{k_3k_2k_1j} = \mathscr{T}_{k_1k_2;k_3j} - \mathscr{T}_{k_1k_3;k_2j} \tag{5.20}$$

and

$$\overset{-1}{R}_{k_3k_2k_1j} = \mathscr{T}_{k_3j;k_1k_2} - \mathscr{T}_{k_2j;k_1k_3}. \tag{5.21}$$

In statistical theory the tensors $\mathscr{T}_;$ up to degree four corresponding to the expected and to the observed yoke, respectively, appear again and again in discussions of invariant asymptotic theory. The tensorial components of $\overset{1}{\not{g}}_0$ corresponding to the expected yoke, i.e. the mixed moments of the so-called Möbius derivatives of the log likelihood function, were introduced in McCullagh and Cox (1986) and used to express the Bartlett adjustment of the log likelihood ratio statistic as a sum of invariant terms. A similar expression in the case of observed geometry was given by Barndorff-Nielsen (1986*b*). The formal analogy between these expressions was established in Blæsild (1987*b*) using the analogy between what is now known as the double strings $\not{g}_;$ corresponding to the expected and observed yokes, respectively. More specifically, if w and w' denote, respectively, the original and the Bartlett adjusted log likelihood ratio statistic for testing a point hypothesis about the parameter ω, it was shown in Blæsild (1987*b*) that both for the expected and observed geometries one has

$$w' = w(1 + B/d),$$

where d is the dimension of ω and where

$$\begin{aligned} 12B =& \{3\mathscr{T}_{ijkm;} + 12\mathscr{T}_{ik;jm}\}\mathscr{T}^{i;j}\mathscr{T}^{k;m} \\ &+ \{3\mathscr{T}_{i;jk}\mathscr{T}_{m;np} + 2\mathscr{T}_{i;kn}\mathscr{T}_{j;mp}\}\mathscr{T}^{i;j}\mathscr{T}^{k;m}\mathscr{T}^{n;p}. \end{aligned} \tag{5.22}$$

The tensors $\mathcal{T}_{;}$ corresponding to the observed yoke also occur in the asymptotic expansions of the conditional distributions of, respectively, the maximum likelihood estimator and the score vector given an (approximately) ancillary statistic derived in Mora (1988) (see also Barndorff-Nielsen (1989)). Finally, the tensors $\mathcal{T}_{;}$ corresponding to the expected yoke appear in the asymptotic expansions for the distributions of the maximum likelihood estimator and the score vector considered in Amari and Kumon (1983) and Barndorff-Nielsen (1986*a*) (cf. also Amari (1985) and Mora (1988)).

One of the topics in the discussion in Dodson *et al.* (1987) is concerned with the need of supplementing the concept of a statistical manifold, introduced in Lauritzen (1987), in order to be able to handle asymptotic theory within a unified framework. The tensor $\mathcal{T}_{i_1i_2;j_1j_2}$ corresponding to the expected yoke was mentioned as a candidate. The formulas (5.19)–(5.22) support this proposal. Despite the formulas (5.20) and (5.21), it does not seem possible to express the tensor $\mathcal{T}_{i_1i_2;j_1j_2}$ in terms of the curvature tensors $\overset{\alpha}{R}$. The present author is not aware of any interpretation of the tensor $\mathcal{T}_{i_1i_2;j_1j_2}$ in terms of classical geometrical quantities. From (5.17) it follows that

$$\mathcal{T}_{k_1k_2;k_3j} = {}_c\overset{1}{\phi}_{jk_1k_2k_3} - \overset{1}{\phi}_{jk_1k_2k_3}$$

showing that the tensor is simply the difference between the third elements of the canonical string ${}_c\overset{1}{\phi}_1$ and of the string $\overset{1}{\phi}_1$, i.e. the tensor is easily expressed in terms of derivative strings.

As a final remark, note that the tensor $\mathcal{T}_{i_1i_2i_3i_4;}$ appears in (5.22) and that, according to (5.11), this tensor can not be expressed in terms of the tensors $\mathcal{T}_{i_1i_2;i_3i_4}$ and $\mathcal{T}_{i_1;i_2i_3}$ only. Consequently, it is not enough to supplement the concept of a statistical manifold with the tensor $\mathcal{T}_{i_1i_2;i_3i_4}$. The basic quantity to consider in relation to asymptotic theory seems to be the expected yoke itself from which the Fisher metric, the skewness tensor and the appropriate tensor of higher order may be derived. Finally, the demand on g being smooth on $M \times M$, i.e. infinitely often differentiable, may be reduced to g being continuously differentiable a suitable number of times; for instance in the case of Bartlett adjustments it suffices that g is four times continuously differentiable.

References

Amari, S. (1985). Differential geometric methods in statistics, *Lecture Notes in Statist.*, **28**, Springer, Heidelberg.

Amari, S. (1987). Differential geometrical theory of statistics—towards new developments, *Differential Geometry in Statistical Inference*, 19–94, IMS Lecture Notes-Monograph Series, Hayward, California.

Amari, S. and Kumon, M. (1983). Differential geometry of Edgeworth expansion in curved exponential family, *Ann. Inst. Statist. Math.*, **35**, 1–24.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families*, Wiley, Chichester.

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

Barndorff-Nielsen, O. E. (1986*b*). Strings, tensorial combinants, and Bartlett adjustments, *Proc. Roy. Soc. London Ser. A*, **406**, 127–137.

Barndorff-Nielsen, O. E. (1987*a*). On some differential geometric concepts and their relations to statistics, *Geometrization of Statistical Theory* (ed. C. T. J. Dodson), 53–90, ULDM Publications, Dept. Math., University of Lancaster, England.

Barndorff-Nielsen, O. E. (1987*b*). Differential and integral geometry in statistical inference, *Differential Geometry in Statistical Inference*, 95–162, IMS Lecture Notes-Monograph Series, Hayward, California.

Barndorff-Nielsen, O. E. (1988*a*). Differential geometry and statistics: Some mathematical aspects, *Indian J. Math.*, **29**, Ramanujan Centenary Volume.

Barndorff-Nielsen, O. E. (1988*b*). Parametric statistical models and likelihood, *Lecture Notes in Statist.*, **50**, Springer, Heidelberg.

Barndorff-Nielsen, O. E. (1989). In the discussion of Kass, R. E.: The geometry of asymptotic inference, *Statist. Sci.*, **4**, 188–234.

Barndorff-Nielsen, O. E. and Blæsild, P. (1987*a*). Strings: Mathematical theory and statistical examples, *Proc. Roy. Soc. London Ser. A*, **411**, 155–176.

Barndorff-Nielsen, O. E. and Blæsild, P. (1987*b*). Strings: Contravariant aspect, *Proc. Roy. Soc. London Ser. A*, **411**, 421–444.

Barndorff-Nielsen, O. E. and Blæsild, P. (1988). Coordinate-free definition of structurally symmetric derivative strings, *Adv. in Appl. Math.*, **9**, 1–6.

Barndorff-Nielsen, O. E., Blæsild, P. and Mora, M. (1988). Derivative strings and higher order differentiation, Memoir 11, Dept. Theor. Statist., Aarhus University, Aarhus, Denmark.

Blæsild, P. (1987*a*). Yokes: Elemental properties with statistical applications, *Geometrization of Statistical Theory* (ed. C. T. J. Dodson), 193–196, ULDM Publications, Dept. Math., University of Lancaster, Lancaster, England.

Blæsild, P. (1987*b*). Further analogies between expected and observed geometries of statistical models, Research Report 160, Dept. Theor. Statist., Aarhus University, Aarhus, Denmark.

Blæsild, P. (1988). Yokes and tensors derived from yokes, Research Report 173, Dept. Theor. Statist., Aarhus University, Aarhus, Denmark.

Dodson, C. T. J., Jupp, P. E., Kendall, W. S. and Lauritzen, S. L. (1987). A summary of points raised in the Discussion Sessions, *Geometrization of Statistical Theory* (ed. C. T. J. Dodson), 235–250, ULDM Publications, Dept. Math., University of Lancaster, Lancaster, England.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in curved exponential families, *Ann. Statist.*, **11**, 793–803.

Eguchi, S. (1985). A differential geometric approach to statistical inference on the basis of contrast functionals, *Hiroshima Math. J.*, **15**, 341–391.

Lauritzen, S. L. (1987). Statistical manifolds, *Differential Geometry in Statistical Inference*, 163–216, IMS Lecture Notes-Monograph Series, Hayward, California.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

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

Mora, M. (1988). Geometrical expansions for the distributions of the score vector and the maximum likelihood estimator, Research Report 172, Dept. Theor. Statist., Aarhus University, Aarhus, Denmark.

Murray, M. K. and Rice, J. W. (1987). On differential geometry in statistics, Research Report, School of Mathematical Sciences, The Flinders University of South Australia, Adelaide, Australia.