OLE E. BARNDORFF-NIELSEN
PETER E. JUPP

Statistics, yokes and symplectic geometry

Annales de la faculté des sciences de Toulouse 6e série, tome 6, nº 3 (1997), p. 389-427
<http://www.numdam.org/item?id=AFST_1997_6_6_3_389_0>

Annales de la Faculté des Sciences de Toulouse Vol. VI, n° 3, 1997

Statistics, Yokes and Symplectic Geometry(*)

OLE E. BARNDORFF-NIELSEN[1]
and
PETER E. JUPP[2]

RÉSUMÉ. — Nous établissons et étudions un lien entre jougs et formes symplectiques. Nous montrons que les jougs normalisés correspondent à certaines formes symplectiques. Nous présentons une méthode pour construire de nouveaux jougs à partir de jougs donnés. Celle-ci est motivée en partie par la dualité entre la formulation hamiltonienne et la formulation lagrangienne de la mécanique conservative. Nous nous proposons quelques variantes de cette construction.

MOTS-CLÉS : Application moment, hamiltonien, joug de vraisemblance espérée, joug de vraisemblance observée, lagrangien, sous-variété lagrangienne, tenseurs.

ABSTRACT. — A relationship between yokes and symplectic forms is established and explored. It is shown that normalised yokes correspond to certain symplectic forms. A method of obtaining new yokes from old is given, motivated partly by the duality between the Hamiltonian and Lagrangian formulations of conservative mechanics. Some variants of this construction are suggested.

KEY-WORDS : expected likelihood yoke, Hamiltonian, Lagrangian submanifold, momentum map, observed likelihood yoke, tensors.

AMS Classification : 58F05, 62E20

(*) Reçu le 10 juillet 1995

(1) Department of Mathematical Sciences, Aarhus University, Ny Munkegade, DK-8000 Aarhus C (Denmark)

(2) School of Mathematical and Computational Sciences, University of St Andrews, North Haugh, St Andrews KY16 9SS (United Kingdom)

1. Introduction

In the differential-geometric approach to statistical asymptotics a central concept is that of a yoke, key examples being the observed and expected likelihood yokes.

A yoke on a manifold M is a real-valued function on the "square" $M \times M$ of M, satisfying conditions (2.1) and (2.2) below (Barndorff-Nielsen [7]). Yokes which are restricted to be non-negative and are zero only on the diagonal are known as contrast functions. For uses of contrast functions in statistics see, e.g., Eguchi [16] and Skovgaard [24]. In applications of yokes to statistical asymptotics we are particularly interested in the values of the yoke near the diagonal Δ_M of $M \times M$, where $\Delta_M = \{(\omega, \omega) \mid \omega \in M\}$. A suitable neighbourhood of Δ_M in $M \times M$ can be regarded as the total space of the normal bundle of Δ_M in $M \times M$ and this normal bundle is isomorphic to the tangent bundle τM of M. Furthermore, a yoke on M determines a (possibly indefinite) Riemannian metric, which can be used to identify the tangent bundle τM with the cotangent bundle $\tau^* M$.

One of the main contexts in which cotangent bundles occur is in conservative mechanics (see e.g., Abraham and Marsden [1], and Marsden [19]), where they are known as the phase spaces. The geometrical concept is that of a symplectic structure. A survey of symplectic geometry and its applications is given by Arnol'd and Givental' [3].

The role of Hamiltonians in conservative mechanics and their relation to exponential families suggest that symplectic geometry may have a natural role to play in statistics. It is relevant here also to mention the work of Combet [13] which discusses certain connections between symplectic geometry and Laplace's method for exponential integrals.

In view of the results to be described below we think it fair to say that there exists a natural link, via the concept of yokes, between statistics and symplectic geometry. How useful this link may be seems difficult to assess at present. Other links have been discussed by T. Friedrich and Y. Nakamura. Friedrich [17] established some connections between expected (Fisher) information and symplectic structures. However, as indicated in Remark 3.4, his approach and results are quite different from those considered here. Nakamura ([22], [23]) has shown that certain parametric statistical models in which the parameter space M is an even-dimensional vector space (and so has the symplectic structure of the cotangent space of a vector space) give rise to completely integrable Hamiltonian systems on M. In contrast to these, the Hamiltonian systems considered here arise in the more general context of yokes on arbitrary manifolds M and are defined on some neighbourhood of Δ_M in $M \times M$.

Section 2 reviews the necessary background on yokes, symplectic structures and mechanics. In Section 3, we show that every yoke on a manifold M gives rise to a symplectic form, at least on some neighbourhood of the diagonal of $M \times M$, and in some important cases on all of $M \times M$. By passing to germs, i.e., by identifying yokes or symplectic forms which agree in some neighbourhood of the diagonal, we show that normalised yokes are almost equivalent to symplectic forms of a certain type. (However, from the *local* viewpoint there is nothing special about those symplectic forms which are given by yokes, since Darboux's Theorem shows that locally every symplectic form can be obtained from some yoke; see Example 3.1.) In Section 4, we consider some uses of the symplectic form given by a yoke.

There is an analogy between conservative mechanics and the geometry of yokes in that symplectic forms play an important part in both areas. In Section 5, this analogy is used to transfer the duality between the Hamiltonian and Lagrangian versions of conservative mechanics to a construction for obtaining from any normalised yoke g another yoke $\breve{g}$, called the Lagrangian of g, which has the same metric as g and is, in some sense, close to the dual yoke g^* of g.

Section 6 considers Lie group actions on the manifold and the associated momentum map, which takes values in the dual of the appropriate Lie algebra.

An outline version of some of the material presented in this paper is given in Barndorff-Nielsen and Jupp [11].

2. Preliminaries

2.1 Yokes

Let M be a smooth manifold. We shall sometimes use local coordinates $(\omega^1, \ldots, \omega^d)$ on M and, correspondingly, local coordinates $(\omega^1, \ldots, \omega^d; \omega'^1, \ldots, \omega'^d)$ on $M \times M$. Furthermore, we use the notations

$$g_{i;}(\omega, \omega') = \frac{\partial g(\omega, \omega')}{\partial \omega^i},$$
$$g_{i;j}(\omega, \omega') = \frac{\partial^2 g(\omega, \omega')}{\partial \omega^i \partial \omega'^j}, \quad \text{etc.}$$

For a function on $M \times M$, the corresponding gothic letter will indicate restriction of that function to the diagonal, so that, e.g.,

$$\mathfrak{g}_{i;j} = g_{i;j}(\omega, \omega).$$

In coordinate terms a yoke on M is defined as a smooth function $g : M \times M \to \mathbb{R}$ such that for every ω in M:

$$\text{(i)} \quad \mathfrak{g}_{i;}(\omega) = 0 \tag{2.1}$$

$$\text{(ii)} \quad \text{the matrix } [\mathfrak{g}_{i;j}(\omega)] \text{ is non-singular.} \tag{2.2}$$

The coordinate-free definition of a yoke is as follows. For a vector field X on M, define the vector fields $\overline{X}$ and $\overline{X}'$ on $M \times M$ by $\overline{X} = (X, 0)$ and $\overline{X}' = (0, X)$, i.e.,

$$Tp_1(\overline{X}) = X, \quad Tp_2(\overline{X}) = 0,$$
$$Tp_1(\overline{X}') = 0, \quad Tp_2(\overline{X}') = X,$$

where $p_k : M \times M \to M$ is the projection onto the k^{th} factor. Then, for vector fields X and Y on M, we define $g(X \mid Y) : M \to \mathbb{R}$ by

$$g(X \mid Y)(\omega) = \overline{X}\,\overline{Y}' g(\omega, \omega).$$

A yoke on M may now be characterised as a smooth function $g : M \times M \to \mathbb{R}$ such that:

(i) $\overline{X} g(\omega, \omega) = 0$ for all ω in M,

(ii) the $(0,2)$-tensor $(X, Y) \mapsto g(X \mid Y)$ is non-singular.

An alternative way of expressing (i) and (ii) is that on Δ_M:

(i) $\mathrm{d}_1 g = 0$,

(ii) $\mathrm{d}_1\, \mathrm{d}_2 g$ is non-singular,

where d_1 and d_2 denote exterior differentiation along the first and second factor in $M \times M$, respectively.

Example 2.1

The simplest example of a yoke is the function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$g(\omega, \omega') = -\frac{1}{2} \left\|\omega - \omega'\right\|^2 .$$

For all ω', the function $\omega \mapsto g(\omega, \omega')$ has a (unique) maximum at $\omega = \omega'$, so that (2.1) holds. Since $\left[\mathfrak{g}_{i;j}(\omega)\right]$ is the identity matrix, (2.2) holds. Some generalisations of this example are given in Examples 5.1-5.3.

Two of the most important geometric objects given by a yoke are a (possibly indefinite) Riemannian metric and a one-parameter family of torsion-zero affine connections. The *metric* is the $(0,2)$-tensor $(X,Y) \mapsto g(X \mid Y)$, given in coordinate form by the matrix $[\mathfrak{g}_{i;j}]$. For α in $\mathbb{R}$, the *α-connection* $\overset{\alpha}{\nabla}$ is defined by

$$\overset{\alpha}{\nabla} = \overset{0}{\nabla} - \frac{\alpha}{2} T ,$$

where $\overset{0}{\nabla}$ is the Levi-Civita connection of the metric and T is given by

$$g(T_X Y \mid Z)(\omega) = \overline{X}\,\overline{Y}'\overline{Z}'g(\omega,\omega) - \overline{Y}\,\overline{Z}\,\overline{X}'g(\omega,\omega) .$$

Note that T is a $(1,2)$-tensor. The lowered Christoffel symbols of $\overset{\alpha}{\nabla}$ at ω are given by

$$\overset{\alpha}{\Gamma}_{ijk} = \frac{1+\alpha}{2}\,\mathfrak{g}_{ij;k} + \frac{1-\alpha}{2}\,\mathfrak{g}_{k;ij} \tag{2.3}$$

and the $(0,3)$-tensor corresponding to T is the "skewness" tensor with elements

$$T_{rst} = \mathfrak{g}_{r;st} - \mathfrak{g}_{st;r} . \tag{2.4}$$

For applications to statistics of the metric and the α-connections of expected and observed likelihood yokes, see Amari [2], Barndorff-Nielsen ([6], [8]) and Murray and Rice ([21]).

A *normalised yoke* is a yoke satisfying the additional condition

$$g(\omega, \omega) = 0 . \tag{2.5}$$

Except where explicitly stated otherwise, we shall consider only normalised yokes. For any yoke g, the corresponding normalised yoke is the yoke $\overline{g}$ defined by

$$\overline{g}(\omega, \omega') = g(\omega, \omega') - g(\omega', \omega')$$

and the *dual yoke* is the yoke g^* defined by

$$g^*(\omega,\omega') = \overline{g}(\omega',\omega)\,.$$

Note that g^* is necessarily a normalised yoke.

In the statistical context the two important examples of normalised yokes are the expected likelihood yoke and the observed likelihood yoke. For a parametric statistical model with parameter space M, sample space $\mathcal{X}$ and log-likelihood function $\ell : M \times \mathcal{X} \to \mathbb{R}$ the expected likelihood yoke on M is the function g given by

$$g(\omega,\omega') = E_{\omega'}\left[\ell(\omega;x) - \ell(w';x)\right]\,.$$

Given an auxiliary statistic a such that the function $x \mapsto (\widehat{\omega}, a)$ is bijective, the observed likelihood yoke on M is defined as the function g given by

$$g(\omega,\omega') = \ell(\omega;\omega',a) - \ell(w';\omega',a)\,. \tag{2.6}$$

For more about expected and observed likelihood yokes see Barndorff-Nielsen ([7], [8]), Chapter 5 of Barndorff-Nielsen and Cox [9] and Blæsild [12].

Differentiation of (2.1) yields the important result

$$\mathfrak{g}_{ij;} + \mathfrak{g}_{i;j} = 0\,. \tag{2.7}$$

This shows that the matrix $[\mathfrak{g}_{i;j}]$ is symmetric, a property which is not obvious at first sight. Similarly, differentiation of (2.5) shows that for a normalised yoke g we have

$$\mathfrak{g}_{i;} = 0 = \mathfrak{g}_{;i} \tag{2.8}$$

and

$$\mathfrak{g}_{ij;} = -\mathfrak{g}_{i;j} = \mathfrak{g}_{;ij}\,. \tag{2.9}$$

Differentiation of (2.7) yields

$$\mathfrak{g}_{ijk;} + \mathfrak{g}_{ij;k} + \mathfrak{g}_{ik;j} + \mathfrak{g}_{i;jk} = 0\,. \tag{2.10}$$

Similarly, for a normalised yoke g, we have

$$\mathfrak{g}_{jk;i} + \mathfrak{g}_{k;ij} + \mathfrak{g}_{j;ik} + \mathfrak{g}_{;ijk} = 0\,. \tag{2.11}$$

Differentiation of (2.10) and (2.11) yields

$$\mathfrak{g}_{ijkl;} + \mathfrak{g}_{ijk;l} + \mathfrak{g}_{ikl;j} + \mathfrak{g}_{ilj;k} + \mathfrak{g}_{ij;kl} + \mathfrak{g}_{ik;lj} + \mathfrak{g}_{il;jk} + \mathfrak{g}_{i;jkl} = 0 \quad (2.12)$$

and

$$\mathfrak{g}_{ijk;l} + \mathfrak{g}_{ij;kl} + \mathfrak{g}_{jk;il} + \mathfrak{g}_{ki;jl} + \mathfrak{g}_{i;jkl} + \mathfrak{g}_{j;kil} + \mathfrak{g}_{k;ijl} + \mathfrak{g}_{;ijkl} = 0\,. \quad (2.13)$$

Formulae (2.7) and (2.10)-(2.13) are special cases of the general balance relations for yokes. See formula (5.91) of Barndorff-Nielsen and Cox [9].

For the applications so far made of yokes to statistical asymptotics the important part of a yoke is its germ round the diagonal, or even just its ∞-jet at the diagonal. In particular, this ∞-jet has been used to construct the associated tensors and affine connections. See, for instance, Blæsild [12] and references given there. (Recall that two functions on $M \times M$ have the same germ at Δ_M if they agree on some neighbourhood of Δ_M and that two such functions have the same ∞-jet at Δ_M if at each point of Δ_M their derivatives of any given order agree.)

By using partial maximisation, it is possible to define a concept of "profile" yoke. In Remark 3.3 we shall discuss briefly how this concept fits together with the idea of the symplectic form of a yoke, another concept to be introduced in Section 3.

2.2 Symplectic structures

Let N be a manifold of dimension k. A *symplectic structure* or *symplectic form* η on N is a non-singular closed 2-form on N, i.e.:

(i) η is a 2-form (skew-symmetric $(0,2)$-tensor) on N;

(ii) at each point n of N, the map $T_nN \to T_n^*N$ given by $X \mapsto \iota_X\eta$ is nonsingular (where $\iota_X\eta$ denotes the 1-form defined by $\iota_X\eta(Y) = \eta(X,Y)$ for Y in T_nN);

(iii) $\mathrm{d}\eta = 0$.

The standard example of a symplectic form is

$$\eta = \sum_{i=1}^{d} \mathrm{d}x^i \wedge \mathrm{d}y_i \quad (2.14)$$

on $\mathbb{R}^d \times \mathbb{R}^{d*}$, where $x^1, \ldots, x^d$ are the standard linear coordinates obtained from a base of $\mathbb{R}^d$ and $y_1, \ldots, y_d$ are the coordinates on $\mathbb{R}^{d*}$ obtained from the dual base. Locally, this is the only example of a symplectic form, because Darboux's Theorem (e.g. Abraham and Marsden [1, p. 175]) states that a 2-form η on N is a symplectic form if and only if round each point of N there is some coordinate system $(x^1, \ldots, x^d, y_1, \ldots, y_d)$ in which (2.14) holds. Note that this implies that k is necessarily even, $k = 2d$, and that the d-fold exterior product

$$\Omega = \eta \wedge \cdots \wedge \eta \tag{2.15}$$

is nowhere zero and so is a volume on N.

Formula (2.14) can be used on manifolds more general than $\mathbb{R}^d$. If $x^1, \ldots, x^d$ are coordinates on a manifold M and $y_1, \ldots, y_d$ are corresponding dual coordinates on the fibres of T^*M then (2.14) defines a canonical symplectic form η on the total space T^*M of the cotangent bundle of M. The symplectic form η can also be derived from the canonical 1-form θ_0 on T^*M given by

$$\theta_0(v) = \tau_{T^*M}(v)\big(T\tau_M^*(v)\big)$$

for every v in TT^*M. Here $\tau_{T^*M} : TT^*M \to T^*M$ denotes the projection map of the tangent bundle of T^*M and $T\tau_M^* : TT^*M \to TM$ denotes the tangent map of $\tau_M^* : T^*M \to M$. In terms of the coordinates $(x^1, \ldots, x^d, y_1, \ldots, y_d)$ on T^*M, the canonical 1-form is

$$\theta_0 = \sum_{i=1}^{d} y_i \, \mathrm{d}x^i , \tag{2.16}$$

and so

$$\eta = -\mathrm{d}\theta_0 .$$

Remark 2.1.— Aspects of symplectic structures such as:

(i) existence, i.e., determining when a given closed 2-form can be deformed into a symplectic structure,

(ii) classification, i.e., determining when two symplectic forms are equivalent,

(iii) the study of Lagrangian manifolds (submanifolds of maximal dimension on which the restriction of η is zero),

appear to be mainly of mathematical/mechanical interest and not of any direct statistical relevance (in this connection, see e.g. McDuff [20] and Weinstein [25]-[26]).

2.3 Hamiltonians and Lagrangians

There are two formulations of conservative mechanics: the Hamiltonian formulation takes place on the cotangent bundle and uses the canonical symplectic form $-\mathrm{d}\theta_0$, whereas the Lagrangian formulation takes place on the tangent bundle and leads to the use of second order differential equations. It is possible to pass from one formulation to the other by means of fibrewise Legendre transformation, i.e., performing Legendre transformation in each cotangent space $T^*_\omega M$ of M. More precisely, let H be a *Hamiltonian* on M, by which in the present context we mean a real-valued function on T^*M. Then the fibre derivative $\mathrm{F}H$ of H is just the derivative of H along the fibre. In terms of the coordinates $(x^1, \ldots, x^d, y_1, \ldots, y_d)$ on T^*M used in (2.16) to define θ_0,

$$\mathrm{F}H_i = \frac{\partial H}{\partial y_i}$$

where $x^1, \ldots, x^d$ are kept fixed. Suppose first, for simplicity, that H is *hyperregular*, i.e., $\mathrm{F}H : T^*M \longrightarrow TM$ is a diffeomorphism. Then the fibrewise Legendre transformation of H is the function $\check{H} : TM \longrightarrow \mathbb{R}$ defined by

$$\check{H}(v) = \alpha(v) - H(\alpha)\,, \tag{2.17}$$

where α is defined by $\mathrm{F}H_{(\alpha)} = v$. In coordinate terms,

$$\check{H}(x^1, \ldots, x^d, z^1, \ldots, z^d) = \sum_{i=1}^{d} z^i y_i - H(x^1, \ldots, x^d, y_1, \ldots, y_d)\,,$$

where

$$z^i = \frac{\partial H}{\partial y_i}(x^1, \ldots, x^d, y_1, \ldots, y_d)\,.$$

The function $\check{H}$ is usually called the *Lagrangian* corresponding to the Hamiltonian H.

In the context of yokes we need to relax the condition of hyper-regularity and assume just that $\mathrm{F}H$ is a diffeomorphism from some neighbourhood U of the zero section of T^*M to a neighbourhood $\mathrm{F}H(U)$ of the zero section of TM. Then we can define $\check{H}$ on $\mathrm{F}H(U)$ by (2.17).

3. The symplectic form given by a yoke

A yoke yields a symplectic form. Let $g : M \times M \to \mathbb{R}$ be a yoke on M. Then differentiation of g along the first copy of M yields a mapping

$$\varphi = d_1 g : M \times M \longrightarrow T^*M\,,$$

given in coordinate terms by

$$\varphi : (\omega, \omega') \longmapsto \big(\omega, g_{i;}(\omega, \omega')\mathrm{d}\omega^i\big)\,. \tag{3.1}$$

If g is the observed likelihood yoke (2.6) of a parametric statistical model then φ is essentially the score. For a general yoke g, the 2-form η on $M \times M$ is defined by

$$\eta = \varphi^*(-\mathrm{d}\theta_0)\,,$$

i.e., as the pull-back to $M \times M$ by φ of the canonical 2-form $-\mathrm{d}\theta_0$ on T^*M, so that for tangent vectors X, Y at some common point of $M \times M$,

$$\eta(X, Y) = (-\mathrm{d}\theta_0)\big(T\varphi(X), T\varphi(Y)\big)\,,$$

where $T\varphi$ denotes the derivative of φ.

Let $(\omega^1, \ldots, \omega^d)$ be local coordinates on M. Then, taking $(\omega^1, \ldots, \omega^d)$ as coordinates on the first factor in $M \times M$ and $(\omega'^1, \ldots, \omega'^d)$ as coordinates on the second factor, the coordinate expression for η is

$$\eta(\omega, \omega') = g_{i;j}(\omega, w')\,\mathrm{d}\omega^i \wedge \mathrm{d}\omega'^j\,. \tag{3.2}$$

It follows from (2.2) that there is a neighbourhood W of the diagonal of $M \times M$ on which η is non-singular. Thus η is a symplectic form on W. In fact, for a large class of statistical models the expected and observed likelihood yokes g have matrices $[g_{i;j}]$ of mixed partial derivatives which are non-singular everywhere on $M \times M$, so that the corresponding symplectic forms η are defined on all of $M \times M$. Note from (3.2) that η is special in containing no terms involving $\mathrm{d}\omega^i \wedge \mathrm{d}\omega^j$ or $\mathrm{d}\omega'^i \wedge \mathrm{d}w'^j$. This special feature is the basis of the characterisation, in Theorem 3.1 below, of those symplectic forms which arise from yokes.

Example 3.1

For the yoke considered in Example 2.1, the corresponding symplectic form η is the canonical symplectic form on $T^*\mathbb{R}^d$ given by (2.14). It follows from Darboux's Theorem that every symplectic form arises locally from some yoke.

Because, in general, η is non-singular only on some neighbourhood of Δ_M, rather than on all of $M \times M$, it is appropriate to consider germs of symplectic forms. Two symplectic forms, each defined on a neighbourhood of Δ_M, have the same germ at Δ_M if they agree in some neighbourhood of Δ_M. Similarly, two yokes on M have the same germ at Δ_M if they agree in some neighbourhood of Δ_M. It follows from (3.2) that the germ $[\eta]$ of η depends only on the germ $[g]$ of g. Thus we have a function

$$\Phi : [g] \longmapsto [\eta] \tag{3.3}$$

from the space of germs of yokes on M to the space of germs of symplectic forms around Δ_M.

As mentioned in Remark 2.1, a submanifold L of a symplectic manifold N is called *Lagrangian* if

(i) $i^*\eta = 0$, where $i : L \to N$ is the inclusion and η is the symplectic form on N,

(ii) $\dim N = 2 \dim L$.

It is useful to call a symplectic form on a neighbourhood W of Δ_M in $M \times M$ hv-*Lagrangian* (or horizontal and vertical Lagrangian) if all the submanifolds $\big(\{\omega\} \times M\big) \cap W$ and $\big(M \times \{\omega'\}\big) \cap W$ are Lagrangian, i.e., if the projections $p_1 : W \to M$ and $p_2 : W \to M$ are Lagrangian foliations in the sense of Arnol'd and Givental' [3, p. 36]. We shall call a symplectic form η on W dhv-*Lagrangian* (or diagonally-symmetric horizontal and vertical Lagrangian) if it is hv-Lagrangian and satisfies the symmetry condition

$$\eta\big((X,0),(0,Y)\big) = \eta\big((Y,0),(0,X)\big) \tag{3.4}$$

for all tangent vectors X, Y at the same point of M.

THEOREM 3.1.— *Let* $[\mathcal{Y}](M)$, $[\mathcal{N}\mathcal{Y}](M)$ *and* $[\mathrm{dhv}\, L](M)$ *denote the spaces of germs of yokes on* M, *of germs of normalised yokes on* M, *and of germs of* dhv-*Lagrangian symplectic forms around* Δ_M, *respectively. Then the function* Φ *given by (3.3) has the following properties:*

(i) $\Phi : [\mathcal{Y}](M) \to [\mathrm{dhv}\, L](M)$;

(ii) $\Phi[g] = \Phi([\overline{g}])$;

(iii) the restriction of Φ *to* $[\mathcal{NY}](M)$ *is a one-to-one map from* $[\mathcal{NY}](M)$ *to* $[\mathrm{dhv}\, L](M)$;

(iv) if M *is simply-connected then the restriction of* Φ *to* $[\mathcal{NY}](M)$ *maps onto* $[\mathrm{dhv}\, L](M)$ *and so is a bijection from* $[\mathcal{NY}](M)$ *to* $[\mathrm{dhv}\, L](M)$.

Proof

(i) If ι is the inclusion $(\{\omega\} \times M) \cap W \to W$ or $(M \times \{\omega'\}) \cap W \to W$ then $\iota^*(g_{i;j}\, \mathrm{d}\omega^i \wedge \mathrm{d}\omega'^j) = 0$. Property (3.4) follows from symmetry of $\mathfrak{g}_{i;j}$.

(ii) This follows from $g_{i;j} = \overline{g}_{i;j}$.

(iii) Let $\overline{g}$ and $\overline{g}'$ be normalised yokes which give rise to the same symplectic form. Then

$$\overline{g}_{i;j}(\omega, \omega') = \overline{g}'_{i;j}(\omega, \omega'),$$

so that

$$\overline{g}_{i;}(\omega, \omega') = \overline{g}'_{i;}(\omega, \omega') + \alpha_i(\omega),$$

for some functions $\alpha_1, \ldots, \alpha_d$. It follows from (2.1) that $\alpha_1 = \ldots = \alpha_d = 0$. Then

$$\overline{g}(\omega, \omega') = \overline{g}'(\omega, \omega') + \beta(\omega')$$

for some function β. From (2.5) it follows that $\beta = 0$ and so $\overline{g} = \overline{g}'$.

(iv) Any 2-form η on an open set in $M \times M$ can be expressed locally as

$$\eta = a_{ij}\, \mathrm{d}\omega^i \wedge \mathrm{d}\omega^j + b_{ij}\, \mathrm{d}\omega^i \wedge \mathrm{d}\omega'^j + c_{ij}\, \mathrm{d}\omega'^i \wedge \mathrm{d}\omega'^j .$$

If η is dhv-Lagrangian then $a_{ij} = c_{ij} = 0$ and so

$$\eta = b_{ij}\, \mathrm{d}\omega^i \wedge \mathrm{d}\omega'^j . \tag{3.5}$$

Since a symplectic form is closed, we have $\mathrm{d}\eta = 0$ and so

$$\frac{\partial b_{ij}}{\partial \omega^k} - \frac{\partial b_{kj}}{\partial \omega^i} = 0 \tag{3.6}$$

$$\frac{\partial b_{ij}}{\partial \omega'^k} - \frac{\partial b_{ik}}{\partial \omega'^j} = 0 . \tag{3.7}$$

Given ω and ω' in M, choose paths ξ and ζ in M such that $\xi(0) = \omega'$, $\xi(1) = \omega$, $\zeta(0) = \omega$ and $\zeta(1) = \omega'$. Define α_i by

$$\alpha_i(\omega, \omega') = \int_0^1 b_{ij}\big(\omega, \zeta(u)\big) \frac{\mathrm{d}\zeta^j}{\mathrm{d}u} \, \mathrm{d}u \, .$$

It follows from (3.7) and the simple-connectivity of M that α_i does not depend on the choice of ζ. Also,

$$\frac{\partial \alpha_i}{\partial \omega^k}(\omega, \omega') = \int_0^1 \frac{\partial b_{ij}}{\partial \omega^k}\big(\omega, \zeta(u)\big) \frac{\mathrm{d}\zeta^j}{\mathrm{d}u} \, \mathrm{d}u + b_{ik}(\omega, \omega') \, . \tag{3.8}$$

Define g by

$$g(\omega, \omega') = \int_0^1 \alpha_i\big(\xi(t), \omega'\big) \frac{\mathrm{d}\xi^i}{\mathrm{d}t} \, \mathrm{d}t \, .$$

It follows from (3.8), (3.4), (3.6) and the simple-connectivity of M that $g(w, \omega')$ does not depend on the choice of ξ, so that g is well-defined. It is simple to verify that g is a normalised yoke with η as its symplectic form.

Remark 3.1.— Yokes and 2-forms satisfying (3.5) are related to preferred point geometries. These geometries were introduced by Critchley *et al.* ([14], [15]) in order to provide a geometrical structure which reflects the statistical considerations that (i) the distribution generating the data has a distinguished role and (ii) this distribution need not belong to the statistical model used. Preferred point geometries are of statistical interest both because of their relevance to mis-specified models and because various geometrical objects which arise in statistics are preferred point metrics. It is convenient here to define a *preferred point metric* on a manifold M to be a map from M to $(0, 2)$-tensors on M, which is non-degenerate on the diagonal. (This is a slight weakening of the definition of Critchley *et al.*, who require the tensor to be symmetric and to be positive-definite on the diagonal.) In terms of coordinates $(\omega^1, \ldots, \omega^d)$, a preferred point metric b on a manifold M has the form

$$b_{ij}(\omega, \omega') \, \mathrm{d}\omega^i \otimes \mathrm{d}\omega^j \, .$$

A yoke g on M determines a preferred point metric on M by

$$g \longmapsto g_{i;j}(\omega, \omega') \, \mathrm{d}\omega^i \otimes \mathrm{d}\omega^j \, .$$

Note that the map from yokes on M to preferred point metrics on M is not onto (even at the level of germs), because a general preferred point metric

b does not satisfy the integrability conditions (3.6) and (3.7). However, it follows as in the proof of Theorem 3.1(iii) that the restriction of this map to the set of normalised yokes is one-to-one.

A preferred point metric b on M determines a 2-form on $M \times M$ satisfying (3.5) by

$$b_{ij}(\omega, \omega') \, \mathrm{d}\omega^i \otimes \mathrm{d}\omega^j \longmapsto b_{ij}(\omega, \omega') \, \mathrm{d}\omega^i \wedge \mathrm{d}\omega'^j \, .$$

Since this 2-form need not be closed, it is not in general a symplectic form. It is a consequence of the proof of Theorem 3.1(iv) that if M is simply connected then the preferred point metrics for which the corresponding 2-form is symplectic and satisfies (3.4) are precisely those which come from yokes.

Remark 3.2.— The construction of the symplectic form of a yoke behaves nicely under inclusion of submanifolds. Let $\iota : N \to M$ be an embedding (or more generally, an immersion) and let g be a yoke on M. If the pull-back to N by ι of the $(0,2)$-tensor $g_{i;j}$ on M is non-singular (as happens, in particular, if the metric of g is positive-definite) then g pulls back to a yoke $g \circ (\iota \times \iota)$ on N. It is easy to verify that the corresponding symplectic form η_N satisfies $\eta_N = (\iota \times \iota)^* \eta_M$.

Remark 3.3.— The construction of the symplectic form behaves nicely under taking profile yokes. Let $p : M \to N$ be a surjective submersion (= fibred manifold) and let g be a yoke on M satisfying $g(m_1, m_2) \leq 0$ (so that $-g$ is a contrast function). The *profile yoke* of g is the function $\widetilde{g} : N \times N \to \mathbb{R}$, defined by

$$\widetilde{g}(n_1, n_2) = \sup\{g(m_1, m_2) \mid p(m_i) = n_i \, , \ i = 1, 2\} \, .$$

We can choose implicitly a section $s : N \times N \to M \times M$ of $p \times p$ (i.e., a function s with $(p \times p) \circ s$ the identity of $N \times N$) such that

$$\widetilde{g} = g \circ s \, .$$

Then the symplectic form η_N of $\widetilde{g}$ satisfies

$$\eta_N = s^* \eta_M \, .$$

For example, let M be the parameter space of a composite transformation model with group G, let p be the quotient map $p : M \to M/G$, and let g be the expected likelihood yoke. Then the corresponding profile yoke

$\widetilde{g}$ is $-\widetilde{I}$, where $\widetilde{I}$ is the Kullback-Leibler profile discrimination considered by Barndorff-Nielsen and Jupp [10]. The tensors $\mathfrak{T}_{i;j}$ and $\mathfrak{T}_{i;jk}$ on M/G obtained by applying Blæsild's [12] general construction to the yoke $\widetilde{g}$ are the transferred Fisher information $p_! i$ and the transferred skewness tensor $p_! D$ of Barndorff-Nielsen and Jupp [10].

Remark 3.4.— A rather different connection between statistics and symplectic structures is given by Friedrich [17]. His construction requires a manifold M, a vector field X on M and an X-invariant volume form λ on M. These give rise to a 2-form η on $P(M, \lambda)$, the space of probability measures on M which are absolutely continuous with respect to λ. If X has a dense orbit then η is a symplectic form on $P(M, \lambda)$.

4. Some uses of the symplectic form of a yoke

In symplectic geometry important ways in which a symplectic structure η is used are:

(i) to raise and lower tensors;

(ii) to transform (the derivatives of) real-valued functions into vector fields;

(iii) to provide a volume Ω, thus enabling integration over the manifold (in mechanics, over the phase space, i.e., the cotangent space).

We now apply these to the symplectic form η of a yoke g on M. Recall that η is defined on some neighbourhood W of Δ_M in $M \times M$.

4.1 Raising and lowering tensors

A symplectic form η enables raising and lowering of tensors in the same way that a Riemannian metric does. In particular, the "musical isomorphism" $\sharp$, which raises 1-forms to vector fields, and its inverse $\flat$ (which lowers vector fields to 1-forms) are defined by

$$\eta(\alpha^\sharp, X) = \alpha(X)$$
$$Y^\flat(X) = \eta(Y, X)$$

for 1-forms α and vector fields X and Y. For the symplectic form (3.2) of a yoke, the coordinate expressions of $\sharp$ and $\flat$ are

$$(\mathrm{d}\omega^i)^\sharp = -g^{i;j}\frac{\partial}{\partial \omega'^j}, \quad (\mathrm{d}\omega'^j)^\sharp = g^{i;j}\frac{\partial}{\partial \omega^i},$$
$$\left(\frac{\partial}{\partial \omega^i}\right)^\flat = g_{i;j}\,\mathrm{d}\omega'^j, \quad \left(\frac{\partial}{\partial \omega'^j}\right)^\flat = -g_{i;j}\,\mathrm{d}\omega^i,$$

where $g^{i;j}$ is the (2.0)-tensor inverse to $g_{i;j}$. The construction in Section 5 of the Lagrangian of a yoke can be expressed in terms of $\sharp$; see (5.2).

4.2 The vector field of a yoke

A yoke g on M gives rise as follows to a vector field X_g on W. The derivative $\mathrm{d}g$ of g is a 1-form. Raising this using $\sharp$ yields the vector field $(\mathrm{d}g)^\sharp = X_g$. In coordinate terms, $X_g(\omega, \omega')$ is given by

$$g_{;j}(\omega,\omega')g^{i;j}(\omega,\omega')\frac{\partial}{\partial \omega^i} - g^{i;j}(\omega,\omega')g_{i;}(\omega,\omega')\frac{\partial}{\partial \omega'^j}. \tag{4.1}$$

By Liouville's Theorem on locally Hamiltonian flows (see, e.g., Abraham and Marsden [1, pp. 188-189]), the flow of X_g preserves the volume Ω on W defined by (2.15). If g is an observed likelihood yoke, then the vector field X_g describes a joint evolution of the parameter and the observation.

The vector field X_g can be used to differentiate functions. Let h be a real-valued function on $M \times M$. Then the derivative of h along X_g is $\mathrm{d}h(X_g)$. An alternative expression for $\mathrm{d}h(X_g)$ is

$$\mathrm{d}h(X_g) = \{h, g\}, \tag{4.2}$$

where $\{h, g\}$ is the *Poisson bracket* (with respect to the symplectic form η of g) defined by

$$\{h, g\} = \eta(X_h, X_g). \tag{4.3}$$

(See, e.g., Abraham and Marsden [1, p. 192].) It is clear from (4.3) that $\{g, g\} = 0$. This suggests that one way of comparing two normalised yokes h and g is by their Poisson bracket. In coordinate terms we have

$$\{h, g\} = g_{;j}g^{i;j}h_{i;} - h_{;j}g^{i;j}g_{i;}. \tag{4.4}$$

Remark 4.1.— The Poisson bracket with respect to a fixed symplectic structure is skew-symmetric. However, because the symplectic structure η used in (4.3) depends on g, the skew-symmetry property

$$\{h, g\} = -\{g, h\}$$

does not hold in general here.

The behaviour of $\{h, g\}$ near the diagonal of $M \times M$ is explored in Proposition 4.1. It shows that the 3-jet of $\{h, g\}$ at the diagonal cannot detect differences in metrics between two normalised yokes h and g but that some differences in skewness tensors can be detected.

PROPOSITION 4.1.— *Let h and g be normalised yokes on M with skewness tensors*

$$S_{rst} = \mathfrak{h}_{r;st} - \mathfrak{h}_{st;r}$$

and

$$T_{rst} = \mathfrak{g}_{r;st} - \mathfrak{g}_{st;r} .$$

Put $k = \{h, g\}$. Then:

(o) $\mathfrak{k} = 0$;

(i) $\mathfrak{k}_{r;} = \mathfrak{k}_{;r} = 0$;

(ii) $\mathfrak{k}_{rs;} = \mathfrak{k}_{r;s} = \mathfrak{k}_{;rs} = 0$;

(iii) $\mathfrak{k}_{rst;} = -\mathfrak{k}_{rs;t} = \mathfrak{k}_{r;st} = -\mathfrak{k}_{;rst} = T_{rst}(h, g)$, *where $T_{rst}(h, g)$ is the tensor defined by*

$$\begin{aligned} T_{rst}(h, g) &= -\{(\mathfrak{h}_{rs;t} + \mathfrak{h}_{t;rs}) - (\mathfrak{g}_{rs;j} + \mathfrak{g}_{j;rs})\mathfrak{g}^{i;j}\mathfrak{h}_{i;t}\}[3] \\ &= -2\{\mathfrak{h}_{rs;t} - \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{h}_{i;t}\}[3] - \{S_{rst} - T_{rsj}\mathfrak{g}^{i;j}\mathfrak{h}_{i;t}\}[3] , \end{aligned}$$

[3] *indicating a sum of three terms obtained by appropriate permutation of indices.*

Furthermore, if

$$\mathfrak{g}_{rs;} = \mathfrak{h}_{rs;} \tag{4.5}$$

and

$$\mathfrak{g}_{rst;} = \mathfrak{h}_{rst;} \tag{4.6}$$

then

$$\mathfrak{k}_{rst;} = \mathfrak{k}_{rs;t} = \mathfrak{k}_{r;st} = \mathfrak{k}_{;rst} = 0 \tag{4.7}$$

and

$$\begin{aligned} \mathfrak{k}_{rstu;} &= -\mathfrak{k}_{rst;u} = \mathfrak{k}_{rs;tu} = -\mathfrak{k}_{r;stu} = \mathfrak{k}_{;rstu} \\ &= 4(\mathfrak{h}_{rstu;} - \mathfrak{g}_{rstu;}) + (\mathfrak{h}_{rst;u} - \mathfrak{g}_{rst;u})[4]\,. \end{aligned} \tag{4.8}$$

Proof.— Part (o) is immediate from (4.4) and (2.8). Differentiating (4.4) with respect to ω and applying (2.8) yields (i). Repeated differentiation of (i) then shows that k satisfies (2.9)-(2.11). Differentiating (4.4) twice and three times with respect to ω, using Leibniz' rule for differentiation of products. and applying (2.8), (2.9) and (2.10), we obtain (ii) and

$$\begin{aligned} \mathfrak{k}_{rst;} &= \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{h}_{it;}[3] - \mathfrak{h}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{g}_{it;}[3] + \\ &\quad + \mathfrak{g}_{r;j}(-\mathfrak{g}^{i;u}\mathfrak{g}_{vs;u}\mathfrak{g}^{v;j})\mathfrak{h}_{it;}[6] - \mathfrak{h}_{r;j}(-\mathfrak{g}^{i;u}\mathfrak{g}_{vs;u}\mathfrak{g}^{v;j})\mathfrak{g}_{it;}[6] + \\ &\quad + \mathfrak{g}_{r;j}\mathfrak{g}^{i;j}\mathfrak{h}_{ist;}[3] - \mathfrak{h}_{r;j}\mathfrak{g}^{i;j}\mathfrak{g}_{ist;}[3] \\ &= (\mathfrak{h}_{rs;t} + \mathfrak{h}_{rst;})[3] - \mathfrak{h}_{r;j}\mathfrak{g}^{i;j}(-\mathfrak{g}_{st;i} + \mathfrak{g}_{ist;} + \mathfrak{g}_{is;t} + \mathfrak{g}_{it;s})[3] \\ &= -(\mathfrak{h}_{rs;t} + \mathfrak{h}_{t;rs})[3] + \mathfrak{h}_{r;j}\mathfrak{g}^{i;j}(\mathfrak{g}_{i;st} + \mathfrak{g}_{st;i})[3]\,. \end{aligned} \tag{4.9}$$

Differentiation of (ii) now yields (iii).

Differentiation of (4.5) together with (4.6) and (2.9)-(2.11) yields

$$\mathfrak{g}_{rs;t} = \mathfrak{h}_{rs;t} \quad \text{and} \quad \mathfrak{g}_{r;st} = \mathfrak{h}_{r;st}\,, \tag{4.10}$$

from which (4.7) follows.

Differentiating (4.4) four times with respect to ω, using Leibniz' rule for differentiation of products, and applying (4.10), we obtain

$$\mathfrak{k}_{rstu;} = (-\mathfrak{g}_{rst;u} + \mathfrak{h}_{rstu;})[4] - (-\mathfrak{h}_{rst;u} + \mathfrak{g}_{rstu;})[4]\,.$$

Differentiation of (4.7) now yields (4.8). □

Example 4.1

Let g be the expected likelihood yoke of the family of samples of size n from a normal distribution with unknown mean μ and unknown variance σ^2. Then

$$g(\omega, \omega') = n\left\{\log \sigma' - \log \sigma - \frac{1}{2}\frac{\sigma'^2}{\sigma^2} - \frac{1}{2}\frac{(\mu - \mu')^2}{\sigma^2} + \frac{1}{2}\right\}.$$

Consider a location-scale model with probability density functions

$$p(x; \mu, \sigma) = \frac{1}{\sigma}\exp\left\{q\left(\frac{x - \mu}{\sigma}\right)\right\}$$

for some function q. Suppose that for samples of size n from a distribution in this family there is no non-trivial sufficient statistic (as happens, for example, in a family of hyperbolic distributions with given shape). Take as an ancillary statistic $(a_1, \ldots, a_n)$, with

$$a_i = \frac{x_i - \widehat{\mu}}{\widehat{\sigma}},$$

where $\widehat{\mu}$ and $\widehat{\sigma}$ denote the maximum likelihood estimates of μ and σ. Let h be the corresponding observed likelihood yoke. Then

$$h(\omega, \omega') = n(\log \sigma' - \log \sigma) + \sum_{i=1}^{n}\left\{q\left(\frac{a_i\sigma' + \mu' - \mu}{\sigma}\right) - q(a_i)\right\}$$

and

$$\{h, g\} = g_{;\mu}g^{\sigma;\mu}h_{\sigma;} - h_{;\mu}g^{\sigma;\mu}g_{\sigma;} + g_{;\sigma}g^{\sigma;\sigma}h_{\sigma;} - h_{;\sigma}g^{\sigma;\sigma}g_{\sigma;}\,,$$

where

$$g_{;\mu} = n\frac{\mu - \mu'}{\sigma^2}$$

$$g_{\sigma;} = n\frac{1}{\sigma}\left\{\frac{\sigma'^2}{\sigma^2} + \frac{(\mu - \mu')^2}{\sigma^2} - 1\right\}$$

$$g_{;\sigma} = n\frac{1}{\sigma}\left\{\frac{\sigma^2 - \sigma'^2}{\sigma\sigma'}\right\}$$

$$h_{;\mu} = \frac{1}{\sigma}\sum_{i=1}^{n} q'(z_i)$$

$$h_{\sigma;} = -\frac{1}{\sigma}\sum_{i=1}^{n}\{1 + z_i q'(z_i)\}$$

$$h_{;\sigma} = \frac{1}{\sigma}\sum_{i=1}^{n}\left\{\frac{\sigma}{\sigma'} + a_i q'(z_i)\right\}$$

$$g^{\sigma;\mu} = \frac{\sigma^2(\mu - \mu')}{n\sigma'}$$

$$g^{\sigma;\sigma} = \frac{\sigma^3}{2n\sigma'},$$

with

$$z_i = \frac{a_i\sigma' + \mu' - \mu}{\sigma}.$$

Another way in which vector fields can arise from statistical models is given by Nakamura ([22], [23]). He considers parametric statistical models (e.g., regular exponential models) in which the parameter space M is an even-dimensional vector space and the Fisher information metric i is given by

$$i_{rs} = \frac{\partial^2\psi}{\partial\omega^r\partial\omega^s}$$

for some potential function ψ. The metric can be used to raise the derivative of ψ to a vector field $(\mathrm{d}\psi)^\sharp$ on M, given in coordinates by

$$i^{rs}\frac{\partial}{\partial\omega^s},$$

where $[i^{rs}]$ is the inverse of the matrix $[i_{rs}]$. If M has dimension $2k$ then it can be given the symplectic structure of the cotangent space T^*V of a k-dimensional vector space V. Nakamura shows that the vector field $(\mathrm{d}\psi)^\sharp$ is Hamiltonian and is completely integrable, i.e., there are functions $f_1, \ldots, f_k$ on M which are constant under the flow of the vector field, satisfy $\{f_i, f_j\} = 0$ for $1 \leq i, j \leq k$, and such that $\mathrm{d}f_1, \ldots, \mathrm{d}f_k$ are linearly independent almost everywhere (see Abraham and Marsden [1, pp. 392-393]). We have found no non-trivial analogue in our symplectic context of Nakamura's complete integrability result.

4.3 The volume of a yoke

Recall from (2.15) that a yoke g on M gives rise to a volume Ω on some neighbourhood W of the diagonal in $M \times M$. For many yokes of statistical interest W can be taken to be all of $M \times M$. However, for a general yoke, W will be a proper subset of $M \times M$. The main way in which volumes are used is as objects to be integrated over the manifolds on which they live, so in such cases there is no obvious choice of manifold over which Ω should be integrated. (One possibility is to integrate Ω over the maximal such W but the interpretation of the value of this integral is not clear.)

An alternative way of using the volume Ω is to compare it with other volumes on W, in particular with the restriction to W of the geometric measure obtained from the metric of g. More precisely, the yoke g gives a (possibly indefinite) Riemannian metric on M and so a volume on M given in coordinate terms as

$$\left|g_{i;j}(\omega,\omega)\right|^{1/2},$$

where $|\cdot|$ denotes the absolute value of the determinant. The corresponding product measure on $M \times M$ is the volume with coordinate form

$$\left|g_{i;j}(\omega,\omega)\right|^{1/2}\left|g_{i;j}(\omega',\omega')\right|^{1/2}. \tag{4.11}$$

It follows from (2.9) that (4.11) can also be written as

$$\left|g_{ij;}(\omega,\omega)\right|^{1/2}\left|g_{;ij}(\omega',\omega')\right|^{1/2}.$$

Since Ω is given in coordinate form as $d!\left|g_{i;j}(\omega,\omega')\right|$, volume (4.11) can be written as $d!\,h\Omega$, where the function $h : W \longrightarrow \mathbb{R}$ is given in coordinate form by

$$h(\omega,\omega') = \left|g_{i;j}(\omega,\omega)\right|^{1/2}\left|g_{i;j}(\omega,\omega')\right|^{-1}\left|g_{i;j}(\omega',\omega')\right|^{1/2}. \tag{4.12}$$

Note that, in general, h is not a yoke. However, h can be used to obtain further yokes from g. For any real λ define $g^{[\lambda]} : M \times M \longrightarrow \mathbb{R}$ by

$$g^{[\lambda]} = h^{\lambda} g\,.$$

Since $h = 1$ on the diagonal, the following proposition shows that $g^{[\lambda]}$ is a normalised yoke with the same metric as g. Note that the germs at Δ_M of h and $g^{[\lambda]}$ depend only on the germ of g.

PROPOSITION 4.2.— *Let g be a normalised yoke on M and $h : M \times M \to \mathbb{R}$ be any function such that $h = 1$ on the diagonal Δ_M. Then*

(i) the product hg is a normalised yoke on M having the same metric as g;

(ii) the function

$$k = g + \log h$$

satisfies

$$k_{i;}(\omega; \omega) + k_{;i}(\omega; \omega) = 0 . \tag{4.13}$$

Proof.— These are simple calculations. □

Remark 4.2.— Note that if k in Proposition 4.2 satisfies also

$$\text{the matrix } \left[k_{i;j}(\omega; \omega) + k_{j;i}(\omega; \omega)\right] \text{ is non-singular} \tag{4.14}$$

then k is rather like a yoke but need not satisfy (2.1). Note also that conditions (4.13) and (4.14) provide a generalisation of the concept of a normalised yoke. Also, in contrast to the definition of a yoke, (4.13) and (4.14) involve the two arguments in a symmetrical way.

Remark 4.3.— Let k be a function on $M \times M$ satisfying (4.14). Define $g_k : M \times M \to \mathbb{R}$ by

$$g_k(\omega, \omega') = k(\omega, \omega') + k(\omega', \omega) - k(\omega, \omega) - k(\omega', \omega') .$$

Then g_k is a normalised yoke on M.

Remark 4.4.— One context in which the "adjustment factor" h of (4.12) occurs in statistics is in a variant of the p^*-formula for the distribution of the score

$$s_* = (s_1, \ldots, s_d) = \left(\frac{\partial \ell}{\partial \omega^1}, \ldots \frac{\partial \ell}{\partial \omega^d}\right) .$$

A suitable starting point for this is the p^*-formula

$$p^*(\widehat{\omega}; \omega \mid a) = c(\omega, a) \left|\,\widehat{\jmath}\,\right|^{1/2} e^{\ell - \widehat{\ell}} \tag{4.15}$$

for the distribution of the maximum likelihood estimator. Here $|\widehat{\jmath}|$ is the determinant of the observed information matrix $j = j(\omega; \widehat{\omega}, a)$ evaluated at $\widehat{\omega}$. For an extensive discussion of this formula and its applications see Barndorff-Nielsen and Cox [9]. The expression $p^*(\widehat{\omega};\omega \mid a)\,\mathrm{d}\widehat{\omega}^1\cdots\mathrm{d}\widehat{\omega}^d$ represents a volume on M. Changing the variable in (4.15) from $\widehat{\omega}$ to the score s_* gives the p^*-formula

$$p^*(s_*;\omega \mid a) = c\big|\,\widehat{\jmath}\,\big|^{1/2}\big|\ell_;\big|^{-1}\,e^{\ell-\widehat{\ell}}$$

for the distribution of s_*. The expression $p^*(s_*;\omega \mid a)\,\mathrm{d}s_1\cdots\mathrm{d}s_d$ represents a volume on the cotangent space $T^*_\omega M$ to M at ω. Now $\mathfrak{j}$ defined by $\mathfrak{j}(\omega) = \jmath(\omega,\omega)$ is an inner-product on $T_\omega M$, so that $\mathfrak{j}^{-1}$ is an inner-product on $T^*_\omega M$. The geometric measure of $\mathfrak{j}^{-1}$ is a measure on $T^*_\omega M$ represented by $\big|\mathfrak{j}\big|^{-1/2}$. Then

$$\frac{p^*(s_*;\omega \mid a)}{\big|\mathfrak{j}\big|^{-1/2}} \tag{4.16}$$

is a ratio of measures on $T^*_\omega M$. Let $\mathfrak{j}^{-1/2}$ be any square root of $\mathfrak{j}^{-1}$ and define the *standardised score* $\overline{s}_*$ by

$$\overline{s}_* = \mathfrak{j}^{-1/2}s_*\,.$$

Barndorff-Nielsen [8, sect. 7.3] derived

$$p^*(\overline{s}_*;\omega \mid a) = c\big|\,\widehat{\jmath}\,\big|^{1/2}\big|\ell_;\big|^{-1}\big|\mathfrak{j}\big|^{1/2}\,e^{\ell-\widehat{\ell}}$$

as an approximation to the density of $\overline{s}_*$. Note that $p^*(\overline{s}_*;\omega \mid a)$ is (4.16).

Define $k : M \times M \to \mathbb{R}$ by

$$k(\omega,\widehat{\omega}) = \log\big(p^*(\overline{s}_*;\omega \mid a)\big)\,.$$

Since the observed likelihood yoke (2.6) is g, given by

$$g(\omega,\widehat{\omega}) = \ell - \widehat{\ell} = \ell(\omega,\widehat{\omega}) - \ell(\widehat{\omega},\widehat{\omega})\,,$$

we have, neglecting an additive constant,

$$k = g + \log h\,,$$

where h is the "adjustment factor" defined in (4.12). From Proposition 4.2, k satisfies (4.13). By differentiating (4.12) twice, we obtain

$$k_{i;j}(\omega,\omega) = \mathfrak{g}_{i;j} - \mathfrak{T}_{ik;j\ell}\mathfrak{g}^{k;\ell}\,,$$

where

$$\mathfrak{T}_{rs;tu} = \mathfrak{g}_{rs;tu} - \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{g}_{i;tu} \tag{4.17}$$

is one of the tensors introduced by Blaesild [12].

5. The Lagrangian of a yoke

Let g be a yoke on M. As discussed after equation (3.2), we can choose a neighbourhood W of Δ_M in $M \times M$ such that the restriction to W of φ, where φ is defined by (3.1), is a diffeomorphism onto $\varphi(W)$. We define the *Hamiltonian* of g as the function $H : \varphi(W) \to \mathbb{R}$ given by

$$H = g \circ \varphi^{-1} .$$

In coordinate terms, H is given by

$$H(\omega, a_i\, \mathrm{d}\omega^i) = g(\omega, \omega') ,$$

where ω' is determined by

$$a_i = g_{i;}(\omega, \omega') .$$

It follows from (2.2) that we can choose W such that the restriction to $\varphi(W)$ of the fibre derivative $\mathrm{F}H$ of H is a diffeomorphism onto its image. By the definition in subsection 2.3, the Lagrangian corresponding to H is $\check{H} : \mathrm{F}H(\varphi(W)) \to \mathbb{R}$, the fibrewise Legendre transform of H. However, it is convenient to refer to

$$\check{g} = \check{H} \circ \mathrm{F}H \circ \varphi : W \to \mathbb{R} \tag{5.1}$$

as the *Lagrangian* of g. An alternative expression for $\check{g}$ is

$$\check{g} = \mathrm{d}_1 g\left(\left(\mathrm{d}_2 g\right)^{\sharp}\right) - g = -\mathrm{d}_2 g\left(\left(\mathrm{d}_1 g\right)^{\sharp}\right) - g . \tag{5.2}$$

In coordinate terms, $\check{g}$ is given by

$$\check{g}(\omega, \omega') = g_{;j}(\omega, \omega') g^{i;j}(\omega, \omega') g_{i;}(\omega, \omega') - g(\omega, \omega') . \tag{5.3}$$

Because the restriction to W of $\mathrm{F}H \circ \varphi$ is a diffeomorphism, it follows from (5.1) that $\check{g}$ is equivalent to $\check{H}$. Since $\check{g}$ is similar to g in being a function defined on a neighbourhood of Δ_M in $M \times M$, it is often convenient to consider $\check{g}$ rather than $\check{H}$.

Example 5.1

Let g be the expected likelihood yoke of the family of d-variate normal distributions with known covariance matrix Σ. Then

$$g(\omega,\omega') = -\frac{1}{2}(\omega-\omega')^t\Sigma^{-1}(\omega-\omega'),$$

so that g is quadratic and $\breve{g} = g$.

Example 5.2

Consider a regular exponential model with log-likelihood function

$$\ell(\omega;x) = \omega^t t(x) - \kappa(\omega),$$

where the canonical parameter ω and the canonical statistic t take values in (appropriate subsets of) a finite-dimensional inner-product space. Then the expected (or observed) likelihood yoke g of this model is

$$g(\omega,\omega') = (\omega-\omega')^t\tau(\omega') - \kappa(\omega) + \kappa(\omega'),$$

where

$$\tau(\omega') = \frac{\partial\kappa}{\partial\omega}(\omega').$$

Calculation gives

$$\begin{aligned} g_{i;} &= \tau_i(\omega') - \tau_i(\omega) \\ g_{;j} &= (\Sigma_{\omega'})_{ij}(\omega-\omega')^i \\ g_{i;j} &= (\Sigma_{\omega'})_{ij}, \end{aligned}$$

where

$$\Sigma_{\omega'} = \frac{\partial^2\kappa}{\partial\omega\partial\omega^t}(\omega').$$

Then

$$g_{;j}g^{i;j}g_{i;} = (\omega-\omega')^t(\tau(\omega') - \tau(\omega))$$

so

$$\breve{g}(\omega,\omega') = g(\omega',\omega),$$

i.e., $\breve{g} = g^*$ (g^* being the dual yoke of g).

Example 5.3

In quantum statistical mechanics the state space is a space M of non-negative Hermitian operators with trace 1 (see. e.g., Balian [4, p. 75]). The *von Neumann relative entropy* is the function $S : M \times M \to \mathbb{R}$ defined by

$$S(\widehat{D} \mid \widehat{D}') = \mathrm{tr}(\widehat{D}' \ln \widehat{D}' - \widehat{D}' \ln \widehat{D}),$$

where if

$$\widehat{D} = U \operatorname{diag}(\lambda_1, \ldots, \lambda_d) U^*$$

with U unitary then

$$\ln \widehat{D} = U \operatorname{diag}(\ln(\lambda_1), \ldots, \ln(\lambda_d)) U^* .$$

The function g, defined by

$$g(\widehat{D}, \widehat{D}') = -S(\widehat{D}' \mid \widehat{D}),$$

is a normalised yoke on M with Lagrangian $\breve{g} = g^*$.

Remark 5.1.— Examples 5.1, 5.2 and 5.3 are special cases of the following. Let V be an open subset of a vector space and let $\psi : V \to \mathbb{R}$ be a smooth function with non-singular second derivative matrix. Then g defined by

$$g(x, y) = (x - y)^t \frac{\partial \psi}{\partial x}(y) - \psi(x) + \psi(y),$$

is a yoke on V. A straight-forward calculation shows that $\breve{g} = g^*$. Thus, in particular, $\breve{g} = g^*$ for the expected or observed likelihood yoke of a regular exponential model. However, Example 5.4 shows that this pleasant property does not always hold even for transformation models. Note that V with the metric of the yoke g is dually flat in the sense of Amari [2, p. 80].

Example 5.4

Consider the set of von Mises distributions with given concentration parameter κ. The parameter space and sample space are the unit circle and the model function is

$$f(\theta; \omega) = \exp\{\kappa \cos(\theta - \omega) - a(\kappa)\},$$

where ω is the mean direction and $a(\kappa) = \log(I_0(\kappa))$, I_0 denoting the modified Bessel function of the first kind and order zero. Then the expected likelihood yoke g is given by

$$g(\omega, \omega') = \kappa A(\kappa) \exp\{\cos(\omega - \omega') - 1\},$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ with I_1 denoting the modified Bessel function of the first kind and order 1. Putting

$$s = \sin(\omega - \omega'), \quad c = \cos(\omega - \omega')$$

and using straight-forward calculations, we obtain

$$\begin{aligned} g_{1;} &= -\kappa A(\kappa) s \\ g_{;1} &= \kappa A(\kappa) s \\ g_{1;1} &= \kappa A(\kappa) c, \end{aligned}$$

so

$$\breve{g} = \kappa A(\kappa)\left\{1 - \frac{1}{c}\right\} = c^{-1} g .$$

Since $g(\omega, \omega') = g(\omega', \omega)$, we have $g = g^*$ and so $\breve{g} \neq g^*$. Also,

$$\begin{aligned} \breve{g}_{1;} &= -\kappa A(\kappa)\frac{s}{c^2} \\ \breve{g}_{;1} &= \kappa A(\kappa)\frac{s}{c^2} \\ \breve{g}_{1;1} &= \kappa A(\kappa)\frac{2 - c^2}{c^3}, \end{aligned}$$

so

$$\breve{\breve{g}} = \kappa A(\kappa)\frac{1 - 2c + c^3}{c(2 - c^2)}$$

and $\breve{\breve{g}} \neq g$.

Remark 5.2.— The first term on the right hand side of (5.3) is reminiscent of

$$g_{j;}(\omega, \omega')g^{i;j}(\omega, \omega)g_{i;}(\omega, \omega'),$$

which is the (quadratic) score statistic in the case where g is an observed or expected likelihood yoke. Thus

$$g_{;j}(\omega, \omega')g^{i;j}(\omega, \omega')g_{i;}(\omega, \omega')$$

can be regarded as a "mixed score statistic" and

$$g_{;i}(\omega, \omega')g^{i;j}(\omega', \omega')g_{;j}(\omega, \omega')$$

can be regarded as a "dual score statistic".

Now we consider how close $\breve{g}$ is to g and show that $\breve{g}$ is closer to g^* than to g.

THEOREM 5.1.— *Let g be a normalised yoke and denote by g^* the dual yoke of g. Define the "skewness tensor" T_{rst} of the yoke g by*

$$T_{rst} = \mathfrak{g}_{r;st} - \mathfrak{g}_{st;r}$$

as in (2.4) and denote by $\nabla^{(\alpha)}$, $\nabla^{*(\alpha)}$ and $\breve{\nabla}^{(\alpha)}$ the α-connections of g, g^* and $\breve{g}$, respectively. Following Blæsild [12], define the tensor $\mathfrak{T}_{rs;tu}$ by

$$\mathfrak{T}_{rs;tu} = \mathfrak{g}_{rs;tu} - \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{g}_{i;tu}\,,$$

as in (4.17). Then:

(i) $\breve{g}$ *is a normalised yoke;*

(ii) $\breve{g}$ *has the same metric as g and g^*;*

(iii)
$$\begin{aligned} \breve{\mathfrak{g}}_{rst;} &= \mathfrak{g}^*_{rst;} = \mathfrak{g}_{rst;} - T_{rst} \\ \breve{\mathfrak{g}}_{rs;t} &= \mathfrak{g}^*_{rs;t} = \mathfrak{g}_{rs;t} + T_{rst} \\ \breve{\mathfrak{g}}_{r;st} &= \mathfrak{g}^*_{r;st} = \mathfrak{g}_{r;st} - T_{rst} \\ \breve{\mathfrak{g}}_{;rst} &= \mathfrak{g}^*_{;rst} = \mathfrak{g}_{;rst} + T_{rst}\ ; \end{aligned}$$

(iv) $\breve{\mathfrak{g}}_{r;st} - \breve{\mathfrak{g}}_{st;r} = -T_{rst}$*;*

(v) $\breve{\nabla}^{(\alpha)} = \nabla^{*(\alpha)} = \nabla^{(-\alpha)}$*;*

(vi)
$$\begin{aligned} \breve{\mathfrak{g}}_{rstu;} - \mathfrak{g}^*_{rstu;} &= -\mathfrak{T}_{rs;tu}[6] \\ \breve{\mathfrak{g}}_{rst;u} - \mathfrak{g}^*_{rst;u} &= \mathfrak{T}_{rs;tu}[6] \\ \breve{\mathfrak{g}}_{rs;tu} - \mathfrak{g}^*_{rs;tu} &= -\mathfrak{T}_{rs;tu}[6] \\ \breve{\mathfrak{g}}_{r;stu} - \mathfrak{g}^*_{r;stu} &= \mathfrak{T}_{rs;tu}[6] \\ \breve{\mathfrak{g}}_{;rstu} - \mathfrak{g}^*_{;rstu} &= -\mathfrak{T}_{rs;tu}[6]\,. \end{aligned}$$

Proof.— Differentiation of (5.3) with respect to ω^r yields

$$\breve{g}_{r;} = g_{r;j}g^{i;j}g_{i;} + g_{;j}(-g^{i;u}g_{vr;u}g^{v;j})g_{i;} + g_{;j}g^{i;j}g_{ir;} - g_{r;} \tag{5.4}$$

and so

$$\breve{\mathfrak{g}}_{r;} = 0\,,$$

by (2.8). Differentiation of (5.4) with respect to ω^s and evaluation at $\omega = \omega'$ gives

$$\breve{\mathfrak{g}}_{rs;} = \mathfrak{g}_{r;j}\mathfrak{g}^{i;j}\mathfrak{g}_{is;} + \mathfrak{g}_{s;j}\mathfrak{g}^{i;j}\mathfrak{g}_{ir;} - \mathfrak{g}_{rs;} = \mathfrak{g}_{rs;} = \mathfrak{g}^{*}_{;rs} \,. \tag{5.5}$$

Similarly,

$$\breve{\mathfrak{g}}_{;rs} = \mathfrak{g}_{;rs} = \mathfrak{g}^{*}_{rs;} \,. \tag{5.6}$$

Since g is a normalised yoke, we see from (2.9) that

$$\mathfrak{g}_{rs;} = -\mathfrak{g}_{r;s} = -\mathfrak{g}^{*}_{s;r} = \mathfrak{g}^{*}_{rs;} \tag{5.7}$$

and so (i) and (ii) follow. Evaluating the second derivative of (5.4) at $\omega = \omega'$, we obtain

$$\begin{aligned}\breve{\mathfrak{g}}_{rst;} &= \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}\mathfrak{g}_{it;}[3] + \mathfrak{g}_{r;j}(-\mathfrak{g}^{i;u}\mathfrak{g}_{vs;u}\mathfrak{g}^{v;j})\mathfrak{g}_{ti;}[6] + \mathfrak{g}_{r;j}\mathfrak{g}^{i;j}\mathfrak{g}_{ist;}[3] - \mathfrak{g}_{rst;}\\ &= \mathfrak{g}_{rs;t}[3] + 2\mathfrak{g}_{rst;} \,.\end{aligned} \tag{5.8}$$

From (2.10), (2.11) and (2.4), we have

$$\mathfrak{g}_{rst;} + \mathfrak{g}_{rs;t}[3] + T_{rst} = 0 \tag{5.9}$$

and

$$\mathfrak{g}_{;rst} + \mathfrak{g}_{r;st}[3] - T_{rst} = 0$$

and so

$$\mathfrak{g}_{rst;} - \mathfrak{g}^{*}_{rst;} = T_{rst} \,. \tag{5.10}$$

Combining (5.8) with (5.9) and (5.10) gives

$$\breve{\mathfrak{g}}_{rst;} = \mathfrak{g}^{*}_{rst;} = \mathfrak{g}_{rst;} - T_{rst} \,. \tag{5.11}$$

Differentiation of (5.5), (5.6) and (5.7) gives

$$\breve{\mathfrak{g}}_{rst;} + \breve{\mathfrak{g}}_{rs;t} = \mathfrak{g}_{rst;} + \mathfrak{g}_{rs;t} \,, \tag{5.12}$$

$$\breve{\mathfrak{g}}_{rt;s} + \breve{\mathfrak{g}}_{r;st} = \mathfrak{g}_{rt;s} + \mathfrak{g}_{r;st} \tag{5.13}$$

and

$$\breve{\mathfrak{g}}_{t;rs} + \breve{\mathfrak{g}}_{;rst} = \mathfrak{g}_{t;rs} + \mathfrak{g}_{;rst} \,. \tag{5.14}$$

Combining (5.12), (5.13) and (5.14) with (5.11) yields

$$\breve{\mathfrak{g}}_{rs;t} = \mathfrak{g}_{rs;t} + T_{rst}$$
$$\breve{\mathfrak{g}}_{r;st} = \mathfrak{g}_{r;st} - T_{rst}$$
$$\breve{\mathfrak{g}}_{;rst} = \mathfrak{g}_{;rst} + T_{rst}\,.$$

Together with the derivatives of (5.6) and (5.7), these give (iii). Equations (iv) and (v) now follow from (iii) and (2.3).

Evaluation of the third derivative of (5.4) at $\omega = \omega'$ gives

$$\breve{\mathfrak{g}}_{rstu;} = 4\mathfrak{g}_{rstu;} + 2\mathfrak{g}_{rst;u}[4] - \mathfrak{g}_{rs;j}\mathfrak{g}^{i;j}(\mathfrak{g}_{it;u}[2] + \mathfrak{g}_{itu;})[6] - \mathfrak{g}_{rstu;}\,. \quad (5.15)$$

Using (2.12) and (2.13) in (5.15), we obtain

$$\breve{\mathfrak{g}}_{rstu;} - \mathfrak{g}^*_{rstu;} = -\mathfrak{T}_{rs;tu}[6]\,.$$

The other equations in (vi) follow on differentiating the left-hand equations in (iii). □

Remark 5.3.— If construction (5.3) is applied to a yoke g which is not normalised then, in general, the resulting function $\breve{g}$ is not a yoke.

Remark 5.4.— Theorem 5.1 shows that $\breve{g}$ is "3^{rd} order close" to g^* on the diagonal, i.e., they have the same 3-jet there. This closeness can also be measured by the derivative $\mathrm{d}\breve{g}(X_{g^*})$ of $\breve{g}$ along the vector field X_{g^*} defined as in (4.1). Because $\breve{g}$ and g^* have the same 3-jet on the diagonal, it follows from Proposition 4.1 and Theorem 5.1 that the 3-jet on the diagonal of $\mathrm{d}\breve{g}(X_{g^*})$ is zero, i.e., if we put $h = \mathrm{d}\breve{g}(X_{g^*})$ then

(o) $\mathfrak{h} = 0$;

(i) $\mathfrak{h}_{r;} = \mathfrak{h}_{;r} = 0$;

(ii) $\mathfrak{h}_{rs;} = \mathfrak{h}_{r;s} = \mathfrak{h}_{;rs} = 0$;

(iii) $\mathfrak{h}_{rst;} = \mathfrak{h}_{rs;t} = \mathfrak{h}_{r;st} = \mathfrak{h}_{;rst} = 0$.

Furthermore,

$$\mathfrak{h}_{rstu;} = -\mathfrak{h}_{rst;u} = \mathfrak{h}_{rs;tu} = -\mathfrak{h}_{r;stu} = \mathfrak{h}_{;rstu} = -8\mathfrak{T}_{rs;tu}[6]\,.$$

Remark 5.5.— The expression (5.3) for $\breve{g}$ suggests the variants $\breve{g}^+$ and $\breve{g}^-$ of $\breve{g}$ defined by

$$\breve{g}^+(\omega, \omega') = g^*_{;j}(\omega, \omega') g^{i;j}(\omega, \omega') g_{i;}(\omega, \omega') - g(\omega, \omega')$$

and

$$\breve{g}^-(\omega, \omega') = g_{;j}(\omega, \omega') g^{i;j}(\omega, \omega') g^*_{i;}(\omega, \omega') - g(\omega, \omega'),$$

where g^* is the dual yoke of g.

Calculations similar to those in Theorem 5.1 show that, if g is a normalised yoke, then $\breve{g}^+$ and $\breve{g}^-$ are normalised yokes with the same metric as g and

$$\breve{\mathfrak{g}}^+_{rst;} = \breve{\mathfrak{g}}^-_{rst;} = \mathfrak{g}_{rst;} - 4T_{rst}\,.$$

Thus $\breve{g}^+$ and $\breve{g}^-$ are "further away" from g than $\breve{g}$ is.

The difference between $\breve{g}^+$ and $\breve{g}^-$ can be expressed in terms of the Poisson bracket (4.3) as

$$\breve{g}^- - \breve{g}^+ = \{g^*, g\}\,.$$

Remark 5.6.— A simple calculation shows that the operations of dualisation and taking the Lagrangian of a yoke commute. i.e.,

$$(g^*)\breve{} = (\breve{g})^*\,.$$

6. Group actions and momentum maps

In both mechanics and statistics, important simplifications occur in the presence of the extra structure provided by a group action. In mechanics a group action represents the symmetries of a Hamiltonian system; in statistics a group action on a parametric statistical model represents the symmetries of the model, which is called a (composite) transformation model. In both mechanics and statistics the presence of such a group action can be used to great advantage in simplifying calculations. In mechanics the principal construction associated to a group action is that of a momentum map from the cotangent bundle to the dual of the Lie algebra of the group. Momentum maps provide invariants of the system under the flow given by the Hamiltonian. We now consider the construction of the momentum map for a yoke.

Let G be a group acting on the manifold M, let LG denote the Lie algebra of G and let LG^* denote the dual of LG. Given a yoke g on M, we can use $\mathrm{d}_1 g : M \times M \to T^*M$ to pull the momentum map (in the sense of mechanics) from T^*M back to $M \times M$. For $X \in LG$, consider a path β in G with $\beta(0)$ the identity element of G and with $\beta'(0) = X$. The *momentum map* of the action is

$$J : M \times M \longrightarrow LG^*$$

given by

$$J(\omega, \omega')(X) = \frac{\mathrm{d}}{\mathrm{d}t}\, g\big(\beta(t)\omega, \omega'\big)\big|_{t=0} \,.$$

Thus

$$J = \psi \circ \mathrm{d}_1 g \,,$$

where $\psi : T^*M \to LG^*$ is defined by

$$\psi(\alpha)(X) = \alpha\left(\frac{\mathrm{d}}{\mathrm{d}t}\, \beta(t)\omega\big|_{t=0}\right)$$

for α in $T^*_\omega M$ and X, β are as above.

An alternative expression for J is

$$J(\omega, \omega') = \mathrm{d}_1 g(\omega, \omega') \circ a_\omega \,,$$

where

$$a_\omega(X) = \frac{\mathrm{d}}{\mathrm{d}t}\, \big(\beta(t)\omega\big)\big|_{t=0}$$

with β as above. In coordinate terms J is given by

$$J(\omega, \omega')\left(\frac{\partial}{\partial\theta^i}\right) = g_{j;}(\omega, \omega')\frac{\partial \omega^j}{\partial \theta^i}\bigg|_{\theta=0} \,,$$

where $\theta^1, \ldots, \theta^n$ are coordinates on G around the identity.

Note that if g is the observed likelihood yoke (2.6) of a parametric statistical model then $\mathrm{d}_1 g$ is essentially the score and so J is given by taking the score on 1-parameter subfamilies which arise through 1-parameter families of transformations. This is reminiscent of the definition of score in semi-parametric and non-parametric models.

Recall that G acts on LG by the adjoint action, in which γ in G takes X in LG to

$$\gamma \cdot X = \frac{\mathrm{d}}{\mathrm{d}t} \left(\gamma\beta(t)\gamma^{-1}\right)\Big|_{t=0} ,$$

where β is as above. The dual action of G on LG^* is the coadjoint action. defined by

$$(\gamma \cdot \alpha)(X) = \alpha(\gamma^{-1} \cdot X)$$

for α in LG^* and X in LG. If the yoke g is G-invariant (as is the case for the expected or observed likelihood yoke of a composite transformation model) then the momentum map is G-equivariant with respect to the product action on $M \times M$ and the coadjoint action on LG^*. In this case, the neighbourhood W of Δ_M in $M \times M$ on which η is defined can be chosen to be G-invariant. Then the symplectic form η and the vector field X_g are also g-invariant. Further, by Noether's Theorem (Marsden [19, p. 34]), the momentum J is preserved under the flow of X_g.

Let μ be an element of LG^* and denote by G_μ the isotropy group of μ, i.e., $G_\mu = \{\gamma \in G \mid \gamma \cdot \mu = \mu\}$. Suppose that μ is a regular value of J (i.e., the derivative of J at (ω, ω') is surjective at all points of $J^{-1}(\mu)$) and that G_μ acts freely and properly on $J^{-1}(\mu)$. Then $J^{-1}(\mu)$ is a submanifold of $M \times M$ and the reduced phase space $J^{-1}(\mu)/G_\mu$ is a manifold. Let $i_\mu : J^{-1}(\mu) \to M \times M$ denote the inclusion and $\pi_\mu : J^{-1}(\mu) \to J^{-1}(\mu)/G_\mu$ denote the projection. It is a simple consequence of the Symplectic Reduction Theorem (Abraham and Marsden [1, p. 299], Marsden [19, p. 36]) that there is a unique symplectic structure ζ on $J^{-1}(\mu)/G_\mu$ such that

$$i_\mu^* \eta = \pi_\mu^* \zeta . \tag{6.1}$$

This says that the (possibly singular) 2-form $i_\mu^* \eta$ obtained by restricting the symplectic form η to the G_μ-invariant set $J^{-1}(\mu)$ can be obtained by pulling back a (unique) symplectic form ζ on the corresponding G_μ-orbit space $J^{-1}(\mu)/G_\mu$.

Example 6.1

The bivariate normal distributions with unknown mean and covariance matrix equal to the unit matrix form a composite transformation model with parameter space $M = \mathbb{R}^2$ and group $G = \mathrm{SO}(2)$ with the usual action of the rotation group SO(2) on $\mathbb{R}^2$. The expected (or observed) likelihood yoke is

$$g(\omega, \omega') = -\frac{1}{2} \left\|\omega - \omega'\right\|^2 , \quad \omega,\, \omega' \in \mathbb{R}^2 .$$

It is useful to identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$, writing

$$\omega = z = x + iy\,, \quad \omega' = w = u + iv\,,$$

and to identify SO(2) with $U(1)$, writing a typical element of SO(2) as $e^{i\theta}$. Then

$$g(z, w) = -\frac{1}{2}\left|z - w\right|^2$$

and $e^{i\theta}$ takes (z, w) to $(e^{i\theta}z, e^{i\theta}w)$. Clearly, g is G-invariant. The symplectic form η of g is

$$\eta = \mathrm{d}x \wedge \mathrm{d}u + \mathrm{d}y \wedge \mathrm{d}v\,.$$

The mapping

$$\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \longmapsto a$$

identifies $\mathrm{L}G = \mathrm{so}(2)$ with $\mathbb{R}$ and so $\mathrm{L}G^*$ with $\mathbb{R}^* = \mathbb{R}$. With this identification, the momentum mapping J is

$$J(z, w) = \mathrm{Im}(z\overline{w}) = -xv + yu\,.$$

Since G is abelian, $G_\mu = G$ for all μ in $\mathbb{R}$. Note that the function

$$(z, w) \longmapsto \left|z\right|^{-1}\overline{z}w$$

is G-invariant and that its restriction to $J^{-1}(\mu)$ can be considered as the projection

$$\pi_\mu : J^{-1}(\mu) \longrightarrow J^{-1}(\mu)/G_\mu$$

for any μ in $\mathbb{R}$. Put

$$\left|z\right|^{-1}\overline{z}w = s + it\,.$$

Then a calculation shows that, for $\mu \neq 0$, the symplectic form

$$\zeta = -\frac{\mu}{t^2}\,\mathrm{d}s \wedge \mathrm{d}t$$

on $J^{-1}(\mu)/G_\mu$ satisfies (6.1).

Example 6.2

A model analogous to that considered in Example 6.1 is given by the hyperbola distributions (Barndorff-Nielsen [5]) on the branch $H = \{(x,y) \mid x^2 - y^2 = 1\,,\ x > 0\}$ of the unit hyperbola in $\mathbb{R}^2$. The parameter space is $M = (0,\infty) \times \mathbb{R}$ and the probability density functions have the form

$$f(z;\omega) = \exp\{-\omega * z - a(\omega)\}\,,$$

where

$$\omega * z = xp - yq$$

with

$$z = (p,q)\,, \quad \omega = (x,y)$$

and $a(\omega)$ is an appropriate function of ω. The group $G = \mathbb{R}$ acts on M and on H via the homomorphism

$$\chi \longmapsto \begin{pmatrix} \cosh\chi & \sinh\chi \\ \sinh\chi & \cosh\chi \end{pmatrix}$$

from $\mathbb{R}$ into GL(2). Under this action the hyperbola distributions form a composite transformation model. By restricting the concentration $\sqrt{(\omega * \omega)}$ of ω to be equal to a given positive constant κ, we obtain a submodel with parameter space equivalent to H. The expected (or observed) likelihood yoke is

$$g(\omega,\omega') = c(\kappa)(\omega - \omega') * \omega'\,,$$

where $c(\kappa)$ is an appropriate function of κ. Clearly, g is G-invariant. The symplectic form η of g is

$$\eta = c(\kappa)(\mathrm{d}x \wedge \mathrm{d}u - \mathrm{d}y \wedge \mathrm{d}v)\,,$$

where

$$\omega' = (u,v)\,.$$

The momentum mapping J is

$$J(\omega,\omega') = c(\kappa)(-xv + yu)\,.$$

Since G is abelian, $G_\mu = G$ for all μ in $\mathbb{R}$. Put

$$s = \frac{xu - yv}{\sqrt{x^2 - y^2}}\,, \quad t = \frac{xv - yu}{\sqrt{x^2 - y^2}}\,.$$

Then the function

$$(\omega, \omega') \longmapsto (s, t)$$

is G-invariant and its restriction to $J^{-1}(\mu)$ can be considered as the projection

$$\pi_\mu : J^{-1}(\mu) \longrightarrow J^{-1}(\mu)/G_\mu$$

for any μ in $\mathbb{R}$. A calculation shows that, for $\mu \neq 0$, the symplectic form

$$\zeta = -\frac{\mu}{t^2}\, \mathrm{d}s \wedge \mathrm{d}t$$

on $J^{-1}(\mu)/G_\mu$ satisfies (6.1).

Example 6.3

Now consider the hyperboloid distributions (Barndorff-Nielsen [5], Jensen [18]) on the branch $H_2 = \{(x, y, z) \mid x^2 - y^2 - z^2 = 1\,,\ x > 0\}$ of the unit hyperboloid in $\mathbb{R}^3$. The model function is

$$f(w; \omega) = \exp\{-\omega * w - a(\omega)\}\,,$$

where

$$\omega * w = \omega_1 x - \omega_2 y - \omega_3 z$$

for

$$\omega = (\omega_1, \omega_2, w_3)\,, \quad w = (x, y, z)$$

and $a(\omega)$ is an appropriate function of ω. By restricting the concentration $\sqrt{(\omega * \omega)}$ of ω to be equal to a given positive constant κ, we obtain a submodel with parameter space equivalent to H_2. The group $G = \mathbb{R}$ acts on H_2 via the homomorphism

$$\chi \longmapsto \begin{pmatrix} \cosh\chi & \sinh\chi & 0 \\ \sinh\chi & \cosh\chi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

from $\mathbb{R}$ into GL(3). Under this action the submodel forms a composite transformation model. The expected likelihood yoke is

$$g(\omega, \omega') = c(\kappa)(\omega - \omega') * \omega'\,,$$

where $c(\kappa)$ is an appropriate function of κ. Because the product $*$ and the function c are G-invariant, so is the yoke g. The symplectic form η of g is

$$\eta = c(\kappa)(\mathrm{d}\omega_1 \wedge \mathrm{d}\omega_1' - \mathrm{d}\omega_2 \wedge \mathrm{d}\omega_2' - \mathrm{d}\omega_3 \wedge \mathrm{d}\omega_3'),$$

where

$$\omega' = (\omega_1', \omega_2', \omega_3').$$

The momentum mapping J is

$$J(\omega, \omega') = c(\kappa)(\omega_2\omega_1' - \omega_1\omega_2').$$

Since G is abelian, $G_\mu = G$ for all μ in $\mathbb{R}$. The function

$$(\omega, \omega') \longmapsto (\omega_3, \omega_3')$$

is G-invariant and its restriction to $J^{-1}(\mu)$ can be considered as the projection

$$\pi_\mu : J^{-1}(\mu) \longrightarrow J^{-1}(\mu)/G_\mu$$

for any μ in $\mathbb{R}$. A calculation shows that the symplectic form

$$\zeta = -c(\kappa)\frac{\sqrt{(1+\omega_3^2)(1+\omega_3'^2) + \big(\mu/c(\kappa)\big)^2} - \omega_3\omega_3'}{\sqrt{(1+\omega_3^2)(1+\omega_3'^2) + \big(\mu/c(\kappa)\big)^2}}\,\mathrm{d}\omega_3 \wedge \mathrm{d}\omega_3'$$

on $J^{-1}(\mu)/G_\mu$ satisfies (6.1).

Acknowledgement

This research was supported in part by a twinning grant in the European Community Science programme and by a network contract in the European Union HCM programme.

References

[1] Abraham (R.) and Marsden (J. E.) .— *Foundations of Mechanics*, 2nd ed., Addison-Wesley, Redwood City (1978).

[2] Amari (S.-I.) .— *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Springer-Verlag, Heidelberg, **28** (1985).

[3] ARNOL'D (V. I.) and GIVENTAL' (A. B.) .— *Symplectic Geometry*, In "Dynamical Systems IV: Symplectic Geometry and its Applications", Encyclopaedia of Mathematical Sciences (V. I. Arnol'd and S. P. Novikov, eds), Springer-Verlag, Berlin, **4** (1990), pp. 1-136.

[4] BALIAN (R.) .— *From Microphysics, to Macrophysics*, Springer-Verlag, Berlin, **1** (1991).

[5] BARNDORFF-NIELSEN (O. E.) .— *Hyperbolic distributions and distributions on hyperbolae*, Scand. J. Statist. **5** (1978), pp. 151-157.

[6] BARNDORFF-NIELSEN (O. E.) .— *Likelihood and observed geometries*, Ann. Statist. **14** (1986), pp. 856-873.

[7] BARNDORFF-NIELSEN (O. E.) .— *Differential geometry and statistics: some mathematical aspects*, Indian J. Math. **29** (1987), pp. 335-350.

[8] BARNDORFF-NIELSEN (O. E.) .— *Parametric Statistical Models and Likelihood*, Lecture Notes in Statistics, Springer-Verlag, Heidelberg, **50** (1988).

[9] BARNDORFF-NIELSEN (O. E.) and COX (D. R.) .— *Inference and Asymptotics*, Chapman & Hall, London (1994).

[10] BARNDORFF-NIELSEN (O. E.) and JUPP (P. E.) .— *Differential geometry, profile likelihood, L-sufficiency and composite transformation models*, Ann. Statist. **16** (1988), pp. 1009-1043.

[11] BARNDORFF-NIELSEN (O. E.) and JUPP (P. E.) .— *Yokes and symplectic structures*, J. Statist. Planning and Infce. **63** (1997), pp. 133-146.

[12] BLÆSILD (P.) .— *Yokes and tensors derived from yokes*, Ann. Inst. Statist. Math. **43** (1991), pp. 95-113.

[13] COMBET (E.) .— *Intégrales Exponentielles*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, **937** (1982).

[14] CRITCHLEY (F.), MARRIOTT (P. K.) and SALMON (M.) .— *Preferred point geometry and statistical manifolds*, Ann. Statist. **21** (1993), 1197-1224.

[15] CRITCHLEY (F.), MARRIOTT (P. K.) and SALMON (M.) .— *Preferred point geometry and the local differential geometry of the Kullback-Leibler divergence*, Ann. Statist. **22** (1994), pp. 1587-1602.

[16] EGUCHI (S.) .— *Second order efficiency of minimum contrast estimation in a curved exponential family*, Ann. Statist. **11** (1983), pp. 793-803.

[17] FRIEDRICH (T.) .— *Die Fisher-Information und symplectische Strukturen*, Math. Nachr. **153** (1991), pp. 273-296.

[18] JENSEN (J. L.) .— *On the hyperboloid distribution*, Scand. J. Statist. **8** (1981), pp. 193-206.

[19] MARSDEN (J. E.) .— *Lectures on Mechanics*, London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, **174** (1992).

[20] MCDUFF (D.) .— *Examples of symplectic structures*, Invent. Math. **89** (1987), pp. 13-36.

[21] MURRAY (M. K) and RICE (J. W.) .— *Differential Geometry and Statistics*, Chapman & Hall, London (1993).

[22] NAKAMURA (Y.) .— *Completely integrable gradient systems on the manifolds of Gaussian and multinomial distributions*, Japan J. Industr. Appl. Math. **10** (1993), pp. 179-189.

[23] NAKAMURA (Y.) .— *Gradient systems associated with probability distributions*, Japan J. Industr. Appl. Math. **11** (1994), pp. 21-30.

[24] SKOVGAARD (I. M.) .— *On the density of minimum contrast estimators*, Ann. Statist. **18** (1990), pp. 779-789.

[25] WEINSTEIN (A.) .— *Symplectic manifolds and their Lagrangian submanifolds*, Adv. Math. **6** (1971), pp. 329-346.

[26] WEINSTEIN (A.) .— *Lectures on Symplectic Manifolds*, AMS Regional Conference Series in Mathematics, American Mathematical Society, Providence, Rhode Island, **29** (1977).