The phylon group and statistics

A.L. Carey, P.E. Jupp, M.K. Murray

1. Introduction

Statisticians are concerned with the question of the extent to which statistical inference is independent of the choice of parameters for a family of probability distributions. For example, of the three well-known forms of general hypothesis test, the likelihood ratio test and the score test are parameterisation-invariant but the Wald test is not. (See e.g., Cox and Hinkley, 1974, pp. 322–323.) It has been clear to statisticians for some time that differential geometry provides a convenient language for studying questions of parameterisation-independence and a fruitful interaction has arisen between the two subjects. In particular, statisticians have developed a theory of 'higher-order' tensors or 'strings' to study these questions. (See Barndorff-Nielsen *et al.*, 1994, for a review of string theory.) In Carey and Murray (1990) two of the present authors began to develop a mathematical framework for 'strings' involving an infinite-dimensional group. This group, later called the phylon group $\mathcal{P}(d)$, plays the same role in the theory of strings as $GL(d)$, the general linear group, plays in the theory of tensors. The theory of the phylon group and its representations was further developed in Barndorff-Nielsen *et al.* (1992). That work concentrated on the finite-dimensional representation theory of the phylon group, utilising in part results of Terng (1978). In the work presented here we consider two related extensions of this work: (i) twisted phylon representations, and (ii) infinite-dimensional phylon representations.

Twisted phylon representations arise naturally in statistical asymptotics and shed light on the relationship between the two senses in which 'order' is used in parametric statistics. These are: order in the sense of the order of a term in a Taylor expansion in some parameters, and order in the sense of an asymptotic expansion in powers of the sample size.

We shall see that twisted phylon representations are by their nature infinite-dimensional, so we are naturally led to study representations of the phylon group that are infinite-dimensional. In Barndorff-Nielsen *et al.* (1992) we defined a certain class of finite-dimensional representation of the phylon group which we called phylon representations. The statistical context and the fact that the phylon group has a Fréchet Lie group structure suggests that the appropriate class of infinite-dimensional phylon representations will have to be a subset of those representations arising from continuous actions on Fréchet spaces. We define this subset as follows. The phylon group $\mathcal{P}(d)$ contains a one-dimensional subgroup of dilations of $\mathbf{R}^d$, that is, maps of the form $x \mapsto \lambda x$ where λ is a non-zero real number. Following Terng (1978), we say that a vector v in a representation of the phylon group has homogeneous degree n if when we act on it by the dilation λ it is multiplied by λ^n. The subspace of all vectors of homogeneous degree n is called the homogeneous subspace of degree n. A representation arising from a continuous action of the phylon group on a Fréchet space is called a phylon representation if it decomposes into a direct product of finite-dimensional homogeneous subspaces whose degrees are bounded above. In Section 4 we prove that any phylon representation is a projective limit of finite-dimensional representations of finite-dimensional quotients $\mathcal{P}_k(d)$ of the phylon group $\mathcal{P}(d)$. The complete theory of the finite-dimensional algebraic representations of the finite-dimensional phylon groups $\mathcal{P}_k(d)$ has already been developed by Terng (1978). The primary example of an infinite-dimensional phylon representation is the adjoint representation of the phylon group $\mathcal{P}(d)$ on its Lie algebra. We shall see that the coadjoint action is not a phylon representation in our sense, illustrating the fact that infinite-dimensional strings and costrings are different.

2. Fréchet vector spaces

A Fréchet space is a complete, locally convex, metrisable, topological vector space. We shall be interested in a particular kind of Fréchet space, so we shall consider these rather than work with the full theory. Readers interested in the general framework should consult the book by Treves (1967).

Recall that if we have a countable collection $\{V_i \mid i = 1, 2, \dots\}$ of finite-dimensional vector spaces then we distinguish between the direct product

$$\prod_{i=1}^{\infty} V_i \tag{2.1}$$

and the direct sum

$$\sum_{i=1}^{\infty} V_i, \tag{2.2}$$

where the latter is the subset of the direct product consisting of all sequences $(v_1, v_2, \dots)$ with only a finite number of non-zero components. The direct product can be made into a Fréchet space by giving it the product topology, i.e. the topology of 'pointwise convergence'. Thus a sequence converges in this topology if and only if each of the sequences of components converges. Notice that there is no assumption of uniformity on the convergence of the components so, for example, $(a_1, 0, 0, \dots), (0, a_2, 0, \dots), (0, 0, a_3, \dots), \dots$ converges to $0 = (0, 0, 0, \dots)$ for *any* sequence $a_1, a_2, a_3, \dots$ of vectors. If $\sum_{i=1}^{\infty} V_i$ (with the product topology) were a Banach space then we could choose $a_1, a_2, a_3, \dots$ to get a sequence of elements of $\sum_{i=1}^{\infty} V_i$ with unit norm but converging to 0. Since this is impossible, $\sum_{i=1}^{\infty} V_i$ is not a Banach space.

If W_i is a subset of V_i let us denote by $W = W_1 \times W_2 \times W_3 \dots$ the subset of $\prod_{i=1}^{\infty} V_i$ defined by

$$W = \{(w_1, w_2, \dots) | w_i \in W_i \ \forall i = 1, 2, \dots\}. \tag{2.3}$$

A basis of neighbourhoods of zero for the topology of pointwise convergence is given by sets of the form $U_1 \times U_2 \times U_3 \times \ldots$, where each of the U_i is an open neighbourhood of 0 in V_i and only a *finite number* of the U_i are not equal to V_i.

The direct sum $\sum_{i=1}^{\infty} V_i$ inherits a topology as a subspace of the direct product but is not complete. For this and other reasons the subspace topology is not the natural topology on $\sum_{i=1}^{\infty} V_i$ and instead we define the topology which has as a basis of neighbourhoods of the identity all sets of the form

$$(U_1 \times U_2 \times U_3 \times \ldots) \cap \sum_{i=1}^{\infty} V_i, \tag{2.4}$$

where each of the U_i is open in V_i. This is also the topology obtained by saying that a set is open if and only if its intersection with each of the subsets

$$V_1 \oplus 0 \oplus 0 \oplus \ldots \subset V_1 \oplus V_2 \oplus 0 \oplus \ldots \subset V_1 \oplus V_2 \oplus V_3 \oplus \ldots \subset \ldots \tag{2.5}$$

is open. As such, the direct sum of the V_i is the inductive limit in the category of topological vector spaces of a sequence of Fréchet spaces (Treves, 1967, p. 126).

The relationship between the direct sum and the direct product is best understood as one of 'duality'. More precisely, if $V = \prod V_i$ with the Fréchet topology just defined then the dual space V^* of continuous linear functionals is the direct sum $\sum V_i^*$. If the dual space is given the strong topology then this is an isomorphism of topological vector spaces. For example, let $\mathbf{R}[[x^1, \ldots, x^d]]$ be the space of all formal power series in d variables. Then this is a direct product of the kind we have been discussing if we let V_i be the space of all homogeneous polynomials of degree i. That is, $V_i = S^i(\mathbf{R}^d)^*$, the ith symmetric product of the dual of $\mathbf{R}^d$. The dual of $\mathbf{R}[[x^1, \ldots, x^d]]$ is the space of all polynomials, $\mathbf{R}[x^1, \ldots x^d]$, that is, formal power series of finite order. This is essentially shown in Treves (1967, pp. 227–228) and that proof can be adapted to the case at hand.

To understand this duality more geometrically, recall that the space of formal power series, $\mathbf{R}[[x^1, \ldots, x^d]]$, can be identified with $J^\infty(\mathbf{R}^d)_0$, the space of all infinite jets at the origin of smooth, real-valued functions on $\mathbf{R}^d$. By definition, the latter space is the quotient of the space $\mathbf{C}^\infty(\mathbf{R}^d)$ of all smooth functions by the ideal of functions which have all their partial derivatives at the origin equal to zero. Then, by a theorem of Borel (Treves, 1967, p. 390), the map

$$C^\infty(\mathbf{R}^d) \to \mathbf{R}[[x^1, \ldots, x^d]] \tag{2.6}$$

which sends a smooth function to its Taylor expansion at the origin is a surjection. This surjection is continuous if we give $C^\infty(\mathbf{R}^d)$ the smooth

(Fréchet) topology (Treves, 1967, p. 86) and hence induces an isomorphism from $J^\infty(\mathbf{R}^d)_0 \to \mathbf{R}[[x^1, \ldots, x^d]]$. Corresponding to the surjection in (2.6) there is an inclusion of the dual spaces. The dual of the space of all smooth functions with the smooth topology is just the space of all distributions $\mathcal{D}(\mathbf{R}^d)$. The dual of $J^\infty(\mathbf{R}^d)_0$ is the subspace of distributions that have support at the origin. As is well-known (Treves, 1967, p. 266), a distribution with support at the origin is a *finite* sum of derivatives of the Dirac delta distribution δ. This space is isomorphic to the space of polynomials. Indeed, we can map the polynomial p to the distribution $p(\partial/\partial x^1, \ldots, \partial/\partial x^d)\delta$. Notice that this isomorphism depends on choosing coordinates to define the partial derivative operators.

In the next section we study the phylon group as a Fréchet Lie group and, in particular, as a Fréchet manifold. The theory of finite-dimensional manifolds extends to manifolds modelled on any topological vector space. Manifolds modelled on Banach or Hilbert spaces behave in a very similar way to finite-dimensional manifolds. Manifolds modelled on Fréchet spaces, however, behave differently because two important theorems of multivariable calculus, the uniqueness of solutions to ordinary differential equations and the inverse function theorem, extend to Banach and Hilbert spaces but not to Fréchet spaces. The geometric consequence of this for the theory of Fréchet Lie groups is that there may not be an exponential map from the Lie algebra to the group and even if there is it need not be a local diffeomorphism. Useful references for the theory of Fréchet manifolds and Fréchet Lie groups are the articles by Hamilton (1982), Kobayashi *et al.* (1985) and Milnor (1984).

3. The infinite-dimensional phylon group

The infinite-dimensional phylon group $\mathcal{P}(d)$ is defined as either the set of infinite jets of germs of local diffeomorphisms of $\mathbf{R}^d$ which fix the origin or, equivalently, as all $\mathbf{R}^d$-valued formal power series in d variables with no constant term and invertible linear term. It is clear from these definitions that we can place on $\mathcal{P}(d)$ the topology of pointwise convergence of the coefficients of the power series. Inside the infinite phylon group we have a sequence of normal subgroups $\mathcal{P}^{(k)}(d)$, which are the sets of jets of diffeomorphisms that agree with the identity up to order k. The quotient groups $\mathcal{P}_k(d)$ are the groups of k-jets of local diffeomorphisms of $\mathbf{R}^d$ which fix the origin. For every k there is a projection map $\mathcal{P}(d) \to \mathcal{P}_k(d)$, which we shall denote by j^k. Moreover, for any $m \geq k$, there is a projection $\mathcal{P}_m(d) \to \mathcal{P}_k(d)$, which we shall also denote by j^k. The groups $\mathcal{P}_k(d)$ are finite-dimensional Lie groups. If $\mathcal{P}(d)$ is given the topology of pointwise convergence then each of the maps j^k is continuous and this topology is, in fact, the projective limit topology. That is, it is the coarsest topology for which each of the projections j^k is continuous.

With this topology the phylon group is a Fréchet Lie group. Indeed, we can say more than this; Omori (1980) shows that the phylon group is a regular Fréchet Lie group. The precise definition of a regular Fréchet Lie group is a little complicated and can be found, along with a survey of their properties, in Kobayashi *et al.* (1985). It suffices for our purposes to note that regular Fréchet Lie groups are a class of Fréchet Lie groups for which the exponential map from the Lie algebra to the group exists and is smooth. Recall (for example, from Kolář *et al.*, 1993, p. 36) how we construct the exponential map $\exp : LG \to G$ for a finite dimensional Lie group G. Given an element X of the Lie algebra, we consider the integral curve γ of the left invariant vector field defined by X for which $\gamma(0)$ is the identity. The exponential map is then defined by $\exp(X) = \gamma(1)$ and the curve γ turns out to be a one-parameter subgroup, that is, $\gamma(t)\gamma(s) = \gamma(s+t)$, and satisfies $\gamma(t) = \exp(tX)$. The integral curve γ can be characterised by the fact that it is the unique one-parameter subgroup whose tangent at zero is X. As we have remarked at the end of Section 2, it is not always possible, even locally, to integrate vector fields on a Fréchet manifold, so this construction need not work for a Fréchet Lie group. Whether or not the exponential map exists for an arbitrary Fréchet Lie group is an open question.

The Lie algebra of $\mathcal{P}(d)$ is the tangent space at the identity and this is just the space of jets at the origin of vector fields on $\mathbf{R}^d$ which are zero at the origin. We denote this space by $L\mathcal{P}(d)$. An element X of $L\mathcal{P}(d)$ can be written as

$$X = \sum_i X^i \frac{\partial}{\partial x^i}, \tag{3.1}$$

where X^i is an element of $\mathbf{R}[[x^1, \ldots, x^d]]$ with constant term zero. The Lie bracket of any two elements X and Y of $L\mathcal{P}(d)$ is

$$[X, Y] = \sum_{i,j} \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i} \right) \frac{\partial}{\partial x^j}. \tag{3.2}$$

On a finite-dimensional Lie group we can prove that the exponential map is a diffeomorphism of a neighbourhood of the identity by using the inverse function theorem. The derivative of the exponential map is easily seen to be invertible. However, as we remarked at the end of Section 2, the inverse function theorem does not apply to Fréchet spaces and the exponential map on a Fréchet Lie group need not be a local diffeomorphism. See Milnor (1984) for a discussion of a counter-example. In the case of the phylon group $\mathcal{P}(d)$, we can exploit the fact that it is a projective limit of finite-dimensional Lie groups to show that exponential map $L\mathcal{P}^{(1)}(d) \to \mathcal{P}^{(1)}(d)$ is a smooth bijection, as follows.

Notice that $\mathcal{P}(d)/\mathcal{P}^{(1)}(d)$ is the group $GL(d)$ and, in fact, $\mathcal{P}(d)$ is the semi-direct product of $\mathcal{P}^{(1)}(d)$ and $GL(d)$. If we denote by $\mathcal{P}_r^{(m)}(d)$ the kernel of $\mathcal{P}_r(d) \to \mathcal{P}_m(d)$ then it is also true that $\mathcal{P}_k(d)$ is the semi-direct product of

$\mathcal{P}_k^{(1)}(d)$ and $GL(d)$. It is pointed out by Kolář *et al.* (1993, p. 131) that the group $\mathcal{P}_k^{(1)}(d)$ is a connected, simply connected, nilpotent finite-dimensional Lie group and that therefore the exponential map $\exp\colon L\mathcal{P}_k^{(1)}(d) \to \mathcal{P}_k^{(1)}(d)$ is a diffeomorphism. In fact, the same is true of the groups $\mathcal{P}_k^{(r)}(d)$ for any $r \geq 1$, so the exponential maps $\exp\colon L\mathcal{P}_k^{(r)}(d) \to \mathcal{P}_k^{(r)}(d)$ are also diffeomorphisms. We shall use this fact to prove that the exponential map $L\mathcal{P}^{(r)}(d) \to \mathcal{P}^{(r)}(d)$ is a smooth bijection (for $r \geq 1$).

We have already noted that, because the phylon group $\mathcal{P}(d)$ is regular, it has a smooth exponential map. We have not yet considered the existence of the exponential map $L\mathcal{P}^{(r)}(d) \to \mathcal{P}^{(r)}(d)$. We would expect that it is the restriction of the exponential map on $L\mathcal{P}(d)$ and we show this as follows. Notice that the exponential map on $L\mathcal{P}(d)$ certainly restricts to a smooth map on $L\mathcal{P}^{(r)}(d)$. The question is: does its image lie in $\mathcal{P}^{(r)}(d)$? Let $X \in L\mathcal{P}^{(r)}(d)$ and consider $f(t) = \exp(tX)$. As we remarked earlier, this curve is a one-parameter subgroup whose tangent at zero is X. Because $j^r\colon \mathcal{P}(d) \to \mathcal{P}_r(d)$ is a group homomorphism, $g(t) = j^r(f(t))$ is also a one-parameter subgroup and if we differentiate it at $t = 0$ we see that the tangent is the zero vector. The point is that the tangent map to j^r going from $L\mathcal{P}(d)$ to $L\mathcal{P}^{(r)}(d)$ is also the corresponding jet map j^r and $j^r(X) = 0$ by assumption. Standard facts about one-parameter subgroups in a finite-dimensional Lie group (as, for example, in Kolář *et al.*, 1993, pp. 35–36) now apply and it follows that $g(t)$ is a constant function equal to the identity and hence $f(t)$ is in $\mathcal{P}^{(r)}(d)$ for all t, as required.

Because the map $j^k : \mathcal{P}^{(r)}(d) \to \mathcal{P}_k^{(r)}(d)$ is a group homomorphism for every k, its derivative is a homomorphism of the corresponding Lie algebras and this is, in fact, also the map j^k acting on jets of vector fields. Moreover, these two maps commute with the exponential map. Thus we have commutative diagrams

$$\begin{array}{ccc} L\mathcal{P}^{(r)}(d) & \xrightarrow{j^k} & L\mathcal{P}_k^{(r)}(d) \\ \exp\downarrow & & \exp_k\downarrow \\ \mathcal{P}^{(r)}(d) & \xrightarrow{j^k} & \mathcal{P}_k^{(r)}(d) \end{array} \tag{3.3}$$

where we temporarily adopt the notation $\exp_k$ for the exponential map on $L\mathcal{P}_k^{(r)}(d)$. Let g be an element of $\mathcal{P}^{(r)}(d)$. Then, for every k, $j^k(g) = \exp_k(X_k)$ for some (unique) X_k in $L\mathcal{P}_k^{(r)}(d)$. If $m \geq k$ then $j^m(g) = \exp_m(X_m)$ and hence

$$\begin{aligned} j^k(g) &= j^k(j^m(g)) \\ &= j^k(\exp_m(X_m)) \\ &= \exp_k(j^k(X_m)). \end{aligned} \tag{3.4}$$

Thus by the uniqueness of X_k we have that $X_k = j^k(X_m)$. Because $L\mathcal{P}^{(r)}(d)$ is the projective limit of the $L\mathcal{P}_k^{(r)}(d)$, it follows that there is a unique X

in the Lie algebra of $\mathcal{P}^{(r)}(d)$ such that $j^k(X) = X_k$ and that $j^k(\exp(X)) = j^k(g)$, so that $\exp(X) = g$. Similarly, it is possible to show that exp is one to one.

Hence we have proved:

Proposition 3.1 *The exponential map* $\exp\colon L\mathcal{P}^{(r)}(d) \to \mathcal{P}^{(r)}(d)$ *is a smooth bijection.*

We can now prove the result that we need for the next section. Assume that we have a representation of $\mathcal{P}(d)$ on a Fréchet space V and that the map $\mathcal{P}(d) \times V \to V$ is smooth. By differentiating, it follows that we have a representation of the Lie algebra and that the map $L\mathcal{P}(d) \times V \to V$ is smooth. Let W be a closed subspace of finite codimension in V. We want to show that (i) if $L\mathcal{P}^{(k)}(d)$ stabilises W then so also does $\mathcal{P}^{(k)}(d)$, (ii) if $L\mathcal{P}^{(k)}(d)(V) \subset W$ then $\mathcal{P}^{(k)}(d)$ acts trivially on V/W. We just use the exponential map as we would in the finite-dimensional case. For X in $L\mathcal{P}^{(k)}(d)$ consider the curve $g_t = \exp(tX)$ and apply it to v in V. Differentiating gives

$$\frac{d}{dt}(g_t.v) = X(g_t.v). \tag{3.5}$$

Let $\pi\colon V \to V/W$ be the projection. Then

$$\frac{d}{dt}\pi(g_t.v) = \pi X(g_t.v). \tag{3.6}$$

If either $L\mathcal{P}^{(k)}(d)$ stabilises W and $v \in W$, or $L\mathcal{P}^{(k)}(d)(V) \subset W$ then $\pi X(g_t.v) = 0$. Since $g_0 = 1$, uniqueness of solutions of first-order differential equations with values in V/W implies that $\pi(g_t.v) = \pi(v)$ for all t.

Hence we have proved:

Proposition 3.2 *Let W be a closed subspace of finite codimension in a vector space V on which the phylon group $\mathcal{P}(d)$ acts smoothly. Then, for* $k = 1, 2, \ldots$,
(i) if $L\mathcal{P}^{(k)}(d)$ stabilises W then so does $\mathcal{P}^{(k)}(d)$,
(ii) if $L\mathcal{P}^{(k)}(d)(V) \subset W$ then the induced action of $\mathcal{P}^{(k)}(d)$ on V/W is trivial.

4. Infinite-dimensional phyla

In Barndorff-Nielsen *et al.* (1992) we defined a phylon to be a certain type of finite-dimensional representation of the infinite-dimensional phylon group $\mathcal{P}(d)$. In this section we shall define a quite general class of infinite-dimensional representations V of $\mathcal{P}(d)$ to be phylon representations and show that they are actually projective limits of (finite-dimensional) representations V_k of $\mathcal{P}_k(d)$.

Before defining an infinite-dimensional phylon representation, let us note that the dilations of $\mathbf{R}^d$, i.e. the maps $x \mapsto \lambda x$ for λ a real non-zero number, form a subgroup of the phylon group isomorphic to the non-zero reals.

Indeed, the dilations are actually the subgroup of $GL(d)$ of all non-zero multiples of the identity. Thus if the phylon group acts linearly on a vector space V then the dilations also act. We say that an element $v \in V$ has homogeneous degree n if $\lambda.1.v = \lambda^n.v$ (Terng, 1978). Then we have:

Definition 4.1 *An (infinite-dimensional) phylon representation on a Fréchet space V is a group homomorphism*

$$\mathcal{P}(d) \to GL(V) \tag{4.1}$$

such that:
(i) $V = \prod_{i=1}^{\infty} V_{n_i}$,
(ii) V_{n_i} is the subspace of all elements of homogeneous degree n_i, and $n_1 > n_2 > n_3 > \ldots$,
(iii) for $i = 1, 2, \ldots$, V_{n_i} is a finite-dimensional $GL(d)$-module,
(iv) the map

$$\mathcal{P}(d) \times V \to V \tag{4.2}$$

is smooth as a map between Fréchet manifolds.

We call the decomposition in (i) the homogenous decomposition of V. An important example of a phylon representation is the adjoint action of the phylon group $\mathcal{P}(d)$ on the Lie algebra $L\mathcal{P}(d)$, in which $\mathcal{P}(d)$ acts on $L\mathcal{P}(d)$ by conjugation. More precisely, for g in $\mathcal{P}(d)$ and X in $L\mathcal{P}(d)$, the adjoint action is

$$(g, X) \mapsto gXg^{-1}.$$

In this case the homogeneous decomposition is

$$\prod_{k=1}^{\infty} [\mathbf{R}^d \otimes S^k(\mathbf{R}^d)^*]. \tag{4.3}$$

The homogeneous degrees are $0, -1, \ldots$ with $L\mathcal{P}(d)_{-k} = \mathbf{R}^d \otimes S^{k+1}(\mathbf{R}^d)^*$ for all k. The vector space structure of the subalgebra $L\mathcal{P}^{(k)}(d)$ of $L\mathcal{P}(d)$ is

$$L\mathcal{P}^{(k)}(d) = \prod_{j \geq k} L\mathcal{P}(d)_{-j}. \tag{4.4}$$

This example is important in understanding the structure of general phylon representations, as we shall now see.

Proposition 4.1 *Every phylon representation is a projective limit of finite-dimensional representations of the phylon group.*

The proof of this result is quite straightforward. Consider the dilation $\lambda.1$. As the non-zero real number λ varies, this defines a curve in the phylon group $\mathcal{P}(d)$ and we can differentiate it at $\lambda = 1$ to obtain the element

$$Z = \sum_i x^i \frac{\partial}{\partial x^i} \tag{4.5}$$

in the Lie algebra $L\mathcal{P}(d)$.

Let v be an element of homogeneous degree n in a phylon representation. Then by definition we have

$$\lambda.v = \lambda^n v, \tag{4.6}$$

where we have written a dot on the left hand side to remind the reader that here λ is acting as an element of the phylon group, whereas on the right hand side λ^n is acting by scalar multiplication. Differentiating shows us that

$$Zv = nv. \tag{4.7}$$

In particular, if X is an element of $L\mathcal{P}(d)$ of homogeneous degree $-k$ then we have $[Z, X] = -kX$. Putting these facts together, we can show that Xv has a homogeneous degree and calculate it from

$$\begin{aligned} Z(Xv) &= [Z, X]v + XZv \\ &= -kXv + nXv \\ &= (n-k)Xv. \end{aligned} \tag{4.8}$$

In other words,

$$L\mathcal{P}(d)_{-k} V_n \subset V_{n-k}. \tag{4.9}$$

From the structure (4.4) of $L\mathcal{P}^{(k)}(d)$ it follows that

$$L\mathcal{P}^{(k)}(d) \left(\prod_{n<m} V_n \right) \subset \prod_{n<m-k} V_n. \tag{4.10}$$

By Proposition 3.2 (i) we therefore also have

$$\mathcal{P}^{(k)}(d) \left(\prod_{n<m} V_n \right) \subset \prod_{n<m} V_n. \tag{4.11}$$

Recalling that $n_1 > n_2 > n_3 > \dots$ and using (4.10) and Proposition 3.2 (ii), we deduce that the quotient spaces

$$W_k = V / \prod_{n \le n_1 - k} V_n \tag{4.12}$$

are acted upon by $\mathcal{P}_k(d)$. Since V is the projective limit of the W_k, which are representations of the finite phylon groups $\mathcal{P}_k(d)$, this completes the proof.

5. Jets of densities

In the next section we wish to work with the 'Taylor expansion' of a measure. For an arbitrary measure such a notion makes little sense. However, if the measure is a smooth function times Lebesgue measure then we can sensibly Taylor expand the smooth function. We shall assume, therefore, that the sample space Ω is a smooth manifold of dimension r, say an open subset of some $\mathbf{R}^r$. We are interested in the space $\Gamma(\Delta(\Omega))$ of smooth densities on Ω. A density on a finite-dimensional vector space V is a map ω from the set of all bases of V into $\mathbf{R}$ with the property that if $v = (v^1, \ldots, v^r)$ is a basis and $w = (w^1, \ldots, w^r)$ is another basis related by a linear transformation $X \in GL(r)$ then $\omega(v) = \omega(Xw) = |\det(X)|\omega(w)$. The set $\Delta(V)$ of all densities on V is a one-dimensional vector space. We can form a vector bundle $\Delta(\Omega)$ over Ω whose fibre at a point x in Ω is $\Delta(T_x\Omega)$, the space of all densities on the tangent space ot Ω at x. A (smooth) density on Ω is therefore a (smooth) choice of density on each tangent space, that is, an element of $\Gamma(\Delta(\Omega))$, the space of all smooth sections of $\Delta(\Omega)$. In the case that the sample space is an open subset of $\mathbf{R}^r$ this is just the space of all smooth functions times Lebesgue measure, i.e. of all measures of the form

$$f(x)dx^1 \ldots dx^r. \tag{5.1}$$

More generally, the densities will have such a representation locally. Because $\Delta(\Omega)$ is a vector bundle, we can form the space $J_x^\infty(\Delta(\Omega))$ of all infinite jets of its sections at any point x. Locally, the objects in here are expressions of the form (5.1) where $f(x)$ is now an infinite polynomial in $x^1 \ldots x^r$. More information on densities can be found in Section 6.9 of Murray and Rice (1993) The reader should note carefully that this use of the word 'density' is different from its use in 'probability density'.

Readers familiar with densities and measures will notice one immediate problem with this substitution. If $\phi\colon M \to N$ is a smooth map between manifolds of the same dimension then it induces both a push-out map ϕ_* which maps measures on M to measures on N and a pull-back map ϕ^* which maps densities on N to densities on M. Thus ϕ_* and ϕ^* behave quite differently. Moreover, whereas pulling back densities is something which can be defined at the level of jets, it is not clear that pushing out is. The solution to these dilemmas is the easily verified fact that if ω is a density on N then

$$\phi_*\phi^*\omega = \omega. \tag{5.2}$$

Thus if the map

$$j_m^\infty(\phi^*)\colon J_{\phi(m)}^\infty(\Delta(N)) \to J_m^\infty(\Delta(M)) \tag{5.3}$$

is invertible then we can just define $j_m^\infty(\phi_*)$ to be $j_m^\infty(\phi^*)^{-1}$. The condition for $j_m^\infty(\phi^*)$ to be invertible is that the derivative of ϕ, i.e. the tangent map

$$T_m(\phi) : T_mM \to T_{\phi(m)}N \tag{5.4}$$

is invertible. In coordinates this is just the condition that the Jacobian of ϕ at m be invertible. Moreover, we do not need to start with a function ϕ defined on all of M. It suffices to take some invertible element of

$$J^\infty(M,N)_m^{\phi(m)}, \tag{5.5}$$

the space of all infinite jets of maps from M to N that map m to $\phi(m)$. An invertible element of this space is an element that has its first order term invertible or equivalently, by the inverse function theorem, it is the jet of a local diffeomorphism.

Let P be a family of (mutually absolutely continuous) probability measures on the sample space Ω. Denote by $f(\omega;p)$ the value at ω of the Radon-Nikodym derivative of the probability measure p with respect to some reference measure. The maximum likelihood estimator based on random samples of size N is the function $M^N\colon\Omega^N \to P$ defined by $M^N(\omega_1,\ldots,\omega_N)$ being the element p (assumed unique) of P which maximises $\prod_{i=1}^N f(\omega_i;p)$. If p is any point in P then $M_*^N(p)$ is a measure on P and we are interested in its asymptotic behaviour as $N\to\infty$. It is a result due to Wald (1949) that $M_*^N(p)$ approaches a delta measure at p as $N\to\infty$. Of more interest is the measure

$$(N^{1/2}\phi)_* M_*^N(p), \tag{5.6}$$

where

$$\phi : P \to T_pP \tag{5.7}$$

is a map sending p to 0 which is a local diffeomorphism. This measure is asymptotically a normal distribution on the tangent space at p, with mean 0 and variance the dual of the Fisher information metric. (See, for instance, Cox and Hinkley (1974), pp. 294–295.) We want to understand the way in which the asymptotics of $(N^{1/2}\phi)_* M_*^N(p)$ depend on the choice of ϕ.

To do this, we assume that $M_*^N(p)$ is a density on P and consider its infinite jet

$$j_p^\infty(M_*^N(p)) \in J_p^\infty(\Delta(P)). \tag{5.8}$$

Then we choose a set of coordinates at p, or more precisely the infinite jet of a set of coordinates at p. This is an element

$$j_p^\infty(\phi) \in J_p^\infty(P,\mathbf{R}^d)_p^0, \tag{5.9}$$

which for convenience we shall denote by just ϕ. We can then define the element

$$j_0^\infty((N^{1/2}\phi)_* M_*^N(p)) \in J_0^\infty(\Delta(\mathbf{R}^d)). \tag{5.10}$$

Let us denote by $\mathcal{M}$ the space $J_0^\infty(\Delta(\mathbf{R}^d))$.

Thus $(N^{1/2}\phi)_* M_*^N(p)$ is a sequence of measures μ on $\mathbf{R}^d$ depending on N. Associated to this sequence of measures is a formal asymptotic series

$$\mu = \mu_0 + \frac{1}{N^{1/2}}\mu_1 + \frac{1}{N}\mu_2 + \ldots \tag{5.11}$$

where the μ_i are signed measures on $\mathbf{R}^d$.

If we change the choice of coordinates from ϕ to χ we have $\chi = g.\phi$ for g in the infinite phylon group $\mathcal{P}(d)$. The change in the asymptotic series can be computed from

$$\begin{aligned}(N^{1/2}\chi)_* M_*(p) &= (N^{1/2} g\phi)_* M_*(p) \\ &= (N^{1/2} g \frac{1}{N^{1/2}})_* N^{1/2}(\phi)_* M_*(p).\end{aligned} \tag{5.12}$$

Thus if μ is the asymptotic series of measures (5.11) induced by the ϕ coordinates, that induced by the χ coordinates is

$$(N^{1/2} g \frac{1}{N^{1/2}})_* \mu. \tag{5.13}$$

On passing to jets we obtain an action of $\mathcal{P}(d)$ on $\mathcal{M}$, which we shall call the *twisted phylon action*. This is to distinguish it from the action where g just acts on each μ_i in the natural way. Recall from the proof of Proposition 4.1 that every non-zero $t \in \mathbf{R}$ defines a dilation $t.1 \in GL(d) \subset \mathcal{P}(d)$. Thus $(N^{1/2} g \frac{1}{N^{1/2}})_*$ is the composition of three elements of $\mathcal{P}(d)$ acting on $\mathcal{M}$. To obtain further insight into this action it is useful to consider more general phylon representations and develop a theory of *twisted phylon representations*.

6. Twisted phylon representations

Let V be a phylon representation as in Section 4. Decompose it into subspaces V_i of homogeneous degree n_i

$$V = \prod_{i=1}^{\infty} V_{n_i} \tag{6.1}$$

As an example, let $\mathcal{M}$ be the space $J_0^\infty(\Delta(\mathbf{R}^d))$ of infinite jets of measures considered in the previous section. Thinking of these as infinite formal power series multiplied by Lebesgue measure, we see that they can be decomposed into sums of homogeneous terms of degree k for each $k \geq 0$. The subspaces C_k consisting of homogeneous functions of degree k are the irreducible $GL(d)$-factors of $\mathcal{M}$. Recalling that the action of an element $g: \mathbf{R}^d \to \mathbf{R}^d$ in the phylon group when applied to a function $f: \mathbf{R}^d \to \mathbf{R}$ is $f \circ g^{-1}$ and that the action of a dilation $\lambda.1$ on Lebesgue measure is to multiply it by λ^d, we see that the action of such a dilation on C_k is multiplication by λ^{d-k}, that is, $n_k = d - k$. This example will be our model.

Given the representation V, we can form the space $V[[t]]$ of all formal asymptotic power series in t:

$$v = v_0 + v_1 t + v_2 t^2 + \ldots, \tag{6.2}$$

where each v_i is in V. As V is a direct product we see that $V[[t]]$ is, in fact, bi-graded as

$$V = \prod_{i\geq 1} \prod_{m\geq 0} V_{n_i} t^m. \tag{6.3}$$

Thus any element $v \in V[[t]]$ can be written as

$$v = \sum_{i\geq 1} \sum_{m\geq 0} v_i t^m \tag{6.4}$$

with $v_i \in V_{n_i}$. Given an element $w = v_i t^a$ with $v_i \in V_{n_i}$, let us call n_i the homogeneous degree of w and a the asymptotic degree of w.

We define the twisted action of the phylon group on $V[[t]]$ by

$$g \star v = t^{-1} g t.v. \tag{6.5}$$

To understand this action in more detail it is useful to consider the corresponding twisted Lie algebra action. Let $X \in L\mathcal{P}(d)_{-k}$ and $v \in V_{n_i}$. Then

$$X \star v t^a = t^{-1} X v t^{a+n_i} = (Xv) t^{a+k}, \tag{6.6}$$

so that the homogeneous degree decreases and the asymptotic degree increases. The action on subspaces is therefore

$$L\mathcal{P}(d)_{-k} \star V_{n_i} t^m \subset V_{n_i - k} t^{m+k}. \tag{6.7}$$

If we filter $V[[t]]$ by the subspaces $V(a) = V[[t]].t^a$ so that

$$V[[t]] = V(0) \supset V(1) \supset V(2) \supset \dots \tag{6.8}$$

then

$$L\mathcal{P}(d)_{-k} \star V(a) \subset V(a+k). \tag{6.9}$$

It follows from Proposition 3.2 that there is an action of $\mathcal{P}(d)$ on the quotient spaces

$$V(a)/V(a+k) \tag{6.10}$$

and that $L\mathcal{P}^{(k)}(d)$ acts trivially on $V(a)/V(a+k)$. Thus the action of $\mathcal{P}(d)$ on $V(a)/V(a+k)$ factors through an action of $\mathcal{P}_k(d)$.

In the application to the maximum likelihood estimator that we are interested in we just need to replace t by $1/N^{1/2}$. Then the action of $\mathcal{P}(d)$ on $\mathcal{M}(0)/\mathcal{M}(k)$ factors through an action of $\mathcal{P}_k(d)$. This says that if we want to understand the behaviour of the asymptotic expansion (5.11) up to and including order $1/N^{k/2}$ under change of parameterisation (i.e. under the action of the phylon group), we need only consider the Taylor expansion of the change of parameterisation up to order k. This fact is the connection between the two types of order in asymptotic expansions mentioned in the introduction. In particular, considering the action of $\mathcal{P}(d)$ on $\mathcal{M}(0)/\mathcal{M}(1)$, we deduce the familiar fact that the leading order term μ_0 in the asymptotic expansion (5.11) of $(N^{1/2}\phi)_* M_*^N(p)$ is acted on only by the matrix

$$\frac{\partial \chi^i}{\partial \phi^a}(p) \tag{6.11}$$

when we change ϕ to a new set of coordinates χ. Since this action is the usual action of $GL(d)$ on $\mathbf{R}^d$, this is another way of saying that μ_0 should be thought of as a measure on the tangent space to P at p.

7. The coadjoint action

We conclude this paper with some remarks on the coadjoint action of $\mathcal{P}(d)$ on the dual $L\mathcal{P}(d)^*$ of the Lie algebra. For infinite-dimensional Lie groups it is known that the orbit theory of Kirillov (1976) often provides a guide to constructing an interesting class of representations. In the case of the phylon group we obtain yet another link with Terng's (1978) results.

Firstly, we note from the discussion in Section 2 that (i) because the Lie algebra of the phylon group is a direct product, the dual is a direct sum

$$L\mathcal{P}(d)^* = \oplus_{i \geq 1}[\mathbf{R}^d \otimes S^i(\mathbf{R}^d)^*], \tag{7.1}$$

(ii) $L\mathcal{P}(d)$ is not a Banach space. Since the dual of a Fréchet space which is not a Banach space is never a Fréchet space (Hamilton, 1982, p. 69), the coadjoint action is not a phylon representation in the sense of Definition 4.1. Clearly, this is a manifestation of the general fact that, in the language of string theory, infinite-dimensional strings and co-strings are quite different.

Let ξ be an element of $L\mathcal{P}(d)^*$. We want to consider the orbit of ξ under the coadjoint action of the phylon group. Let $\langle\ ,\ \rangle$ denote the pairing of a vector space and its dual. Then under the coadjoint action g in $\mathcal{P}(d)$ maps ξ in $L\mathcal{P}(d)^*$ to $g^{-1}\xi g$, where

$$\langle g^{-1}\xi g, X\rangle = \langle \xi, g^{-1}Xg\rangle$$

for all X in $L\mathcal{P}(d)$. Recall that we have projections

$$j^k \colon L\mathcal{P}(d) \to L\mathcal{P}_k(d) \tag{7.2}$$

and therefore dual inclusions

$$j^{k*} \colon L\mathcal{P}_k(d)^* \to L\mathcal{P}(d)^*. \tag{7.3}$$

Because the dual of the Lie algebra is the direct sum (7.1), if $\xi \neq 0$ then for some k it has the form

$$\xi = (\xi_1, \xi_2, \xi_3, \ldots, \xi_k, 0, 0, \ldots) \tag{7.4}$$

with $\xi_i \in \mathbf{R}^d \otimes S^i(\mathbf{R}^d)^*$ and $\xi_k \neq 0$. It follows that $\xi = j^{k*}(\xi')$, where

$$\xi' = (\xi_1, \xi_2, \xi_3, \dots, \xi_k) \in L\mathcal{P}_k(d)^*, \tag{7.5}$$

and that k is the smallest integer for which this is true. We call k the order of ξ.

For g in the phylon group and X in its Lie algebra we have

$$\begin{aligned}\langle g^{-1}\xi g, X\rangle &= \langle g^{-1} j^{k*}(\xi') g, X\rangle \\ &= \langle \xi', j^k(g^{-1}Xg)\rangle \\ &= \langle \xi', j^k(g^{-1}) j^k(X) j^k(g)\rangle \\ &= \langle j^{k*}(j^k(g)^{-1}\xi' j^k(g)), X\rangle. \end{aligned} \tag{7.6}$$

Hence we have the identity:

$$g^{-1}\xi g = j^{k*}(j^k(g)^{-1}\xi' j^k(g)). \tag{7.7}$$

Thus if ξ is an element of order k then its stabiliser contains $\mathcal{P}^{(k)}(d)$ and hence the group $\mathcal{P}_k(d)$ acts transitively on its $\mathcal{P}(d)$-orbit, which must therefore be finite-dimensional. Moreover, the coadjoint orbit of ξ is the image under $j^{k*}: L\mathcal{P}_k(d)^* \to L\mathcal{P}(d)^*$ of the coadjoint orbit of ξ'. It follows that the representations of $\mathcal{P}(d)$ obtained from coadjoint orbits are actually representations of the finite-dimensional phylon groups $\mathcal{P}_k(d)$.

Note that we have not explored here the question of whether or not the representations of $\mathcal{P}_k(d)$ described by Terng (1978) arise from orbits of the coadjoint representation. All we have observed is that the orbit theory for the infinite phylon group is no more than the union of the theories for its finite-dimensional quotient groups. Thus the orbit theory will not give rise to the infinite-dimensional examples which we have been discussing.

References

O.E. Barndorff-Nielsen, P. Blæsild, A.L. Carey, P.E. Jupp, M. Mora and M.K. Murray: *Finite dimensional algebraic representations of the infinite phylon group*. Acta. Appl. Math., **28**, 219–252, 1992.

O.E. Barndorff-Nielsen, P.E. Jupp, and W.S. Kendall: *Stochastic calculus, statistical asymptotics, Taylor strings and phyla*. Ann. Fac. Sci. Toulouse, **III**, No 1, 5–62, 1994.

A.L. Carey and M.K. Murray: *Higher order tensors, strings and new tensors*. Proc. R. Soc. Lond. A., **430**, 423-432, 1990.

D.R. Cox and D.V. Hinkley: *Theoretical Statistics*. Chapman and Hall, London, 1974.

R.S. Hamilton: *The inverse function theorem of Nash and Moser*. Bull. Am. Math. Soc., **7**, 65–222, 1982.

A.A. Kirillov: *Elements of the Theory of Representations*. Springer-Verlag, Berlin, 1976.

O. Kobayashi, A. Yoshioka, Y. Maeda and H. Omori: *The theory of infinite-dimensional Lie groups and its applications*. Acta Appl. Math., **3**, 71–106, 1985.

I. Kolář, P.W. Michor and J. Slovák: *Natural Operations in Differential Geometry*. Springer-Verlag, London, 1993.

J. Milnor: *Remarks on infinite-dimensional Lie groups*. In *Les Houches, Session XL, 1983, Relativity, Groups and Topology II*. B.S. DeWitt and R. Stora, eds. Elsevier, 1984.

M.K. Murray and J.W. Rice: *Differential Geometry and Statistics*. Chapman and Hall, London, 1993.

H. Omori: *A method of classifying expansive singularities*. J. Diff. Geom., **15**, no. 4, 493–512, 1980.

C.L. Terng *Natural vector bundles and natural differential operators*. Amer. J. Math., **100**, 775–828, 1978.

F. Treves: *Topological Vector Spaces, Distributions and Kernels*. Academic Press, New York, 1967.

A. Wald: *Note on the consistency of the maximum likelihood estimate*. Ann. Math. Stat., **20**, 595–601, 1949.