Annales de la faculté des sciences de Toulouse

Ole E. Barndorff-Nielsen

Peter E. Jupp

Wilfrid S. Kendall

# Stochastic calculus, statistical asymptotics, Taylor strings and phyla

*Annales de la faculté des sciences de Toulouse 6e série*, tome 3, nº 1 (1994), p. 5-62

<http://www.numdam.org/item?id=AFST_1994_6_3_1_5_0>

Annales de la Faculté des Sciences de Toulouse Vol. III, n° 1, 1994

# Stochastic calculus, statistical asymptotics, Taylor strings and phyla(*)

OLE E. BARNDORFF-NIELSEN(1), PETER E. JUPP(2) et WILFRID S. KENDALL(3)

RÉSUMÉ. — Cet article fournit un exposé sans démonstration de la théorie du calcul d'ordre supérieur que l'on appelle la théorie des cordes (cordes de Taylor ou cordes statistiques). Cette théorie était introduite comme outil pour traiter les complexités géométriques de la statistique asymptotique mais elle pourrait aussi jouer un rôle dans les approches géométriques aux études asymptotiques. (La théorie des cordes de Taylor n'a rien à voir avec le concept de corde tel qu'on l'utilise en physique mathématique.) L'exposé est divisé en trois parties. La première partie introduit la théorie des cordes au second ordre via une description du formalisme du calcul stochastique. La deuxième partie décrit des idées d'invariance de la statistique asymptotique et la géométrie des jougs statistiques, ainsi que la manière dans laquelle elles conduisent à la formation de la théorie des cordes de Taylor. Dans la troisième partie, on passe en revue les rapports entre la théorie des cordes de Taylor et d'autres concepts mathématiques, tels que les fibrés de jets et les fibrés naturels. La théorie des cordes de Taylor s'introduit pas à pas comme on en a besoin.

MOTS-CLÉS : Algèbre symbolique, calcul d'Itô, calcul d'Itô symbolique, calcul d'ordre supérieur, calcul stochastique, champ de cordes, connexion, différentielle du deuxième ordre, fibré de repères d'ordre $T$, fibré vectoriel naturel, corde de connexions, corde de coordonées, corde de dérivées, corde de différentielles, corde scalaire, formule-$p^*$, groupe de phyla, jets, jets semi-holonomes, joug, phyla, REDUCE, semi-martingale, série invariante de Taylor, statistique asymptotique, vraissemblance.

ABSTRACT. — This paper provides an exposition without proofs of the theory of higher order calculus known as (Taylor or statistical) string theory. This theory was introduced as a tool to deal with the geometric

(*) Reçu le 15 décembre 1992

(1) Depart. Theor. Statistics, Aarhus University, DK-8000 Aarhus C (Denmark)

(2) Dept. of Math. and Comp. Sciences, University of St. Andrews, North Haugh, St. Andrews KY16 9SS (Scotland, UK)

(3) Department of Statistics, University of Warwick, Coventry CV4 7AL (UK)

complexities of statistical asymptotics, but has a potentially wider role to play in geometric approaches in asymptotic studies. (Note that Taylor string theory is not related to the concept of a string as used in mathematical physics.) The exposition is a survey in three parts, intended to introduce and to motivate the study of Taylor strings. The paper begins by introducing string theory at the second order *via* a description of the formalism of the second-order stochastic calculus, a powerful method of analysis of an important class of random processes. There follows a description of some invariance considerations of statistical asymptotics and the geometry of statistical yokes, and the way in which they lead to the formulation of Taylor string theory. Finally a survey is given of the interrelationships between Taylor string theory and other concepts from mathematics such as jet bundles and natural bundles. The apparatus of Taylor string theory is introduced step-by-step as it becomes necessary, thus providing intuition and applications to motivate this approach to higher-order calculus and higher-order invariance.

KEY-WORDS : computer algebra, connection, connection string, coordinate string, derivative string, differential string, higher-order calculus, invariant Taylor series, Itô calculus, jets, likelihood, natural vector bundle, $p^*$-formula, phyla, phylon group, *REDUCE*, scalar string, second-order differential, semi-holonomic jets, semimartingale, statistical asymptotics, stochastic calculus, string field, symbolic Itô calculus, $T^{th}$-order frame bundle, yoke.

**AMS classification :** 62A10, 62E20, 60H05, 58A20

---

## 1. Introduction

The use in statistics of geometric intuition and concepts has a long history and has recently experienced a considerable surge of progress through a number of related developments. See for example Amari [1], Amari *et al.* [2], Barndorff-Nielsen [8], Barndorff-Nielsen, Cox & Reid [16], Critchley, Marriott & Salmon [23], Kass [39], McCullagh [51]. The purpose of this paper is to expound and to survey one of these developments, a fascinating link between four topics which have been developed largely in isolation from each other: second order stochastic calculus; the geometry of statistical asymptotics; some generalizations of tensors suggested by invariance considerations in physics; and the concept of natural vector bundles. The link lies in the behaviour under coordinate transformation of local descriptions of the objects involved in these topics. The feature common to second order stochastic calculus and statistical asymptotics is that the underlying mathematical objects are *derivative strings*, in the sense of Barndorff-Nielsen [7], Barndorff-Nielsen & Blæsild [10]-[11]. Derivative strings, differential strings (Blæsild & Mora [19]) and the "new tensors" (suggested *inter alia* by considerations from theoretical physics; Foster [32]-[35]) all alike have a natural generalization to *phyla*, objects with transformation laws in a very general algebraic class. The class of phyla includes tensors, affine connections, jets of functions and of tensor fields and almost all objects arising in classical and higher order differential geometry. Thus it may be viewed as forming a suitable general context for "higher-order calculus". The idea of "invariants" or "geometrical objects" as objects which transform in almost the most general possible way under smooth coordinate transformation has been formalized using the concept of elements of natural vector bundles. It turns out that phyla are precisely the elements of algebraic natural vector bundles, and that algebraic natural vector bundles provide the ultimate mathematical framework for the study of Taylor strings.

This link provides a powerful and expressive formalism for dealing in a geometric way with the calculations that arise in statistical asymptotics and, to a lesser extent, with those arising in stochastic calculus. Moreover the demands of statistical inference call forth constructions which are of intrinsic geometric interest, such as the geometries defined by *yokes* and their associated string fields, described briefly in section 3. Finally the demands of stochastic calculus have stimulated the implementation within a computer algebra package of an effective framework formalizing the theory of stochastic calculus which is closely related to the higher-order invariance considerations which the theories above hold in common. Similar implementations of statistical asymptotics are currently being sought, using the geometric perspectives described below. It is hoped that this expository paper will stimulate and aid workers in the fields of geometry, statistical asymptotics and stochastic calculus to consider what benefits might accrue from further examination and investigation of these relationships.

The relevant aspects of stochastic calculus and of statistical asymptotics are outlined in sections 2 and 3 respectively, while section 4 gives a brief indication of the theory of strings and phyla and discusses its relations to natural vector bundles and other geometric constructs, including statements without proof of a number of key theorems.

The reader is alerted to the use of the summation convention (summation over repeated dummy indices) in formulae where there are no explicit signs of summation. Associated notation is described in section 2.2. An extension of this summation convention is described in section 3.2 and employed from then on except where indicated otherwise.

## 2. Second order stochastic calculus

Our first example of higher-order calculus arises in the theory of the *stochastic calculus*. This theory allows one to use calculus on a class of random processes known as the *semimartingales*. Semimartingales are continuous-time random processes which can be decomposed in a way analogous to the basic "signal-plus-noise" decomposition of statistics and signal-processing. The classical differential calculus cannot be applied to these processes, because their sample paths are in general too irregular. Nevertheless the signal-plus-noise decomposition means that an extension of differential calculus is available, namely the stochastic calculus. As we will see below, the stochastic calculus is *second-order* in distinction to the classical differential calculus, which is first-order. To be more precise about this distinction, classical differential calculus deals with limits of approximations by first-order truncated Taylor series which may be expressed using finite-difference notation as

$$\Delta f\big(x(t)\big) = f'\big(x(t)\big) \cdot \Delta x(t) + o(\Delta t) \tag{2.1}$$

which holds for sample paths $x$ of locally bounded variation (hence $\Delta x(t) = O(\Delta t)$) while stochastic calculus deals with limits of approximations by second-order truncated Taylor series

$$\Delta f\big(x(t)\big) = f'\big(x(t)\big) \cdot \Delta x(t) + \frac{1}{2} f''\big(x(t)\big) \cdot \big(\Delta x(t)\big)^2 + o(\Delta t) \tag{2.2}$$

which holds for certain classes of random sample paths $x$ satisfying

$$\big(\Delta x(t)\big)^2 = O(\Delta t)\,.$$

The move from first- to second-order truncated Taylor series is the first step towards the theory of higher-order calculus known as the *Theory of phyla*. In this section we summarize the elements and basic operations of stochastic calculus (also called Itô calculus) of continuous semimartingales. We then introduce and discuss notations and definitions used in the stochastic calculus, taking the opportunity also to introduce notational conventions to be used later in the paper. Then we consider the invariant formulation of stochastic calculus, which is a matter of importance not only for aesthetic reasons but also for the purposes of computation and presentation of results. The section is concluded by a brief discussion of issues raised by implementation of stochastic calculus within a computer algebra package. It will be seen that computer algebra and geometrical invariance are closely linked by the notions of second-order calculus.

## 2.1. Elements of stochastic calculus

Basic to stochastic calculus are the notions of *adapted processes* and *filtrations*. Suppose given a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Informally, an adapted process $X$ is one whose value $X(t)$ at any fixed time $t$ depends only on events which are determined before time $t$. Let $\mathcal{F}_t \subset \mathcal{F}$ denote the ensemble of all events deemed to be determined before time $t$. It is reasonable to insist that each of these ensembles is a $\sigma$-field. The specification of all these $\sigma$-fields fixes the *filtration* $\{\mathcal{F}_t : t \geq 0\}$. An adapted process $X$ is defined by the property: $X(t)$ is measurable with respect to $\mathcal{F}_t$ for each $t$. One of the fascinations of stochastic calculus is the way in which specification of the filtration $\{\mathcal{F}_t : t \geq 0\}$, and thence of the adapted processes, lays down the structure of cause and effect in a problem. See Dellacherie & Meyer [25]-[27] for a very thorough discussion of the ramifications involved here, in the case of discontinuous as well as of continuous processes. A brief but informative presentation is to the found in Meyer's appendix to Emery [30]. There are many other useful monographs: for example Protter [59], Rogers & Williams [61] and (for the continuous sample-path theory to which the following discussion is restricted) Revuz & Yor [60].

We may now define a (*continuous sample-path*) *semimartingale* $X$ as an adapted continuous random process expressible as a sum $X = X(0) + M + V$ of a continuous local martingale $M$ and a continuous process $V$ such that the paths of $V$ are of locally bounded variation. This decomposition is actually unique if we take $M(0) = V(0) = 0$. (Recall that a *martingale* $M$ is defined by the following property concerning conditional expectations: $\mathbf{E}[M(t+s) \mid \mathcal{F}_t] = M(t)$ for all $t$, $s \geq 0$. A *local martingale* can be turned into a martingale by stopping it at any one of a sequence of stopping times tending to infinity.) Examples of (continuous sample-path) semimartingales include: Brownian motion B (when $M = B$ and $V = 0$ identically), solutions $y(t)$ to ordinary differential equations $y' = f(y)$ for bounded smooth $f$ (when $M = 0$ identically and $V = y - y(0)$, and smooth functions $f(B)$ of Brownian motion B (when $M$, $V$ may be derived from (SC1)-(SC4) below). One may think of $V$ as being the signal or trend, and $M$ as being the noise, so that $X = X(0) + M + V$ is a "signal-plus-noise" decomposition. We shall call $V$ the (*integrated*) *drift* of $X$. Being of locally bounded variation, $V$ may be viewed as the solution of an ordinary differential equation, perhaps with stochastic coefficients and perhaps subject to a (possibly singular) random time-change. Consequently ordinary calculus applies to the bounded-variation part $V$. However save in trivial cases the local-martingale part $M$ never has paths of locally bounded variation and so the full force of stochastic calculus is required. There are many treatments of stochastic calculus (see for example those mentioned above) so here we merely sketch the elements of the theory in order to illustrate and to explain the invariance considerations which link stochastic calculus to the other topics of this paper.

In summary, the elements of (continuous sample-path) stochastic calculus are:

(SC1) *A smooth invariance result.*— If $X = X(0) + M + V$ is a semimartingale and $f$ is a $C^2$ function then $f(X)$ is a semimartingale, with decomposition say $f(X) = f\big(X(0)\big) + N + U$. In more formal terms, *the class $\mathcal{S}$ of continuous sample-path semimartingales is closed under application of $C^2$ functions.* Let $\mathcal{S}_0 = \big\{X \in \mathcal{S} : X(0) = 0\big\}$, $\mathcal{V}$ be the class of adapted continuous processes of locally bounded variation, and $\mathcal{V}_0 = \big\{V \in \mathcal{V} : V(0) = 0\big\}$.

(SC2) *An integral extending the Lebesgue-Stieljes integral.*— If $X$ is a semimartingale and $H$ is a bounded adapted process then we can make sense of $\int H \, \mathrm{d}X$ as a semimartingale in a way that both extends the Lebesgue-Stieltjes integral and also respects the "signal-plus-noise" decomposition (if $X = M$ is a local martingale then so is $\int H \, \mathrm{d}M$).

(SC3) *A measure of the randomness of the martingale part of a semimartingale $X = X(0) + M + V$.*— The *bracket process* $[X , X] = [M , M]$ is the unique continuous adapted increasing process given by the (Doob-Meyer) decomposition of the semimartingale $M^2$ as $M^2 = N + [M , M]$ where $N$ is a continuous local martingale and where $[M , M](0) = 0$. In fact $[X , X]$ is also known as the *quadratic variation* of $X$, since $[X , X](t) = \lim \sum_{s \leq t} (\Delta X)^2$ taking the limit in probability over dyadic partitions of the time axis. A particular and important example is that of Brownian motion B, for which $[B , B](t) = t$ for all times $t$. Note that $[X , Y]$ can be defined by polarization. The bracket process allows us to relate ordinary and stochastic integration via $\int H \, \mathrm{d}[X , Y] = \big[\int H \, \mathrm{d}X , Y\big]$, holding for all bounded adapted $H$.

(SC4).— A "*change-of-variables*" formula to make (SC1) precise: the famous Itô formula asserts that if $f$ is a $C^2$ function then

$$f\big(X(t)\big) = f\big(X(0)\big) + \int_0^t f'(X) \, \mathrm{d}X + \frac{1}{2} \int_0^t f''(X) \, \mathrm{d}[X , X] . \qquad (2.3)$$

This of course uses (SC3), and also (SC2) to make sense of $\int f'(X)\,\mathrm{d}X$, and allows us to determine the "signal-plus-noise" decomposition of $f(X)$.

(SC5) *The quadratic nature of* $[X\,,\,X]$.— For example if $f$ is $C^2$ then

$$[f(X)\,,\,f(X)] = \int f'(X)^2\,\mathrm{d}[X\,,\,X]\,. \tag{2.4}$$

(SC6).— It can be shown that (subject to suitable regularity conditions) *the statistical behaviour of a semimartingale* $X = X(0) + M + V$ *is determined once one knows its second-order structure*, which is to say, formulae for $X(0)$, $V$ and $[X\,,\,X] = [M\,,\,M]$. For example if $X(0) = 0$, $V$ vanishes and $[X\,,\,X](t) = t$ for all $t$ then a theorem of Lévy asserts that $X$ is actually real-valued Brownian motion. If the second-order structure is more complicated then information may be derived for example from formulation as a *martingale problem*: the process determined by

$$f(X(t)) - f(X(0)) - \int_0^t f'(X)\,\mathrm{d}V - \frac{1}{2}\int_0^t f''(X)\,\mathrm{d}[X\,,\,X] \tag{2.5}$$

is actually a continuous martingale whenever $f$ is $C^2$ and of compact support.

There is of course a more general stochastic calculus dealing with *discontinuous* semimartingales. However it is the continuous stochastic calculus which is dominant in applications, with the important exception of stochastic calculus for adapted locally-bounded variation processes with jumps only of $\pm 1$. Presumably this dominance is due to the relative parsimony of descriptions of continuous semimartingales based on the second-order structure of $X$ referred to in (SC6). (A similar parsimony holds in the exceptional case of $\pm 1$ jumps.)

## 2.2. Notations and definitions

Stochastic integrals such as those written down in (2.3)-(2.5) are notationally tedious. The formalism of *stochastic differentials* provides an extremely useful abbreviation: for example (2.3) and (2.4) are usually written as

$$\mathrm{d}f(X) = f'(X)\,\mathrm{d}X + \frac{1}{2}f''(X)\,\mathrm{d}[X\,,\,X] \tag{2.6}$$

$$\mathrm{d}[f(X)\,,\,f(X)] = f'(X)^2\,\mathrm{d}[X\,,\,X]\,. \tag{2.7}$$

Stochastic differentials are interpreted by integration against bounded adapted processes $H$. As pointed out in Itô [37], this notation treats $\mathrm{d}X$ *et cetera* as generating a space $\mathcal{D}^2$ of stochastic differentials which is an $\mathcal{H}$-module, where $\mathcal{H}$ is the family of bounded adapted processes. Let $\mathcal{D}^1 \subset \mathcal{D}^2$ be the space of conventional differentials generated by $\mathrm{d}V$, $V \in \mathcal{V}$. One makes sense of stochastic differentials by integrating them; thus the stochastic integration of (SC2) provides a linear map

$$\begin{aligned} &\int : \mathcal{D}^2 \to \mathcal{S}_0 \\ \text{extending } &\int : \mathcal{D}^1 \to \mathcal{V}_0 \text{ Lebesgue-Stieltjes integral.} \end{aligned} \tag{2.8}$$

Here $\int$ inverts $\mathrm{d}\restriction_{\mathcal{S}_0}$, where

$$d : \mathcal{S} \to \mathcal{D}^2 \tag{2.9}$$

is the procedure of taking stochastic differentials. So $\int$ is an "anti-derivative". The operator $\int$ is determined in the obvious way for $h\,\mathrm{d}X$ when $h$ is piecewise constant in time, and is defined for general $h \in \mathcal{H}$ as the limit of $\int h_n\,\mathrm{d}X$ for piecewise constant $h_n \in \mathcal{H}$ converging rapidly to $h$. Of course it is a nontrivial matter to show that this leads to a good definition of $\int h\,\mathrm{d}X$.

Similarly the bracket construction of (SC3) yields a symmetric $\mathcal{H}$-bilinear map

$$\begin{aligned} \text{Bracket} : \mathcal{D}^2 \times \mathcal{D}^2 &\to \mathcal{D}^1 \\ (\mathrm{d}X, \mathrm{d}Y) &\mapsto \mathrm{d}[\,X\,,\,Y\,]\,. \end{aligned} \tag{2.10}$$

It is important that Bracket (also called "multiplication of Itô differentials" or "quadratic covariation") vanishes on $\mathcal{D}^1 \times \mathcal{D}^2$ and on $\mathcal{D}^2 \times \mathcal{D}^1$. The basic examples are the products arising from differentials $\mathrm{d}A$, $\mathrm{d}B$ of two independent Brownian motions and the differential of time $\mathrm{d}t$ forming the *Itô multiplication table*:

$$\begin{aligned} \mathrm{d}A^2 = \mathrm{d}B^2 &= \mathrm{d}t \\ \mathrm{d}A \times \mathrm{d}B = \mathrm{d}A \times \mathrm{d}t &= \mathrm{d}t^2 = 0\,. \end{aligned} \tag{2.11}$$

In quite a different way the decomposition $X = X(0) + M + V$ into noise plus drift yields an $\mathcal{H}$-linear map

$$\text{Drift} : \mathcal{D}^2 \to \mathcal{D}^1\,. \tag{2.12}$$

Note Drift is the identity on $\mathcal{D}^1$. Furthermore Drift depends on the underlying probability measure $\mathbb{P}$ while $\int$, Bracket depend on $\mathbb{P}$ only through its equivalence class under absolute continuity of probability measures. For example consider $X = B^2$ the square of Brownian motion B. By (2.6) and (SC3) we have $\mathrm{d}X = 2B\,\mathrm{d}B + \mathrm{d}t$. Since $B$ is a martingale we can use (SC2) to deduce $\mathrm{Drift}(\mathrm{d}X) = \mathrm{d}t$. On the other hand it is possible to make a change of probability measure such that $B(t) = A(t) + t$ for all times $t$, where $A$ is a Brownian motion under the new probability measure, while $B$ is no longer a Brownian motion under the new probability measure. The decomposition $\mathrm{d}X = 2B\,\mathrm{d}B + \mathrm{d}t$ continues to hold under the new probability measure, but now $\mathrm{Drift}(\mathrm{d}X) = \mathrm{Drift}\big(2B(\mathrm{d}A + \mathrm{d}t) + \mathrm{d}t\big) = (2B+1)\mathrm{d}t$.

Together with the invariance property (SC1), the expressive nature of this formulation in terms of differentials provides a strong incentive to try to formulate $\mathcal{D}^2$ and the above constructions for *multivariate* semimartingales in a geometric and intrinsic manner. Property (SC1) makes it clear that the notion of a semimartingale $X = (X^1, \ldots, X^d)$ on a $C^2$ manifold is well-defined by the requirement that $f(X)$ be a semimartingale whenever $f$ is a real-valued $C^2$ function. However, if $X = (X^1, \ldots, X^d)$ is a multivariate semimartingale and if $f$ is twice-differentiable then the multivariate form of (2.6) is

$$\mathrm{d}f(X) = \sum_{i=1}^{d} \frac{\partial f}{\partial x^i}\,\mathrm{d}X^i + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 f}{\partial x^i \partial x^j}\,\mathrm{d}[X^i, X^j]. \tag{2.13}$$

Not only will this lead to complicated formulae for high dimensionality $d$, (generally requiring $d + d(d+1)/2$ summands as opposed to the $d$ summands of the counterpart in ordinary calculus!), but also the presence of second-order terms means that $\mathrm{d}X$ is not at all a tensorially-invariant quantity. Thus $X$ makes sense as a semimartingale on a manifold, $\mathrm{d}X$ makes sense in coordinates, but if $\mathrm{d}X$ makes invariant sense then it must be in some extended sense and *not* as a conventional invariant. Some economy of notation is brought in by introducing the repeated index summation convention of Einstein and the associated convention for partial derivatives, $\partial f/\partial x^i = f_{/i}$, $\partial^2 f/\partial x^i \partial x^j = f_{/ij}$:

$$\mathrm{d}f(X) = f_{/i}\,\mathrm{d}X^i + \frac{1}{2} f_{/ij}\,\mathrm{d}[X^i, X^j]. \tag{2.14}$$

However the lack of tensorial invariance is still a problem. In order further to demonstrate the problem (and incidentally to introduce more conventions used later on in this survey), suppose that $\psi$ is a new system of coordinates and consider the expression for the stochastic differential of $Y = \psi(X)$ in this new set of coordinates. We find from (2.14)

$$\mathrm{d}Y^a = \mathrm{d}\psi^a(X) = \psi^a_{/i}(X)\,\mathrm{d}X^i + \frac{1}{2}\,\psi^a_{/ij}(X)\,\mathrm{d}[X^i\,,\,X^j] \tag{2.15}$$

where indices $i$, $j$, ... refer to derivatives taken with respect to the original set of coordinates and indices $a$, $b$, ... refer to coordinates based on $\psi$. Furthermore if $f$ is a real-valued $C^2$-function as above then

$$\begin{aligned}\mathrm{d}\big(f(Y)\big) &= \mathrm{d}\,\big(f(\psi(X))\big)\\ &= f_{/a}(Y)\psi^a_{/i}(X)\,\mathrm{d}X^i + \frac{1}{2}\,f_{/a}(Y)\psi^a_{/ij}(X)\,\mathrm{d}[X^i\,,\,X^j] + \\ &\qquad + \frac{1}{2}\,f_{/ab}(Y)\psi^{ab}_{/ij}(X)\,\mathrm{d}[X^i\,,\,X^j]\,,\end{aligned} \tag{2.16}$$

$$\begin{aligned}\mathrm{d}[f(Y)\,,\,f(Y)] &= \mathrm{d}\left[f\big(\psi(X)\big)\,,\,f\big(\psi(X)\big)\right]\\ &= f_{/a}(Y)f_{/b}(Y)\psi^{ab}_{/ij}(X)\,\mathrm{d}[X^i\,,\,X^j]\end{aligned} \tag{2.17}$$

where $\psi^{ab}_{/ij} = \psi^a_{/i}\psi^b_{/j}$. The lack of tensorial invariance is clear from the second derivatives occurring in the two trailing terms in (2.16) and the potential for complexity of formulae is evident.

We digress briefly to extend the notational conventions above in a way which will be used extensively when we come to discuss the coordinate-based theory for Taylor strings and phyla. Let $\omega = (\omega^1, \ldots, \omega^d)$ and $\psi = (\psi^1, \ldots, \psi^d)$ be two alternative coordinate systems or parametrizations for a manifold $\mathbb{M}$. (We treat these as though they were global rather than local, and indeed the distinction does not affect the essence of the discussion throughout this survey paper. Of course most of the interesting questions in geometry and stochastic differential geometry, though not necessarily in statistical asymptotics, arise when we consider the issue of extending from local to global considerations!) Generic coordinates of $\omega$ and $\psi$ are indicated by $\omega^i$, $\omega^j$, $\omega^k$, ... and $\psi^a$, $\psi^b$, $\psi^c$, ... respectively. If $I = (i_1 \ldots i_n)$ and $A = (a_1 \ldots a_m)$ are finite sequences of indices and if $f$ is a suitably smooth function then we write

$$f_{/I} = f_{/i_1 \ldots i_n} = \frac{\partial^n f}{\partial\omega^{i_1} \ldots \partial\omega^{i_n}}\,, \tag{2.18}$$

and we also use the following notation for sums over partitions:

$$\psi^A_{/I} = \psi^{a_1 \ldots a_m}_{/i_1 \ldots i_n} = \sum_{I/m} \psi^{a_1}_{/I_1} \cdots \psi^{a_m}_{/I_m} \tag{2.19}$$

where the summation is over all *standard ordered partitions* of $I$ into $m$ nonempty subsequences or blocks $I_1, \ldots, I_m$ such that:

(PS1) for $m \leq n$ the order of the indices in each of $I_1, \ldots, I_m$ is the same as their order within $I$,

(PS2) for $p = 1, \ldots, m-1$ the first index in $I_p$ comes before the first index in $I_{p+1}$ in the ordering within $I$,

(PS3) in the case $m > n$ we interpret both sides of equation (2.19) as equalling zero.

Here is an illustration of this notation. Suppose $A = (a_1 a_2)$ and $I = (i_1 i_2 i_3)$. Then

$$\psi^A_{/I} = \sum_{I/2} \psi^{a_1}_{/I_1} \psi^{a_2}_{/I_2} = \psi^{a_1}_{/i_1} \psi^{a_2}_{/i_2 i_3} + \psi^{a_1}_{/i_1 i_2} \psi^{a_2}_{/i_3} + \psi^{a_1}_{/i_1 i_3} \psi^{a_2}_{/i_2} .$$

Another particular case is

$$\psi^{a_1 \ldots a_n}_{/i_1 \ldots i_n} = \psi^{a_1}_{/i_1} \ldots \psi^{a_n}_{/i_n} \tag{2.20}$$

in accordance with the notation for $\psi^{ab}_{/ij}$ in equations (2.15)-(2.17) above.

## 2.3. Invariant formulation and geometrical considerations

We return to the problem of lack of tensorial invariance. This can be evaded by using the *Stratonovich calculus*; the second-order nature of (2.6)-(2.7) is avoided by defining the Stratonovich integral

$$\int Y \, \mathrm{d}_S X = \int Y \, \mathrm{d}X + \frac{1}{2} [Y , X] \quad \text{for } X, Y \in \mathcal{S} \tag{2.21}$$

which has a first-order or tensorial transformation rule. This has proved of great use in stochastic differential geometry, finding application for example to harmonic maps in Kendall [41], [43], [44]. Stratonovich calculus is enormously convenient in providing intrinsic formulations of geometric constructions. Nevertheless even when using Stratonovich calculus it is still necessary to resort to Itô calculus when a detailed analysis is required of the behaviour of a random process, because Itô calculus provides a direct link with probability via the "signal-plus-noise" decomposition referred to above. For this reason in the computer algebra treatment referred to below it turns out to be preferable to deal directly with Itô rather than Stratonovich calculus, because otherwise the delay in the inevitable resort to Itô calculus renders the algebraic computations susceptible to "intermediate expression swell" as the computer algebra package expands Stratonovich differential upon Stratonovich differential. So, Stratonovich calculus notwithstanding, it is still desirable to find a treatment of the Itô differential $\mathrm{d}X$ which is invariant in some extended sense.

What is to be done? It is desirable to understand $\mathrm{d}X$ invariantly, not only for purely abstract reasons but also to allow geometric reasoning to help tame the complexity of (2.16). It also turns out the invariant approach can be viewed as underlying a successful computer algebra approach to calculations involving expressions such as (2.16).

The key is *Schwartz' principle* (Emery [30], Meyer [53]-[54], Schwartz [63]-[64]): view the differential of $X$ as a (formal) *second-order tangent vector* $\mathbf{d}\boldsymbol{X} = \mathrm{d}X$, concealing its bracket within itself via

$$\mathbf{d}\boldsymbol{X} = \begin{pmatrix} \mathrm{d}X \\ (1/2)\mathrm{d}[X\,,\,X] \end{pmatrix}. \tag{2.22}$$

Here if $X$ is a semimartingale in $\mathbb{R}^n$ then $[X\,,\,X]$ is viewed as a matrix-valued process in $\mathbb{R}^n \otimes \mathbb{R}^n$. So $\mathbf{d}\boldsymbol{X}$ is equipped with a second-order transformation rule

$$\begin{aligned}
\mathbf{d}f(\boldsymbol{X}) &= \begin{pmatrix} \mathrm{d}\big(f(X)\big) \\ (1/2)\mathrm{d}\left[\big(f(X)\big)\,,\,\big(f(X)\big)\right] \end{pmatrix} \\
&= \begin{pmatrix} f'(X) & f''(X) \\ 0 & f'(X)\otimes f'(X) \end{pmatrix} \begin{pmatrix} \mathrm{d}X \\ (1/2)\mathrm{d}[X\,,\,X] \end{pmatrix} \\
&= \overset{2}{\mathbf{f}}(X)\cdot \mathbf{d}\boldsymbol{X}
\end{aligned} \tag{2.23}$$

where (in the terminology of Schwartz *et al.*)

$$\overset{2}{\mathbf{f}} = \begin{pmatrix} f' & f'' \\ 0 & f'\otimes f' \end{pmatrix}$$

is the *second-order form* associated to $f$. That is to say, the components of the top row of $\overset{2}{\mathbf{f}}$ transform according to

$$\begin{aligned}
f_{/a} &= f_{/i}\omega^{i}_{/a} \\
f_{/ab} &= f_{/ij}\omega^{ij}_{/ab} + f_{/i}\omega^{i}_{/ab}
\end{aligned} \tag{2.24}$$

where $\omega$ denotes a coordinate system, while the non-zero component on the bottom right can be deduced from the top left component via a squaring operation. More generally if

$$\mathbf{F} = \begin{pmatrix} f_i & f_{ij} \\ 0 & f_i \otimes f_i \end{pmatrix}$$

transforms in the same way, so that

$$\begin{aligned} f_a &= f_i \omega^i_{/a} \\ f_{ab} &= f_{ij} \omega^{ij}_{/ab} + f_i \omega^i_{/ab}, \end{aligned}$$

Then $\mathbf{F}$ is a second-order form in the terminology of the second-order approach to stochastic calculus (see the papers cited above for more on second-order forms). In the string terminology used in this paper, $\overset{2}{\mathbf{f}}$ and $\mathbf{F}$ possess the invariant character of a *costring field of length* 2 while at least formally $\mathbf{d}\boldsymbol{X}$ possesses the invariant character of a *contrastring field of length* 2. The effect of the transformation laws (2.23)-(2.24) is that if $\mathbf{F}$ is a second-order form and $\mathbf{d}\boldsymbol{X}$ is a stochastic differential as above then the integral $\int \mathbf{F} \cdot \mathbf{d}\boldsymbol{X}$ makes intrinsic sense, giving the same result whatever coordinate system is used for the calculations.

Note how (2.23) encapsulates equations (2.16)-(2.17) on multiplying out the matrices.

In this way we may now work with spaces $\mathcal{D}^2(\mathbb{M})$ of stochastic differentials on a manifold $\mathbb{M}$, with $\mathbf{d}\boldsymbol{X} \in \mathcal{D}^2(\mathbb{M})$ understood *formally* as a second-order tangent vector sitting above its semimartingale $X \in \mathcal{S}(\mathbb{M})$. The formal understanding is made precise by integrating $\mathbf{d}\boldsymbol{X}$ against second-order forms (costring fields of length 2) $\mathbf{F}$ sitting above $X$, with (2.23) showing that when $\mathbf{F}$ is "exact" ($\mathbf{F} = \overset{2}{\mathbf{f}}$ for some smooth $f$) then $\int \mathbf{F} \cdot \mathbf{d}\boldsymbol{X} = \int \mathrm{d}f(X)$ behaves as expected.

This approach has been used by Schwartz, Meyer and Emery to investigate the invariant nature of stochastic differential equations, and most recently to elucidate notions of transfer from classical to stochastic differential geometry.

Note that the Bracket operation now picks out (twice the value of) the second component of $\mathbf{d}\boldsymbol{X}$ as given in the vector form of equation (2.22). Formally $\mathrm{Bracket}(\mathbf{d}\boldsymbol{X}, \mathbf{d}\boldsymbol{X})$ is a tensorial object sitting above $X$. Against it we may integrate conventional first-order forms sitting above $X$. (However these forms must take values in $T^*\mathbb{M} \otimes T^*\mathbb{M}$ since $\mathrm{Bracket}(\mathrm{d}X, \mathrm{d}X)$ is to be understood as a time differential taking values in $T\mathbb{M} \otimes T\mathbb{M}$. For example if $g$ is a Riemannian metric for $\mathbb{M}$ then we can make sense of $\int g(X) \cdot \mathrm{Bracket}(\mathrm{d}X, \mathrm{d}X)$, which is a process of great importance in stochastic differential geometry. Emery [30, 5.2] refers to it as the *Riemannian quadratic variation*, and it measures an "intrinsic time" for the semimartingale $X$.)

Note finally that $\mathrm{Drift}(\mathrm{d}X)$ on its own *still* does not make invariant sense: we want it to live above $X$ in $\mathcal{D}^1(\mathbb{M})$ and yet the formula

$$\mathrm{Drift}\, \mathrm{d}f(X) = f'(X)\mathrm{Drift}(\mathrm{d}X) + \frac{1}{2} f''(X) \cdot \mathrm{Bracket}(\mathrm{d}X, \mathrm{d}X) \tag{2.25}$$

shows that it partakes of a second-order nature. The way to deal with this is to impose additional structure on $\mathbb{M}$ in order to allow the elimination of the second-order part of (2.25). We suppose specified a (symmetric) *connection* on $\mathbb{M}$, expressed in a given coordinate system by the array of *Cristoffel symbols* $[\Gamma^i_{jk}]$. If $\omega$, $\psi$ are two smooth coordinate systems then the Christoffel symbols $\Gamma^i_{jk}$ transform by

$$\Gamma^a_{bc} = \Gamma^i_{jk} \psi^a_{/i} \omega^{jk}_{/bc} + \psi^a_{/i} \omega^i_{/bc} \tag{2.26}$$

and, the connection being symmetric, they are required also to satisfy the symmetry condition

$$\Gamma^i_{jk} = \Gamma^i_{kj} \,. \tag{2.27}$$

For future reference note that a *nonsymmetric* connection has Christoffel symbols $\Gamma^i_{jk}$ transforming according to equation (2.26) but *not* satisfying the symmetry condition (2.27). Ikeda & Watanabe [36, discussion preceding Remark 4.2, chap. V] use nonsymmetric connections in a stochastic differential geometry representation of diffusions with smooth elliptic coefficients: see also the comments in Kendall [40].

We may now define the "intrinsic drift" of $\mathrm{d}X$ by

$$\mathrm{IDrift}(\mathrm{d}X)^i = \mathrm{Drift}(\mathrm{d}X^i) + \frac{1}{2} \Gamma^i_{jk}\, \mathrm{Bracket}(\mathrm{d}X^j, \mathrm{d}X^k) \,. \tag{2.28}$$

It may be checked that IDrift is an $\mathcal{H}$-linear map

$$\mathrm{IDrift} : \mathcal{D}^2(\mathbb{M}) \to \mathcal{D}^1(\mathbb{M}) \,. \tag{2.29}$$

Specification of the intrinsic drift in a way which depends smoothly on the location of the process $X$ is equivalent to choice of a symmetric connection.

In practice one often specifies Christoffel symbols and hence a connection by asserting that one of two special cases applies; either that a particular coordinate system is *flat* (the $\Gamma^i_{jk}$ all vanish in that system) or that the connection is the *Lévi-Civita connection* for a specified Riemannian metric. If one is considering a strictly elliptic diffusion with smooth coefficients then it is often natural to use the metric formed by the inverse of the matrix of the (second-order) diffusion coefficients.

In any case one may now specify Bracket$(\mathbf{d}\boldsymbol{X}, \mathbf{d}\boldsymbol{X})$ and IDrift$(\mathbf{d}\boldsymbol{X})$, by requiring them to satisfy intrinsic equations involving the semimartingale $X$ itself and perhaps other exogeneous random processes. This amounts to the specification of an intrinsic *stochastic differential equation*, directly generalizing the idea of a classical differential equation for an ordinary dynamical system. Just as the classical differential equation specifies the velocity (which is to say, the trend) of a particle given its configuration, so IDrift delivers the infinitesimal trend of $\mathrm{d}X$ and Bracket the infinitesimal variance of the noise part of $\mathrm{d}X$. Moreover the analogy extends to a uniqueness and existence theorem with proof generalizing the classical case: the stochastic differential equation can be solved (at least locally) when the intrinsic drift and the bracket depend on the configuration in a Lipschitzian fashion. (In fact the full theory is rather subtle. There are several different kinds of uniqueness available of various strengths, and conditions for existence and uniqueness can be considerably relaxed depending on the precise flavour of uniqueness required. See the stochastic calculus monographs cited above for a full discussion.)

Of course in practice the specification is often presented in terms of a particular coordinate system, using Drift rather than IDrift. It should also be noted that the above describes *weak-sense* stochastic differential equations; applications often involve *strong-sense* stochastic differential equations in which the noise part of $X$ is required to be directly related to specific Brownian motions or even more general semimartingales. In terms of the above this corresponds to also specifying equations for brackets Bracket$(\mathbf{d}\boldsymbol{X}, \mathbf{d}\boldsymbol{Y})$ for various semimartingales $Y$.

We may now translate (SC6) into a more intrinsic terminoloy: the theory of stochastic differential equations tells us that the statistical behaviour of a multivariate semimartingale $X$ may be determined by specifying

- its initial point $X(0)$,
- a formula for the bracket operation Bracket as it applies to $\mathbf{d}\boldsymbol{X}$,

- a connection for the ambient space of $X$,
- and a formula for the intrinsic drift IDrift($\mathbf{d}X$) of $\mathbf{d}X$.

One is free to choose the connection to clarify the structure, and this leads to various possibilities for description or specification of the behaviour of $X$. Using the extrinsic drift Drift($\mathbf{d}X$) corresponds to choosing the flat connection associated to a particular preferred coordinate system. In many cases it is convenient to use Bracket to determine a Riemannian structure (if $X$ is a strictly elliptic diffusion) and then to use the Lévi-Civita connection. Thus one perhaps specifies $X$ in a fixed flat coordinate system and then searches for a new geometry and new coordinate system leading to enlightening presentation of the behaviour of $X$. This is the basis for the applications of computer algebra to the stochastic calculus of shape, mentioned in the following section. See also Antonelli, Chapin & Voorhees [5] for an application in a genetic context.

## 2.4. Computer algebra and stochastic calculus

Another response to the complexity inherent in (2.13) is to implement stochastic calculus in a computer algebra package, for example the `ito` collection of procedures (also described as *symbolic Itô calculus*) programmed in `REDUCE` and discussed in Kendall [42], [45]-[47]. It is striking, though in retrospect inevitable, that the implementation turns out to have strong resemblances to the approach via second-order geometry given above. The implementation consists of augmenting the computer algebra package with a new type of variable and a number of operations. The new type of variable is of course a (scalar) stochastic differential `dX`, to which is attached a (scalar) semimartingale `X` serving as its primitive. The basic procedures are

- `Introduce(X,dX)` introducing $\mathrm{d}X$ to the system as a basic stochastic differential,
- `d(Y)` returning an expression for the stochastic differential represented by `Y`,
- `Drift(dY)` computing the drift of the stochastic differential represented by `dY`,
- `Add!_Drift(dX,f*dt)` adding to a substitution list the information that the stochastic differential represented by `dX` has drift represented by `f*dt`.

The implementation of `d(Y)` uses Itô's formula and hence ensures that stochastic differentials such as `dX` transform formally as second-order tangent vectors, as they should.

In the current implementation (`ito` version 3 or `itovsn3`, as described in Kendall [45], [47]) there is no bracket operation. Instead the pure second-order structure is implemented as a sequence of `LET` rules or rewrite rules such as

$$\texttt{LET dB**2} = \texttt{dt}\,. \tag{2.30}$$

The above implements *scalar* stochastic differentials. Multivariate stochastic differentials may be specified coordinatewise in some convenient set of coordinates. One means of investigation (used for example in Kendall [42], [46]) corresponds to examining the second-order differential in various coordinate systems endowed with appropriate connections, to search for specific representations displaying useful and informative structure. In Kendall [42] the shape diffusion of three Brownian points in Euclidean $n$-space is thereby represented as Brownian motion on a hemisphere together with an intrinsic drift depending on the dimension $n$ in a simple way. In Kendall [46] the general shape diffusion is thereby represented as a skew product of a drifting Brownian motion on a sphere together with a rotation-group Brownian motion whose statistics depend on the spherical diffusion.

The principal work of Kendall [46] involved overcoming the difficulty of representing symbolic sums in REDUCE. The result is effective but somewhat untidy. It is an interesting and possibly fruitful question whether a still closer implementation of the structures of second-order stochastic differential geometry might lead to a more transparent and efficient solution. Implementation questions of this kind become yet more pressing when one considers the much less well-developed computer algebra of statistical asymptotics.

## 3 Geometry of statistical asymptotics

Our second example of higher-order calculus arises from the differential geometry approach to statistical inference. Differential geometric concepts and ideas are of interest in connection with various aspects of statistical inference and in particular with statistical asymptotics. This is mainly because it is often preferable that statistical procedures for parametric and semiparametric models should be invariant, in the sense that the conclusions reached should be independent of the particular parametrization chosen. It turns out that invariant formulations require generalization of the tensorially invariant calculus of differential geometry in a manner which is closely related to the second-order calculus developed for stochastic calculus and described in the previous section. The resulting "higher-order calculus" also turns out to be helpful in simplifying, understanding and managing the complex expressions which occur in statistical asymptotics.

Furthermore it is desirable to develop corresponding computer algebra programs and this has been found to require certain nontrivial refinements of existing packages for symbolic mathematical calculations (Kendall [48]).

The developments in statistical asymptotics which prompted the formulation of higher-order calculus have their root in a paper by McCullagh & Cox [52]. These authors wished to obtain a better understanding of the character of a certain statistical quantity, the *Bartlett adjustment*, by expressing it as a sum of invariant terms, the idea being that each of these invariant terms would have a simple interpretation. As a step toward the decomposition of the Bartlett adjustment they constructed certain tensors which can be viewed as higher order "symmetric tensorial derivatives" of a particular scalar function, namely the log likelihood function.

The McCullagh-Cox construction was framed in a particular setting, that of "expected likelihood geometry". The questions then arose of whether a similar construction would be possible in "observed likelihood geometry" and, more generally, of what was the mathematical essence underlying their construction. These questions were addressed in Barndorff-Nielsen [7], in which was presented a general definition of *symmetric tensorial derivatives* in a purely differential geometric framework. This definition rests on several other definitions, in particular that of *connection string fields*, which generalize the concept of (affine) connections which was discussed in section 2.3.

The symmetric tensorial derivatives are similar to higher-order covariant derivatives and, in fact, the second-order symmetric tensorial derivative of a scalar function $f$ is nothing more than the covariant derivative of the vector of ordinary first order derivatives (or *Hessian*) of $f$ with respect to a torsion-free connection. However this identification does not hold at higher-order; in contrast to higher-order covariant derivatives the symmetric tensorial derivatives are symmetric in the indices.

Extensions of the concepts of connection string fields and of symmetric tensorial derivatives were proposed in Barndorff-Nielsen & Blæsild [10]-[11] leading to a theory of *derivative strings*. A further theory of *differential strings* has subsequently been developed by Blæsild & Mora [19]. These and further extensions will be surveyed in section 4. Here we restrict consideration to fields of connection strings, fields of scalar strings, and symmetric tensorial derivatives. We will consider all these objects as multi-arrays defined on and varying smoothly over the differentiable manifold. Their geometric nature is determined by the transformation law that specifies how the multi-arrays change under change of coordinates on the manifold.

In the following we first review the basic statistical concepts providing a framework for the McCullagh-Cox construction, then we describe the mathematical developments in the geometry of statistics which are motivated by this construction and their formulation in terms of Taylor string theory, and finally we indicate how these developments may be applied to the theory of statistical asymptotics and consider briefly the prospects for implementation within a computer algebra package.

## 3.1. Elements of likelihood-based inference

The statistical framework to be considered is that of a parametric statistical model with model function $p(x;\omega)$. Here $x$ denotes the data and $\omega$ is a $d$-dimensional parameter whose value determines the probability measure $P_\omega$ which is assumed to govern the stochastic behaviour of $x$. The probability measure $P_\omega$ is determined as follows: $x \mapsto p(x;\omega)$ is it Radon-Nikodym derivative with respect to some fixed $\sigma$-finite nonnegative reference measure $\mu$.

In the statistical theory of likelihood the fundamental tool for inference is the *log likelihood function* $l(\omega)$, defined by $l(\omega) = l(\omega;x) = \log p(x;\omega)$. In fact it is convenient to view the log likelihood as determined only up to an arbitrary additive constant which may depend on the given data $x$ but which does not depend on $\omega$. Any function differing from $\log p(x;\omega)$ by an additive term depending only on the data is said to be the log likelihood function of the model, and all likelihood-based procedures of inference are defined so as to be independent of the choice of version of $l(\omega)$.

Two key concepts which are intimately linked to likelihood are those of sufficiency and ancillarity. There is not scope here for discussion of these concepts, but consideration of them leads to a view of the log likelihood function as depending on the data through a pair of statistics $(\widehat{\omega}, a)$ such that:

(L1) $\widehat{\omega} = \widehat{\omega}(x)$ is the *maximum likelihood estimator* (that value of $\omega$ which maximizes the log likelihood),

(L2) $a = a(x)$ is a distribution constant statistic (the distribution of $a$ does not depend on the parameter $\omega$),

(L3) $(\widehat{\omega}, a)$ is a *sufficient statistic* (there is a version of the log likelihood function which depends on $x$ only through $(\widehat{\omega}, a)$, so that $\log p(x;\omega) = l(\omega; \widehat{\omega}, a) + c(x)$ for some function $c(x)$ of the data $x$ alone).

We will assume regularity conditions obtain which amongst other matters ensure that the maximum likelihood estimator $\widehat{\omega}$ exists and is unique, that the likelihood function is a smooth function of the parameter, and that $\widehat{\omega}$ is the unique solution of the *likelihood equation* $s = 0$, where the *score statistic* $s$ is given by $s = dl = l_*$ (the one-form or vector of derivatives with respect to compoments of $\omega$). A statistic $a$ with the properties (L2) and (L3) above is said to be *ancillary*. Generally speaking, ancillary statistics may not be unique and need not exist, but we shall suppose a statistic $a$ to be chosen which is at least approximately ancillary, in the sense that (L2) and (L3) hold approximately. In subsequent arguments the chosen ancillary will be considered to be held at a fixed value (by conditioning). This will allow us to use $\widehat{\omega}$ to give a one-to-one correspondence between the parameter space and that slice of the data space corresponding to the observed value of $a$.

The probabilistic basis for inference is the statistical law of $\widehat{\omega}$ under conditioning by the value of the ancillary statistic $a$. We denote the model function of this conditional statistical law by $p(\widehat{\omega}; \omega \mid a)$. Equivalently, one may work with the conditional law of the score $s$. The model function of this law is related to that of $\widehat{\omega}$ by the relation

$$p(s; \omega \mid a) = \left| l_{*;*} \right|^{-1} p(\widehat{\omega}; \omega \mid a)$$

where $l_{*;*}$ denotes the Jacobian matrix of the transformation from $\widehat{\omega}$ to $s$ (assumed one-to-one and smooth as part of our global regularity assumptions). We shall return to consideration of $p(s; \omega \mid a)$ after introducing *yoke geometry*, the *observed* and *expected likelihood yokes* built from the log-likelihood, and the higher-order calculus notation of Taylor string theory.

## 3.2. Geometry of statistics: yokes and Taylor string fields

First consider the definition of a general yoke. Let $\Omega$ be a manifold and let $\Omega'$ denote a copy of $\Omega$. In statistical applications $\Omega$ will be the parameter space and $\Omega'$ will be the domain of variation of the maximum likelihood estimator. (Note that this $\Omega$ is the parameter space, in contrast to the sample space $\Omega$ of the triple $(\Omega, \mathcal{F}, \mathbb{P})$ referred to once at the start of section 2 and underlying the discussion there. The clash of traditional notation is unfortunate but should cause no confusion.) We consider functions $g = g(\omega;\omega')$ defined on the product space $\Omega \times \Omega'$, and using local coordinates $(\omega^1, \ldots, \omega^d)$ we write

$$g_{I;J} = g_{i_1 \ldots i_m\, ;\, j_1 \ldots j_n}(\omega;\omega') = \partial_{i_1} \ldots \partial_{i_m} \partial'_{j_1} \ldots \partial'_{j_n} g(\omega;\omega') \tag{3.1}$$

(where $\partial_i = \partial/\partial\omega^i$, $\partial'_j = \partial/\partial(\omega')^j$, and $I$, $J$ are finite sequences of indices $I = (i_1 \ldots i_m)$, $J = (j_1 \ldots j_n)$), and

$$\mathfrak{g}_{I;J} = g_{I;J}(\omega;\omega)\,. \tag{3.2}$$

This last notation will be used more generally later: the gothic version of a symbol (letter) for a function on $\Omega \times \Omega'$ indicates the restriction of that function to the diagonal of $\Omega \times \Omega'$. The function $g$ is said to be a *yoke* if it satisfies the following two conditions for every $\omega \in \Omega$:

(Y1) $\mathfrak{g}_{i;} = g_{i;}(\omega;\omega) = 0$,

(Y2) the matrix $[\mathfrak{g}_{ij;}] = \left[g_{ij;}(\omega;\omega)\right]$ is non-singular.

Repeated differentiation of the equation $\mathfrak{g}_{i;} = 0$ in (Y1) yields the sequence of relations

$$\mathfrak{g}_{ij;} + \mathfrak{g}_{i;j} = 0 \tag{3.3}$$

$$\mathfrak{g}_{ijk;} + \mathfrak{g}_{ij;k} + \mathfrak{g}_{ik;j} + \mathfrak{g}_{i;jk} = 0 \tag{3.4}$$

and so forth, with the general form being

$$\mathfrak{g}_{I;} + \sum_{I/2} \mathfrak{g}_{I_1;I_2} = 0 \tag{3.5}$$

where we have used the notation for summing over ordered partitions introduced in section 2.2 ((2.19) and (PS1)-(PS3)).

Note that (3.3) and condition (Y2) imply that the matrix $[\mathfrak{g}_{i;j}]$ is symmetric and nonsingular, even though the function $g$ has *not* been assumed to be symmetric. This is crucial to the development of a general theory of differential geometries derived from yokes, which we now describe. Any yoke induces a collection of geometrical objects on $\Omega$, including a pseudo-Riemannian metric given by the symmetric non-singular tensor $\mathfrak{g}_{i;j}$ and a family of symmetric connections $\{\overset{\alpha}{\Gamma} : \alpha \in \mathbb{R}\}$, where $\overset{\alpha}{\Gamma}$ is given in terms of lowered Christoffel symbols by

$$\overset{\alpha}{\Gamma}_{ijk} = \frac{1+\alpha}{2}\,\mathfrak{g}_{ij;k} + \frac{1-\alpha}{2}\,\mathfrak{g}_{k;ij}\,. \tag{3.6}$$

In particular

$$\overset{1}{\Gamma}_{ijk} = \mathfrak{g}_{ij;k} \tag{3.7}$$

and

$$\overset{-1}{\Gamma}_{ijk} = \mathfrak{g}_{k;ij}\,. \tag{3.8}$$

We shall be working with the generalization of (3.6) given by

$$\overset{\alpha}{\Gamma}_{j;k_1\ldots k_n} = \frac{1+\alpha}{2}\,\mathfrak{g}_{k_1\ldots k_n;j} + \frac{1-\alpha}{2}\,\mathfrak{g}_{j;k_1\ldots k_n}\,. \tag{3.9}$$

Lifting the index $j$ by means of $\mathfrak{g}^{ij}$, the inverse tensor to $\mathfrak{g}_{i;j}$ these quantities are converted to the quantities

$$\overset{\alpha}{\Gamma}{}^{i}_{k_1\ldots k_n} = \frac{1+\alpha}{2}\,\mathfrak{g}^{;i}_{k_1\ldots k_n} + \frac{1-\alpha}{2}\,\mathfrak{g}^{i}{}_{;k_1\ldots k_n}\,. \tag{3.10}$$

Here

$$\mathfrak{g}^{;i}_{k_1\ldots k_n} = \mathfrak{g}_{k_1\ldots k_n;j}\mathfrak{g}^{ij} \tag{3.11}$$

$$\mathfrak{g}^{i}{}_{;k_1\ldots k_n} = \mathfrak{g}^{ij}\mathfrak{g}_{j;k_1\ldots k_n}\,. \tag{3.12}$$

Note for future reference that the $\overset{1}{\Gamma}{}^{i}_{k_1\ldots k_n}$ are the coefficients of the Taylor series expansion about $\omega$ of

$$g^{;i}(\tilde{\omega};\omega) : \tilde{\omega} \mapsto g_{;j}(\tilde{\omega};\omega)\mathfrak{g}^{ij}(\omega)\,,$$

while the $\overset{-1}{\Gamma}{}^{i}_{k_1\ldots k_n}$ play the same role for the Taylor series expansion about $\omega$ of

$$g^{i}(\omega;\tilde{\omega}) : \tilde{\omega} \mapsto \mathfrak{g}^{ij}(\omega)g_{j;}(\omega,\tilde{\omega})\,.$$

We shall see below that the sequence of arrays $\{[\overset{\alpha}{\Gamma}{}^i_{k_1\ldots k_n}] : n = 1, 2, \ldots\}$ is an example of a *connection string*.

It is of interest to observe that on introducing the normalized yoke

$$\bar{g}(\omega;\omega') = g(\omega;\omega') - g(\omega';\omega'), \tag{3.13}$$

which is again a yoke, and defining $h(\omega;\omega')$ by

$$h(\omega;\omega') = \bar{g}(\omega';\omega) \tag{3.14}$$

we have that $h$ is also a yoke. Indeed $h$ is a dual to $g$ in the sense that

$$\left\{\begin{array}{l} \mathfrak{h}_{i;j} = \mathfrak{g}_{j;i} = \mathfrak{g}_{i;j} \\ \mathfrak{h}_{i;jk} = \mathfrak{g}_{jk;i} \\ \mathfrak{h}_{ij;k} = \mathfrak{g}_{k;ij} \\ \qquad\qquad \textit{et cetera}\ldots \end{array}\right\} \tag{3.15}$$

so that the Riemannian metrics are the same while the connection families correspond under the bijection $\alpha \leftrightarrow 1-\alpha$.

Returning to the statistical framework, let $\Omega$ be the parameter space of the statistical model $p(x;\omega)$ and, as above, denote by $l(\omega) = l(\omega;x)$ the function of $\omega$ which is the corresponding log likelihood function. Under standard mild regularity conditions, the function given by

$$g(\omega;\omega') = E_{\omega'}\big[l(\omega;x) - l(\omega';x)\big] \tag{3.16}$$

is a yoke, the *expected likelihood yoke*. Furthermore, the function

$$g(\omega;\omega') = l(\omega\,;\omega',a) - l(\omega'\,;\omega',a) \tag{3.17}$$

is also a yoke, the *observed likelihood yoke*. Both the observed likelihood yoke and the expected likelihood yoke are normalized. The *observed likelihood geometries* and *expected likelihood geometries* are given by the metric tensors, connections, and higher-order string objects derived from the general yoke geometries discussed in the foregoing using, respectively, the observed and expected likelihood yokes. Note the use of the ancillary statistic in the definition of observed geometry.

Note, incidentally, that formula (3.3) specializes to the identity between the two well-known forms for the expected information when $g$ is taken to be the expected likelihood yoke.

It turns out to be the case that observed likelihood geometry is "more geometrical" than expected likelihood geometry, since the integrations involved in construction of the latter can obscure geometric structure. This enhances the intimate relationship between the theories of statistical inference and differential geometry, because observed likelihood quantities are also more directly related to the likelihood function (and hence of a more basic *statistical* nature) than the corresponding expected likelihood quantities.

Before proceeding with statistical issues it is now convenient to introduce the concepts of scalar and connection string fields, motivated by the need for a convenient notation to use in discussion of such expressions as arise in equations (3.10)-(3.12). A *scalar string field* is defined as a (finite or infinite) sequence of multi-arrays with entries $f_{k_1\ldots k_t}$ depending on $\omega \in \Omega$

$$\begin{aligned} \mathbf{f} &= \left\{[f_K] = [f_{k_1\ldots k_t}] : |K| = t = 1,\, 2,\, \ldots,\, t_{\max}\right\} \\ &= \left\{[f_{k_1}],\, [f_{k_1 k_2}],\, \cdots [f_{k_1\ldots k_t}],\, \cdots\right\}, \end{aligned}$$

where the *length* $t_{\max} = \left|K\right|_{\max}$ of the string $\mathbf{f}$ is a positive integer or infinity, and where $\mathbf{f}$ is required to satisfy the following transformation law:

$$f_C = \sum_{\tau=1}^{|C|} \sum_{K:|K|=\tau} f_K \omega^K_{/C} = f_K \omega^K_{/C}\,. \tag{3.18}$$

Here $|K|$ is the length of the sequence $K$ of dummy suffices, and the right-hand formula uses the *extended summation convention*: since the symbol $K$ is repeated the formula is summed as $K$ runs over all finite sequences of length $|K| = 1, 2, \ldots$ (recall from (PS3) in section 2.2 that $\omega^K_{/C} = 0$ if $|K| > |C|$). We adopt the convention that if $t_{\max} = \infty$ then the notation $1, 2, \ldots, t_{\max}$ is to be interpreted as representing the sequence $1, 2, \ldots$ of all positive intergers, not the set $1, 2, \ldots, \infty$.

This transformation law is obeyed by the sequence of derivatives of a scalar function $f$; if $f$ is defined on a domain $\Omega$ and has $|K|$-fold partial derivatives forming an array $[f_{/K}]$ then (3.18) is satisfied if we write $f_K = f_{/K}$. Not all scalar string fields are generated as the multi-array of derivatives of a scalar function (this generalizes the well-known fact that not all 1-forms arise as derivatives of functions). For example if $g$ is a yoke then the sequence of arrays $\left\{[\,\mathfrak{g}_{k_1\ldots k_n}],\, n = 1, 2, \ldots\right\}$, constitutes a scalar string, but the component arrays are not in general the derivatives of a single scalar function.

A *connection string field* is defined as a (finite or infinite) sequence

$$\begin{aligned}\mathbf{\Gamma} &= \left\{[\Gamma^i_K] : |K| = t = 1,\, 2,\, \ldots,\, t_{\max}\right\} \\ &= \left\{[\Gamma^i_{k_1}],\, [\Gamma^i_{k_1 k_2}],\, \ldots [\Gamma^i_{k_1 \ldots k_t}],\, \ldots\right\},\end{aligned}$$

where the length $t_{\max} = \left|K\right|_{\max}$ of the connection string field $\mathbf{\Gamma}$ is a positive integer or infinity, and where the component matrices of $\mathbf{\Gamma}$ are required to satisfy the following transformation law:

$$\Gamma^a_C = \sum_{\tau=1}^{|C|} \sum_{|K|=\tau} \psi^a_{/i} \Gamma^i_K \omega^K_{/C} = \psi^a_{/i} \Gamma^i_K \omega^K_{/C} \,. \tag{3.19}$$

Here again the extended summation convention is employed in the right-hand formula. We shall employ the extended summation convention without further comment in the remainder of this paper, except where indicated otherwise. The connection string field is said to be *invertible* if the matrix formed by the array at length 1 (which is to say, the first array $[\Gamma^i_{k_1}]$ in the sequence $\left\{[\Gamma^i_{k_1}],\, [\Gamma^i_{k_1 k_2}],\, \ldots [\Gamma^i_{k_1 \ldots k_t}],\, \ldots\right\}$ ) is invertible.

The indices in the sequences $C$ and $K$ in (3.18)-(3.19) above are referred to as *structural indices* because they typically require higher-order derivatives in the transformation law, as opposed to tensorial indices such as $i$ in (3.19) above. The connection string is said to be *symmetric* if its component multi-arrays are symmetric in their structural indices.

A (classical) connection (not necessarily symmetric) having Christoffel symbols $\Gamma^i_{k_1 k_2}$ may be viewed as an element of a short connection string $\mathbf{\Gamma}$ of length 2; define $\Gamma^i_k$ to be given by the Kronecker delta tensor $\delta^i_k$ and let $\mathbf{\Gamma}$ be given by the sequence of two arrays $\left\{[\Gamma^i_{k_1}]\, ,\, [\Gamma^i_{k_1 k_2}]\right\} = \left\{[\delta^i_{k_1}]\, ,\, [\Gamma^i_{k_1 k_2}]\right\}$.

Yokes can be used to produce connection string fields as indicated above; if $g$ is a yoke then the set of array $\left\{[\overset{\alpha}{\Gamma}{}^i_{k_1 \ldots k_n}] : n = 1,\, 2,\, \ldots\right\}$ (as described in equation (3.10)) forms a connection string field $\overset{\alpha}{\mathbf{\Gamma}}$ of infinite length. Note that such a connection string field is symmetric. It is also invertible because of property (Y2).

Given a scalar function $f$ and an invertible connection string field $\mathbf{\Gamma}$ it is possible to define *tensorial derivatives* of $f$ relative to $\mathbf{\Gamma}$ by the operation of *intertwining of strings* (note there is no obvious connection with the concept of intertwining of group representations). The tensorial derivatives are indeed tensors, and are determined in coordinates as certain linear combinations of the ordinary higher partial derivatives of $f$. Specifically, the tensorial derivative of $f$ with respect to $\omega^{i_1}, \ldots, \omega^{i_t}$, denoted by $f_{//I} = f_{//i_1 \ldots i_t}$, is defined implicitly by the system of equations

$$f_{/K} = f_{//I}\Gamma^I_K\,, \tag{3.20}$$

where the symbols $\Gamma^I_K$ are defined, analogously to (2.19), by

$$\Gamma^I_K = \sum_{K/|I|} \Gamma^{i_1}_{K_1} \ldots \Gamma^{i_{|I|}}_{K_{|I|}}\,. \tag{3.21}$$

(When $|K| < |I|$ then we follow (PS3) in defining $\Gamma^I_K = 0$.)

Note that the transformation law (3.19) under changes of coordinates for the arrays $[\Gamma^I_K]$ contains as a special case the transformation law for Christoffel symbols (2.26), if the array of Christoffel symbols $[\Gamma^i_{k_1k_2}]$ is augmented by the Kronecker delta tensor $[\delta^i_k]$ to form a connection string field of length 2 as above. More generally we define a *special connection string field* $\Gamma$ to be a connection string field such that $\Gamma^i_k = \delta^i_k$. For simplicity we assume all connection string fields are special in the rest of this section. Essentially this is no restriction and under this assumption the $\Gamma^i_{k_1k_2}$ constitute the Christoffel symbols of a connection $\nabla$ on the domain $\Omega$, with the connection being symmetric if the special connection string is symmetric.

The first few tensorial derivatives of $f$ are then given explicitly by

$$\begin{aligned}
f_{//k_1} &= f_{/k_1} \\
f_{//k_1k_2} &= f_{/k_1k_2} - \Gamma^i_{k_1k_2} f_{/i} \\
f_{//k_1k_2k_3} &= f_{/k_1k_2k_3} - \Gamma^i_{k_1k_2} f_{/ik_3}[3] - \left(\Gamma^i_{k_1k_2k_3} - \Gamma^i_{jk_1}\Gamma^j_{k_2k_3}[3]\right) f_{/i}\,.
\end{aligned} \tag{3.22}$$

In the last equation the symbol [3] indicates a sum of three terms determined by cyclic permutation of the structural indices $k_1$, $k_2$, $k_3$. The formulae (3.22) show, in particular, that the second order tensorial derivative of $f$ equals the covariant derivative of $f_{/k}$ with respect to the connection $\nabla$. The tensorial derivatives of $f$ will be symmetric in the indices provided the arrays $[\Gamma^i_K]$ are symmetric in the sequences of structural indices $K = (k_1 \ldots k_t)$.

If the connection string is either $\overset{1}{\Gamma}$ or $\overset{-1}{\Gamma}$ derived from some yoke then the tensorial derivatives correspond to conventional derivatives evaluated in the corresponding coordinate system indicated in the remarks after equations (3.11)-(3.12). Indeed for any connection string it is possible to determine *representing coordinate systems* for which this remark is correct; see the second paragraph of section 3.3.

Note that we have already met an example of intertwining in section 2.3; consider the definition of IDrift in equation (2.28) and compare it with the second of the formulae in (3.22). The connection given by $\Gamma^i_{jk}$ is used to convert a non-tensorial quantity $\mathrm{Drift}(\mathrm{d}X)$ into a tensorial quantity $\mathrm{IDrift}(\mathrm{d}X)$ (The sign difference is attributable to the difference between differentials and functions). In the next subsection we turn to statistical applications of intertwining.

### 3.3. Invariant Taylor series expansions and statistical asymptotics

The concept of symmetric tensorial derivatives may be used together with that of yokes to define invariant Taylor expansions. Ordinary Taylor expansions are not invariant; if a function $f$ is approximated by a Taylor expansion up to and including polynomials of degree $n$, say, the approximation will depend on which coordinate system on the domain of definition of $f$ one employs. To define an invariant Taylor expansion using a yoke $g$, let

$$g^i(\omega;\omega') = \mathrm{g}^{ij}(\omega)g_{j;}(\omega;\omega') \tag{3.23}$$

(determining the second of the two coordinate systems described after equations (3.11)-(3.12)) and let

$$g^I = g^{i_1}\dots g^{i_n}\,, \quad \text{where } I = (i_1\dots i_n)\,. \tag{3.24}$$

Then $f$ may be expanded around $\omega \in \Omega$ as

$$\begin{aligned} f(\omega') &= f(\omega) + \frac{1}{|I|!} f_{//I}(\omega) g^I(\omega;\omega') \\ &= f(\omega) + \sum_{\nu=1}^{\infty} \sum_{I:|I|=\nu} \frac{1}{\nu!} f_{//I}(\omega) g^I(\omega;\omega')\,, \end{aligned} \tag{3.25}$$

where the $f_{//I}(\omega)$ are the tensorial derivatives of $f$ with respect to the connection string obtained by taking $\alpha = -1$ in (3.10) and the extended summation convention of the middle formula is written out at length in the last formula. The quantity $g^i$ behaves as a contravariant tensor in $\omega$ and as a scalar in $\omega'$, so that each of the fixed-$\nu$ terms in the last formula,

$$\frac{1}{\nu!} \sum_{I:|I|=\nu} f_{//I}(\omega) g^I(\omega;\omega'), \tag{3.26}$$

is invariant under changes of coordinates. Moreover for each fixed $\nu$ the expression (3.26) is of the same order of magnitude in $\omega' - \omega$ as the corresponding term in the ordinary Taylor expansion of $f$ in $\omega$-coordinates.

The above considerations lead us the consider representation of invertible symmetric connection string fields using coordinate systems, in a manner which is helpful for the purposes of intuition. Suppose given data of fixed $\omega \in \Omega$ and an invertible symmetric connection string field $\mathbf{\Gamma}$. It is possible to determine a special coordinate system using a local diffeomorphism $\phi_\omega : \Omega \to T_\omega\Omega$ (where $T_\omega\Omega$ is the tangent space to $\Omega$ at $\omega$ and $\phi_\omega(\omega) = o$) which *represents* $\mathbf{\Gamma}$ at $\omega$ in the following sense: working in the coordinate system provided by $\phi_\omega$ we find all of the arrays in the multi-array $\mathbf{\Gamma}$ will vanish at $\omega$ except for the first which at $\omega$ will be the Kronecker delta or identity matrix. This corresponds to the "coordinate string" approach of Murray [57]. Blæsild [17] notes an explicit construction of such $\phi_\omega$ in the case of a $(\pm 1)$-connection string field derived from a yoke, which may be deduced from our comments on Taylor series expansions following equations (3.11)-(3.12) above. In such a representing system $\phi_\omega$ the tensorial derivatives $f_{//I}$ will agree with the ordinary derivatives $f_{/I}$ at $\omega$. In particular, the invariant Taylor series of a scalar function $f$ at $\omega$ using a yoke $g$ will agree with the ordinary Taylor series of $f$ at $\omega$ computed using the $\overset{-1}{\mathbf{\Gamma}}$-representing coordinate system $\phi_\omega$. (Incidentally it is clear from the discussion in McCullagh & Cox [52], that they had in mind an intuitive representation of this type when defining the concept of a Möbius derivative as described below.)

A representing coordinate system $\phi_\omega$ is of course determined by $\omega$ and $\mathbf{\Gamma}$ only up to its $N$-jet at $\omega$, where $N$ is the (possibly infinite) length of the connection string field $\mathbf{\Gamma}$. (The $N$-jet of a function $f$ at $\omega$ is essentially the Taylor series of $f$ at $\omega$ truncated at $N^{\text{th}}$-order.) The analogy with normal coordinates is tempting but limited; systems of normal coordinates are defined as geodesic coordinate systems and do of course represent their connections (which correspond to connection strings of length 2) but are determined as functions rather than only as $N$-jets.

On the other hand, the connection string field $\mathbf{\Gamma}$ is represented by the *N-jet field* ($N$ being the length of the connection string field $\mathbf{\Gamma}$) generated by the family of coordinate systems

$$\left\{\phi_\omega : \phi_\omega \text{ represents } \mathbf{\Gamma} \text{ at } \omega,\ \omega \text{ varies through } \Omega\right\}. \tag{3.27}$$

This is essentially the *coordinate string* approach of Murray [57]. We return to this point in section 4.2.

There is a link here with the "preferred point geometry" currently under investigation by Critchley, Marriott & Salmon [23], in which $\Omega$ is endowed with a whole family of Riemannian metrics parametrized by the points of $\Omega$ itself, so that a preferred point selects a metric for $\Omega$ at least in a neighbourhood of the preferred point. We may represent at least the local aspects of a preferred point geometry by the field of orthonormal coordinate systems given by locally-defined inverses to the family of exponential maps of the various metrics:

$$\Big\{\left(\mathrm{Exp}_\omega^{(\omega)}\right)^{-1} : \mathrm{Exp}_\omega^{(\omega)} : T_\omega\Omega \to \Omega \text{ is the exponential map at } \omega$$
$$\text{based on the metric preferred by } \omega,\ \omega \text{ varies through } \Omega\Big\}. \tag{3.28}$$

This induces a unique connection string field $\mathbf{\Gamma}$ of length $n$ for each $n$ under the requirement that the field of orthonormal coordinate systems represent $\mathbf{\Gamma}$. Hence we obtain a many-to-one map from preferred point geometries to connection string fields. The map actually takes values in the subset of *special* symmetric connection string fields (see section 4.3).

We now consider two statistical examples employing invariant Taylor series and hence string theory.

Firstly, consider the sequence of log likelihood derivatives

$$l = \left\{[l_{/K}] : |K| = 1, 2, \ldots\right\}.$$

The symmetric tensorial derivatives of $l$ relative to the connection string field $\overset{1}{\mathbf{\Gamma}}$ (derived by (3.10) with $\alpha = 1$ from the expected likelihood yoke given by (3.16)) are precisely the Möbius derivatives" of the log likelihood function, introduced by McCullagh & Cox [52].

Secondly, consider the $p^*$-formula (Barndorff-Nielsen [6]). This is a generally accurate approximation to the (conditional) distribution $p(s\,;\omega \mid a)$ of the score vector $s$, which is given by

$$p^*(s\,;\omega \mid a) = c\left|\jmath(\widehat{\omega})\right|^{1/2}\left|l_{*;*}\right|^{-1} e^{l(\omega)-l(\widehat{\omega})}, \tag{3.29}$$

where $|l_{*;*}(\widehat{\omega}\,;\,\omega,a)|$ is the determinant of the matrix $\{l_{i;j}(\widehat{\omega}\,;\,\omega,a)\}$ and $\mathrm{j}(\widehat{\omega})$ is the observed information matrix $\mathrm{j} = -\{\mathfrak{l}_{ij;}\}$ (evaluated at $\omega = \widehat{\omega}$), and $c = c(\omega, a)$ is a norming constant. Applying the method of invariant Taylor expansions, using the observed likelihood yoke and the corresponding connection string field $\overset{-1}{\boldsymbol{\Gamma}}$ as determined by (3.12), one finds (Barndorff-Nielsen [9], Mora [56]) that $p^*(s\,;\,\omega \mid a)$ may be expanded around the $d$-dimensional normal distribution with mean 0 and covariance matrix $\mathrm{j}$ as follows:

$$p^*(s\,;\,\omega \mid a) \overset{\star}{=} \varphi(s;\mathrm{j})\{1 + S_1 + S_2\}, \tag{3.30}$$

where the symbol $\overset{\star}{=}$ indicates that the approximation has asymptotic error of order $O(n^{-3/2})$ under ordinary repeated sampling (an order of error meeting the requirements of many statistical applications) and where

$$S_1 = \frac{1}{6}\,\mathfrak{t}_{ijk} h^{ijk}(s;\mathrm{j}) \tag{3.31}$$

$$S_2 = \frac{1}{24}\left\{\mathfrak{t}_{ijkl} - \frac{1}{2}\,\mathfrak{t}_{ij;kl}\right\} h^{ijkl}(s;\mathrm{j}) - \frac{1}{4}\,\mathfrak{t}_{ij;kl}\,\mathrm{j}^{kl} h^{ij}(s;\mathrm{j}) + {} + \frac{1}{72}\,\mathfrak{t}_{ijk}\,\mathfrak{t}_{lmn} h^{ijklmn}(s;\mathrm{j})\,. \tag{3.32}$$

Here $S_1$ and $S_2$ are of order $n^{-1/2}$ and $n^{-1}$, respectively, under ordinary repeated sampling, and may be viewed as arising from third- and fourth-order invariance considerations. In (3.31) and (3.32) the factors $h$ are contravariant tensorial Hermite polynomials and the quantities $\mathfrak{t}$ are tensors given by formulae (3.33)-(3.35) below. Thus $S_1$ and $S_2$ are both parametrization invariant, the only non-invariant part of the right side of (3.30) being the multivariate normal density.

We have

$$\mathfrak{t}_{ijk} = -\left(\mathfrak{l}_{ijk} + \mathfrak{l}_{ij;k}[3]\right) \tag{3.33}$$

$$\mathfrak{t}_{ijkl} = -\left\{\mathfrak{l}_{ijkl} + \mathfrak{l}_{ijk;l}[4] + \frac{1}{2}\left(\mathfrak{l}_{ij;kl} + \mathfrak{l}_{ij;n}\,\mathfrak{t}_{klm}\,\mathrm{j}^{mn}\right)[6]\right\} \tag{3.34}$$

$$\mathfrak{t}_{ij;kl} = \mathfrak{l}_{ij;kl} - \mathfrak{l}_{ij;m}\,\mathrm{j}^{mn}\,\mathfrak{l}_{n;kl}\,. \tag{3.35}$$

The three tensors (3.33)-(3.35) are of a basic statistical importance. They are special cases of tensors derived, by intertwining, from a general yoke $g$ (see Blæsild [18]). The first two tensors are obtainable from the "skewness" tensor and its covariant derivative with respect to the connection $\overset{1}{\boldsymbol{\Gamma}}$.

However, as discussed by Mora [56], the tensor $t_{ij;kl}$ does not appear to be expressible in terms of known curvature tensors (as might have been supposed). It is expressible in terms of the difference at length 3 between the connection string field $\overset{1}{\Gamma}$ and the "canonical" connection string built by using as a representing system the normal geodesic coordinates corresponding to the Christoffel connection $\overset{1}{\Gamma}$. Thus it is a genuinely "string-theoretic" quantity. A further remark on interpretation of these quantities is made in section 4.4.

## 3.4. Computer algebra and statistical asymptotics

A number of workers have used computer algebra in statistical asymptotics. In particular Andrews & Stafford [4] have developed a number of procedures in the computer algebra package *Mathematica*, allowing for direct computation of Edgeworth series and Bartlett adjustments though as yet not addressing the problem of handling the Einstein summation convention. The success of `ito` in implementing stochastic calculus within a computer algebra package, and its close links with second-order stochastic calculus, are an encouragement to try to develop computer algebra approaches to statistical asymptotics which proceed by implementing yoke geometry and Taylor string theory, and this is now under development in Kendall [48]. Statistical asymptotics present a harder challenge to computer algebra than does stochastic calculus, because typically one requires fourth-order rather than second-order expansions and this leads (especially in multivariate cases) to expressions composed of very large numbers of summands. It will therefore be interesting to see what gains might follow from an approach implementing these notions of geometry and invariance.

At the time of writing it appears that success will come from programming the computer algebra package so as to be able to use the language of Taylor string theory to provide succinct specifications to the computer algebra package of the calculations to be performed. The package will use the string-theoretic expressions to produce much larger expressions using ordinary dummy suffices and the ordinary repeated summation convention, and will then substitute in the actual formula obtained in particular examples under investigation. The principal difficulty lies in ensuring that calculations and combinatorial constructions are carried out without repetition, for otherwise in such large expressions there will be considerable inefficiency.

A further complication arises from the large numbers of dummy suffices which arise when elaborating on succinct string-theoretic expressions. It appears that this leads to simplification issues which are linked to difficult questions in algorithmic graph theory; thus it may be necessary to accept some inefficiencies in calculation here.

## 4. Theory of Taylor strings and phyla

The objects arising in stochastic calculus which were discussed in section 2 and those arising in asymptotic statistical inference in section 3 are similar in their behaviour under coordinate changes. We have seen how examination of this behaviour leads to the theory of Taylor strings, which (as the adjective suggests) grew out of a requirement for a theory of invariant Taylor series. The importance of invariance is that, although the choice of a specific coordinate system is almost essential for the purposes of computation, a coordinate-free approach is enormously helpful both to aid our understanding and to facilitate the grouping of terms to ensure that computation is efficient. The search for an invariant approach was the motivation for the work which led to the theory of Taylor strings. In this section we gather together the strands of string theory from the previous sections and outline the general theory which encompasses these strands. We then discuss a natural generalization of strings, the theory of *phyla*, and we describe relationships of string and phyla theory to various other mathematical contexts, including especially the theory of natural bundles, by means of stating (with discussion but with no proofs) a number of key theorems.

### 4.1. Derivative strings and the phylon group

Recall from the previous two sections the following four examples, whose motivations from stochastic calculus and statistical asymptotics make string theory seem inevitable.

*Example 4.1*

A stochastic differential $\mathbf{d}\boldsymbol{X}$, as introduced in formulae (2.6)-(2.7) and (2.22), has coordinate representative $\left(\mathrm{d}X^i\,,\ \frac{1}{2}\mathrm{d}[\,X^i\,,\ X^j\,]\right)$ which transform according to formulae (2.15)-(2.17) and (2.23) by the rule

$$\begin{aligned} \mathrm{d}Y^a &= \psi^a_{/i}\mathrm{d}X^i + \psi^a_{/ij}\frac{1}{2}\,\mathrm{d}[\,X^i\,,\,X^j\,]\,, \\ \frac{1}{2}\,\mathrm{d}[\,Y^a\,,\,Y^b\,] &= \psi^{ab}_{/ij}\frac{1}{2}\,\mathrm{d}[\,X^i\,,\,X^j\,]\,. \end{aligned}$$

In order to emphasize the relationship of stochastic differentials with the other three examples we introduce the notation $\mathrm{d}x^i = \mathrm{d}X^i$, $\mathrm{d}x^a = \mathrm{d}Y^a$, $\mathrm{d}x^{ij} = (1/2)\mathrm{d}[\,X^i\,,\,X^j\,]$, $\mathrm{d}x^{ab} = (1/2)\mathrm{d}[\,Y^a\,,\,Y^b\,]$ and write the stochastic differential $\mathbf{d}X$ as the multi-array of differentials $\{[\mathrm{d}x^i\,]\,,\,[\mathrm{d}x^{ij}\,]\}$. Then the above can be written as

$$\begin{aligned} \mathrm{d}x^a &= \psi^a_{/i}\mathrm{d}x^i + \psi^a_{/ij}\mathrm{d}x^{ij}\,, \\ \mathrm{d}x^{ab} &= \psi^{ab}_{/ij}\mathrm{d}x^{ij}\,. \end{aligned}$$

Note that the drift $\mathrm{Drift}(\mathbf{d}X)$ produces a similar quantity when put together with $\mathrm{Bracket}(\mathbf{d}X, \mathbf{d}X)$: if $[\mathrm{d}y^i\,]$ is the vector of differentials formed by the drift then $\{[\mathrm{d}y^i\,]\,,\,[\mathrm{d}x^{ij}\,]\}$ transforms in a similar manner.

*Example 4.2*

As described in section 3.2, a scalar string field $\mathbf{f}$ of length $T$ is represented in local coordinates by arrays $f_K$ with $1 \leq |K| \leq T$ which according to formula (3.18) transform under coordinate change by

$$f_C = f_K \omega^K_{/C}\,, \tag{4.1}$$

where summation is carried out over the multi-index $K$ according to the extended summation convention discussed in section 3.2. Note that the special case of $T = 2$ corresponds to the transformation law concerning second-order differentials described by equation (2.24), which we repeat here:

$$\begin{aligned} f_{/a} &= f_{/i}\omega^i_{/a} \\ f_{/ab} &= f_{/ij}\omega^{ij}_{/ab} + f_{/i}\omega^i_{/ab}\,. \end{aligned}$$

*Example 4.3*

A connection string field of length $T$ as defined in section 3.2 is represented by arrays $\Gamma^i_K$ with $1 \leq |K| \leq T$ which according to formula (3.19) transform by

$$\Gamma^a_C = \psi^a_{/i}\Gamma^i_K \omega^K_{/C}\,.$$

*Example 4.4*

Given a yoke, consider the quantities $\overset{\alpha}{\Gamma}_{i;K} = \overset{\alpha}{\Gamma}_{i;k_1\dots k_n}$ with $1 \leq |K| \leq T$ which were define in (3.9). These transform by

$$\overset{\alpha}{\Gamma}_{a;C} = \omega^{i}_{/a}\overset{\alpha}{\Gamma}_{i;k}\omega^{K}_{/C}\,.$$

Consider the two remarks:

(a) example 4.1 is dual to example 4.2 (with $T = 2$),

(b) a connection string field transforms as the tensor product of a vector field with a scalar string field.

These suggest the formulation of a general theory encompassing all these objects, namely the theory of *derivative strings* (Barndorff-Nielsen [7], Barndorff-Nielsen & Blæsild [10]-[11]). Derivative strings are defined in coordinate terms as follows. Given a point $m$ in a manifold $\mathbf{M}$, a *derivative string of tensorial degree* $(r, s)$ *and length* $(T, U)$ *at* $m$ assigns to each local coordinates system $\omega$ round $m$ a *multi-array* (sequence of real-valued arrays) $H^{IL}_{JK}$ indexed by multi-indices $I = (i_1 \dots i_r)$, $J = (j_1 \dots j_s)$, and multi-indices $K$, $L$ with $|K| \leq T$, $|L| \leq U$. These arrays are required to transform under coordinate change from $\omega$ to $\psi$ by

$$\begin{aligned} H^{AD}_{BC} &= H^{a_1\dots a_r\,,D}_{b_1\dots b_s\,,C} \\ &= \psi^{a_1\dots a_r}_{/i_1\dots i_r}\,\omega^{j_1\dots j_s}_{b_1\dots b_s}\,H^{i_1\dots i_r\,,L}_{j_1\dots j_s\,,K}\,\omega^{K}_{/C}\,\psi^{D}_{/L} \\ &= \psi^{A}_{/I}\,\omega^{J}_{/B}\,H^{IL}_{JK}\,\omega^{K}_{/C}\,\psi^{D}_{/L}\,, \end{aligned} \tag{4.2}$$

where the lengths of the multi-indices $I$, $J$ are held fixed at $|I| = r$, $|J| = s$, but otherwise the extended summation convention applies.

Note the following points about formula (4.2). The derivatives are evaluated at $\psi(m)$ or $\omega(m)$, as appropriate, and the extended Einstein summation convention (of summing over any multi-index which occurs as both a subscript and a superscript) is used. Note the differences in the behaviour of the various multi-indices; because $|A| = |I|$ and $|B| = |J|$, the multi-indices $I$ and $J$ behave in a tensorial manner. Hence elements $i_1, \dots, i_r$ and $j_1, \dots, j_s$ of $I$ and $J$ are called *tensorial indices*. In contrast, the multi-indices $K$ and $L$ range in length over $1 \leq |K| \leq |C|$ and $|D| \leq |L| \leq U$ and so elements $k_1, \dots, k_t$ and $l_1, \dots, l_u$ of $K$ and $L$ are called *structural indices*. We have emphasized this distinction by writing out the sequences of tensorial indices in full in the inner two formulae of equations (4.2).

The upper indices $I$ and $L$ behave in a "contravariant" fashion, while the lower indices $J$ and $K$ behave "covariantly". The set of strings of tensorial degree $(r, s)$ and length $(T, U)$ at $m$ is denoted by $\mathcal{S}^{rU}_{sT}(\mathbb{M})_m$. Taking the union as $m$ runs through $\mathbb{M}$ yields the spaces $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ of strings of degree $(r, s)$ and length $(T, U)$ on $\mathbb{M}$. Elements of $\mathcal{S}^{r0}_{sT}(\mathbb{M})$ and $\mathcal{S}^{sU}_{r0}(\mathbb{M})$ are called $(r, s)$-*costrings* and $(r, s)$-*contrastrings*, respectively. Note finally that the distinction between tensorial and structural indices is also brought out by a representation of $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ as the tensor product of vector bundles $\mathcal{T}^r_s(\mathbb{M}) \otimes \mathcal{S}^{0U}_{0T}(\mathbb{M})$, where $\mathcal{T}^r_s(\mathbb{M})$ is the bundle of tensors of type $(r, s)$ over $\mathbb{M}$; see the comment before equation (4.5).

The examples 4.1-4.4 correspond to elements of $\mathcal{S}^{02}_{00}(\mathbb{M})$, $\mathcal{S}^{00}_{0T}(\mathbb{M})$, $\mathcal{S}^{10}_{0T}(\mathbb{M})$, $\mathcal{S}^{00}_{1T}(\mathbb{M})$ respectively, except that the sense in which the stochastic differential and drift differential of example 4.1 are both members of $\mathcal{S}^{02}_{00}(\mathbb{M})$ is formal only and has to be made precise via the theory of stochastic integration as described in section 2.

The transformation law (4.2) has two notable features:

(ST1) linearity in the $H^{IL}_{JK}$,

(ST2) dependence on the coordinate change from $\omega$ to $\psi$ only through the derivatives $\omega^k_{/C}$ and $\psi^d_{/L}$ with $|C| \leq T$ and $|L| \leq U$.

This allows us to detect the action of a group of Taylor series of diffeomorphisms which plays a fundamental role in Taylor string theory, and which we now describe.

Assume without loss of generality that $\omega(m) = \psi(m) = \mathrm{o}$. Then the multi-array $\{[\psi^c_{/k_1}], \ldots, [\psi^c_{/k_1 \ldots k_T}]\}$ evaluated at $\mathrm{o}$ is essentially the set of coefficients in the $T^{\text{th}}$-order Taylor series of the coordinate change function $\psi \circ \omega^{-1}$ from some open set in $\mathbb{R}^d$ to $\mathbb{R}^d$. Assume for the moment that $U \leq T$ and $T$ is finite. Then the transformations in (4.2) form a representation of $\mathcal{P}_T(d)$, the *phylon group of order* $T$ *of* $\mathbb{R}^d$, where $\mathcal{P}_T(d)$ is the set of multi-arrays $\{[a^c_{k_1}], \ldots, [a^c_{k_1 \ldots k_T}]\}$ with $a^c_{k_1 \ldots k_t}$ symmetric in $k_1 \ldots k_t$ and with the $[a^c_k]$ forming a non-singular matrix. Under identification of multi-arrays $\{[a^c_{k_1}], \ldots, [a^c_{k_1 \ldots k_T}]\}$ in $\mathcal{P}_T(d)$ with $T^{\text{th}}$-order $\mathbb{R}^d$-valued Taylor series

$$f^c(\omega) = a^c_k \omega^k + \frac{1}{2!} a^c_{k_1 k_2} \omega^{k_1} \omega^{k_2} + \cdots + \frac{1}{T!} a^c_{k_1 \ldots k_T} \omega^{k_1} \ldots \omega^{k_T},$$

the group operation in $\mathcal{P}_T(d)$ corresponds to composition of functions. From the coordinate-free viewpoint, $\mathcal{P}_T(d)$ is the group of $T$-jets at $o$ of local diffeomorphisms of $(\mathbb{R}^d, o)$ with itself, the group operation being composition. (Recall that two functions are said to have the same $T$-jet at a point $x$ if they have the same $T^{\text{th}}$-order Taylor series round $x$.) Thus $\mathcal{P}_T(d)$ is the group of polynomial functions from $\mathbb{R}^d$ to $\mathbb{R}^d$, of degree at most $T$, with zero constant term and invertible linear term. It can be written as

$$\mathcal{P}_T(d) = \left\{(A_1, \ldots, A_T) \mid A_1 \in \odot^i (\mathbb{R}^d)^* \otimes \mathbb{R}^d ,\ A_1 \in \mathrm{GL}(d)\right\},$$

where $\odot$ denotes the symmetric tensor product.

To avoid considering separately the various values of $T$, it is useful to consider Taylor series of infinite order and so we introduce the *infinite phylon group* $\mathcal{P}(d)$ or $\mathcal{P}_\infty(d)$ of $\infty$-jets of invertible $\mathbb{R}^d$-valued formal power series on $\mathbb{R}^d$. Thus

$$\mathcal{P}(d) = \left\{(A_1, A_2, A_3, \ldots) \mid A_i \in \odot^i (\mathbb{R}^d)^* \otimes \mathbb{R}^d ,\ A_1 \in \mathrm{GL}(d)\right\}.$$

In order to describe the way in which $\mathcal{P}(d)$ acts in (4.2) we require some notation corresponding to the convention for summing over ordered partitions described in section 2.2 in equation (2.19) and (PS1)-(PS3). If $E$ and $F$ are vector spaces then we set

$$\mathcal{L}_{T,0}(E;F) = \prod_{t=1}^{T} \left(\bigotimes{}^t E\right)^* \otimes F$$

for $1 \leq T \leq \infty$. Note that $\mathcal{P}_T(d)$ is contained in $\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R}^d)$. Given $B = (B_1, \ldots, B_T)$ in $\mathcal{L}_{T,0}(E;F)$ we define $B_{t/u}$ in $\left(\bigotimes^t E\right)^* \otimes \left(\bigotimes^u F\right)$ by

$$B_{t/u}(e_1 \otimes \cdots \otimes e_t) = \sum B_{|C_1|}(e_{C_1}) \otimes \cdots \otimes B_{|C_u|}(e_{C_u}), \tag{4.3}$$

where $e_{C_i} = (e_{j_1} \otimes \cdots \otimes e_{j_k})$ when $C_i = (j_1, \ldots, j_k)$ and the summation runs over standard ordered partitions $(C_1, \ldots, C_u)$ of $(1, \ldots, t)$ into $u$ subsets, as described in (PS1)-(PS3).

We now define the binary operation

$$\begin{aligned} \square : \mathcal{L}_{T,0}(E;F) \times \mathcal{L}_{T,0}(F;H) &\to \mathcal{L}_{T,0}(E;H) \\ (B, A) &\mapsto A \,\square\, B \end{aligned}$$

where

$$\left(A \,\square\, B\right)_t = \sum_{u=1}^{t} A_u \circ B_{t/u}\,.$$

This definition is motivated by the composition mapping formula (Faà di Bruno's formula), since the restriction of $\square$ from $\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R}^d)\times\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R}^d)$ to $\mathcal{P}_T(d)\times\mathcal{P}_T(d)$ is the group operation on $\mathcal{P}_T(d)$ corresponding to composition of $T$-jets. There is a left action of $\mathcal{P}_T(d)$ on $\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R})$ by

$$\begin{aligned}\mathcal{P}_T(d)\times\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R}) &\to \mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R})\\ (B,A) &\mapsto A\,\square\,B^{-1}\end{aligned}\tag{4.4}$$

and this is the action involved in (4.2) when $r=s=U=0$.

An alternative description of action (4.4) was given by Carey & Murray [21], relating it to jets $\mathcal{J}_0^\infty(\mathbb{R}^d;\mathbb{R}^d)$ of vector fields on $\mathbb{R}^d$ at the origin $o$ of $\mathbb{R}^d$. This idea is that $\mathcal{L}_{T,0}(\mathbb{R}^d;\mathbb{R})$ can be identified with the set of linear operators

$$D:\bigoplus_{k=1}^{T}\bigotimes{}^k\mathcal{J}_0^\infty(\mathbb{R}^d;\mathbb{R}^d)\to\mathbb{R}$$

satisfying

$$\begin{aligned}D\big(X_1\otimes\cdots\otimes X_{i-1}\otimes X_i\otimes fX_{i+1}\otimes X_{i+2}\otimes\cdots\otimes X_k\big) &=\\ =D\big(X_1\otimes\cdots\otimes X_{i-1}\otimes fX_i\otimes X_{i+1}\otimes X_{i+2}\otimes\cdots\otimes X_k\big)&+\\ +D\big(X_1\otimes\cdots\otimes X_{i-1}\otimes X_i(f)X_{i+1}\otimes X_{i+2}\otimes\cdots\otimes X_k\big)&\end{aligned}$$

and

$$D\big(fX_1\otimes\ldots\otimes X_k\big)=f(0)D\big(X_1\otimes\cdots\otimes X_k\big)$$

using vectorfield $\infty$-jets $X_1,\ldots,X_k\in\mathcal{J}_0^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and a function $f:\mathbb{R}\to\mathbb{R}$. Action (4.4) is identified with the action $B:D\mapsto B\dagger D$ (for $B$ in $\mathcal{P}_T(d)$), where

$$(B\dagger D)\big(X_1\otimes\ldots\otimes X_k\big)=D\big(j_0^\infty(Tg^{-1}\circ Y_1\circ g)\otimes\cdots\otimes j_0^\infty(Tg^{-1}\circ Y_k\circ g)\big)$$

using a $T$-jet $B=j_0^T g$ for $g$ a local diffeomorphism of $(\mathbb{R}^d,0)$ with itself and vectorfield $\infty$-jets $X_i=j_0^\infty Y_i$ for $Y_1,\ldots,Y_k$ vector fields on $\mathbb{R}^d$. (Here $j_0^T g$ denotes the $T$-jet at 0 of $g$, *et cetera*, while $Tg^{-1}$ denotes the tangent map of $g^{-1}$.)

Taylor string theory is also related to the theory of vector bundles associated with $T^{\text{th}}$-order frame bundles. Recall (Ehresmann [28], Kobayashi [50]) that the *$T^{\text{th}}$-order frame bundle* $H^T(M)$ on a manifold $\mathbb{M}$ is a principal $\mathcal{P}_T(d)$-bundle having as its fibre over $m$ the set of $T$-jets of local diffeomorphisms of $(\mathbb{R}^d,o)$ with $(\mathbb{M},m)$. Then the bundle $\mathcal{S}_{0T}^{00}(\mathbb{M})$ of $(0,0)$-costrings of length $T$ is the vector bundle associated to $H^T(\mathbb{M})$ by action (4.4). More generally, the bundle $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ is a vector bundle associated to $\mathcal{H}^{\max(T,U)}(\mathbb{M})$ by the tensor product of action (4.4) with its dual and with tensor powers of the usual actions of $\mathrm{GL}(d)$ on $\mathbb{R}^d$ and $(\mathbb{R}^d)^*$.

Another description (Jupp [38]) of $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ is

$$\mathcal{S}^{rU}_{sT}(\mathbb{M}) = \left(\bigotimes{}^r T\mathbb{M}\right) \otimes \left(\bigotimes{}^s T^*\mathbb{M}\right) \otimes \bar{J}^{T,0}(\mathbb{M}) \otimes \bar{J}^{U,0}(\mathbb{M})^*, \qquad (4.5)$$

where $\bar{J}^{T,0}(\mathbb{M})$ denotes the space of zero-truncated semi-holonomic $T$-jets from $\mathbb{M}$ to $\mathbb{R}$, obtained from the space $\bar{J}^T(\mathbb{M})$ of semi-holonomic $T$-jets by quotienting out jets of constant functions. Semi-holonomic jets (Ehresmann [29]) are generalizations of jets whose coordinate expressions lack the symmetry of higher derivatives which is found in jets; a modern treatment can be found in Saunders [62]. In many ways, semi-holonomic jets can be treated just like ordinary (holonomic) jets. The latter are precisely those semi-holonomic jets with coordinate representations which are symmetric in their lower indices.

We have seen that Taylor strings are related to jet theory, to $T^{\text{th}}$-order frame bundles, and in particular to semi-holonomic jets. At this point some readers may feel that there is no point in the theory of strings, since it can be viewed merely as an aspect of the theory of (semi-holonomic) jets. However the point of string theory is that it is an *operational* implementation of jet theory considered at and around a point of the manifold, developed in order to bridge the gap between invariant geometric intuition and coordinate based calculation, as exemplified in section 3. These more abstract geometric contexts are important in providing the setting for Taylor string theory, but do not detract from its actual and potential importance as an aid to computation.

An important class of strings consists of the *structurally symmetric strings* : those strings represented by arrays $H^{IL}_{JK}$ which are symmetric in the elements of $K$ and also symmetric in the elements of $L$. In the description of strings given by Carey & Murray [21], an element of $\mathcal{S}^{00}_{0T}(\mathbb{M})$ is structurally symmetric if it corresponds to a linear operator $D$ which satisfies

$$\begin{aligned} D\big(X_1 \otimes \cdots \otimes X_i \otimes X_{i+1} \otimes \cdots \otimes X_k\big) &+ \\ - D\big(X_1 \otimes \cdots \otimes X_{i+1} \otimes X_i \otimes \cdots \otimes X_k\big) &= \\ = D\big(X_1 \otimes \cdots \otimes [\, X_i\,,\, X_{i+1}\,] \otimes \cdots \otimes X_k\big)\,, \end{aligned}$$

where $[\,\cdot\,,\,\cdot\,]$ denotes the Lie bracket.

Restricting (4.5) to the set $\mathrm{Sym}(\mathcal{S}^{rU}_{sT}(\mathbb{M}))$ of structurally symmetric elements of $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ gives (Barndorff-Nielsen & Blæsild [12]):

$$\mathrm{Sym}(\mathcal{S}^{rU}_{sT}(\mathbb{M})) = \left(\bigotimes{}^r T\mathbb{M}\right) \otimes \left(\bigotimes{}^s T^*\mathbb{M}\right) \otimes J^{T,0}(\mathbb{M}) \otimes J^{U,0}(\mathbb{M})^*,$$

where $J^{T,0}(\mathbb{M})$ denotes the space of zero-truncated $T$-jets of real-valued functions on $\mathbb{M}$, obtained from the space of $T$-jets by quotienting out jets of constant functions. Note that $\mathrm{Sym}(\mathcal{S}^{0U}_{00}(\mathbb{M})) = J^{U,0}(\mathbb{M})^*$ is the $U^{\text{th}}$-order tangent bundle of Ambrose, Palais & Singer [3].

## 4.2. Invertible connection string fields and intertwining

Perhaps the most useful concept in string theory is that of an invertible connection string field, as defined in the next paragraph. The importance of such fields is that they assign to each point $m$ of $\mathbb{M}$ a semi-holonomic jet from $(\mathbb{M}, m)$ to $(T_m\mathbb{M}, o)$ which can be considered as (the jet of) a "semi-holonomic coordinate chart" round $m$ taking values in the tangent space $T_m\mathbb{M}$. (Constrast this with the usual coordinate charts on a manifold which take values in a *fixed* vector space not depending on $m$.)

In order to define invertible connection string fields, note that an important special case of (4.5) gives

$$\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M}) = T\mathbb{M} \otimes \overline{J}^{T,0}(\mathbb{M}).$$

Thus connection strings of length $T$ at $m$ can be identified with semi-holonomic $T$-jets from $(\mathbb{M}, m)$ to $(T_m\mathbb{M}, o)$. A *connection string field of length $T$* is a section of $\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})$. The canonical projection from $\overline{J}^{T,0}(\mathbb{M})$ to $\overline{J}^{1,0}(\mathbb{M}) = J^{1,0}(\mathbb{M}) = T^*\mathbb{M}$ yields a projection of $\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M}) = T\mathbb{M} \otimes \overline{J}^{T,0}(\mathbb{M})$ to $T\mathbb{M} \otimes T^*\mathbb{M} = \mathrm{End}(T\mathbb{M})$, which in coordinate terms sends the multi-array $H$ with component arrays $[H^i_K]$ to the matrix $[H^i_k]$ which is the first array in the sequence of arrays making up the multi-array. A connection string field is called *invertible* if it projects to an invertible element of $\mathrm{Sec}(\mathrm{End}(T\mathbb{M}))$; which is to say in coordinate terms, if it projects to an invertible matrix $[H^i_k]$. The "semi-holonomic coordinate charts" given by an invertible connection string field $\mathbf{\Gamma}$ are most easily understood in the important special case where $\mathbf{\Gamma}$ is structurally symmetric of length $T$. Then $\mathbf{\Gamma}(m)$ is a holonomic jet at each point $m$ of $\mathbb{M}$, and so there is a local diffeomorphism $\phi_m : (\mathbb{M}, m) \to (T_m\mathbb{M}, o)$ such that $\mathbf{\Gamma}(m) = j^{T,0}_m \phi_m$. Thus $\phi_m$ can be considered as a system of local coordinates representing the connection string field $\mathbf{\Gamma}$ at $m$, precisely as described in the discussion of invariant Taylor series expansions in section 3.3. Thus $\phi_m$ is a system of *extended normal coordinates* for $\mathbf{\Gamma}$ at $m$ in the weak local sense that $\phi_m(m) = o$ and the expressions $\Gamma^i_K$ for $\mathbf{\Gamma}$ in these coordinates satisfy $\Gamma^i_K(o) = 0$ for $|K| \geq 2$. Use of such coordinates simplifies many calculations. See Blæsild [17] and Mora [56], and also Murray [57]. For a general invertible connection string field $\mathbf{\Gamma}$ the role of $\phi_m$ is played by the semi-holonomic jet $\mathbf{\Gamma}(m)$ from $(\mathbb{M}, m)$ to $(T_m\mathbb{M}, o)$. Note that $\mathbf{\Gamma}$ has a unique inverse $\mathbf{\Gamma}(m)^{-1}$ in $\overline{J}^T_0(T\mathbb{M}, \mathbb{M})$, defined by

$$\mathbf{\Gamma}(m) \,\Box\, \mathbf{\Gamma}(m)^{-1} = j^T_0 I_{T_m\mathbb{M}} ,$$

where $j^T_0 I_{T_m\mathbb{M}}$ denotes the $T$-jet at $o$ of the identity map of $T_m\mathbb{M}$ and (by slight abuse of notation) $\Box$ denotes composition of semi-holonomic jets.

One of the main uses of invertible connection string fields is that of "intertwining", as in the discussion of invariant Taylor series expansions in section 3.3, with the objective of relating strings to collections of tensor fields (Barndorff-Nielsen & Blæsild [10]-[11] and Blæsild [18]), as described in both coordinate-free and coordinate-based terms in Theorem 4.1 below. We now give a general formulation of the notion of intertwining of strings, following (Barndorff-Nielsen & Blæsild [10]-[11] and Murray [57]) — we repeat from section 3.2 that there is no obvious connection with intertwining of group representations.

The formulation is that once we have fixed on an invertible connection string field of length $\max(T, U)$ then we have a bijection between $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ and $\bigoplus_{n=1}^T \bigoplus_{p=1}^U \mathcal{T}^{r+p}_{s+n}(\mathbb{M})$, where $\mathcal{T}^{r+p}_{s+n}(\mathbb{M})$ denotes the vector bundle $(\bigotimes^{r+p} T\mathbb{M}) \otimes (\bigotimes^{s+n} T^*\mathbb{M})$ of tensors of type $(r+p,\ s+n)$, in which case the resulting tensors arise from repeated differentiation in a representing coordinate system as described in the discussion after equation (3.26). In the language of jets, the intuitive idea behind intertwining can be seen by considering the case of a structurally symmetric invertible connection string field $\mathbf{\Gamma}$ of length $T$. At each point in $m$ of $\mathbb{M}$, $\mathbf{\Gamma}(m) = j^{T,0}_m \phi_m$ for some local diffeomorphism $\phi_m : (\mathbb{M}, m) \to (T_m\mathbb{M}, o)$. For any real-valued function $f$ on $\mathbb{M}$, $j^{T,0}_m f$ in $\mathcal{S}^{0\,0}_{0\,T}(\mathbb{M})_m$ is mapped to $j^{T,0}_o(f \circ \phi_m^{-1})$ in $\mathcal{S}^{0\,0}_{0\,T}(T_m\mathbb{M})_o = \bigoplus_{n=1}^T \bigotimes^n T_m\mathbb{M}^*$, where this last identification uses the vector space structure of $T_m\mathbb{M}$. The effect is to perform the repeated differentiation in the coordinate system given by $\phi_m$. As $m$ runs through $\mathbb{M}$, this yields a function from $\mathcal{S}^{0\,0}_{0\,T}(\mathbb{M})$ to $\bigoplus_{n=1}^T \mathcal{T}^0_n(\mathbb{M})$. If $\mathbf{\Gamma}$ is not symmetric then jets must be replaced by semi-holonomic jets.

Theorem 4.1 describes in general mathematical terms how each invertible connection string field $\mathbf{\Gamma}$ gives a bijection between strings and collections of tensors. At each point $m$ of $\mathbb{M}$, composition with the inverse of the "semi-holonomic coordinate system" $\mathbf{\Gamma}(m)$ transforms $(0,0)$-costrings $B$ at $m$ in $\mathbb{M}$ into $(0,0)$-costrings $B \,\Box\, \mathbf{\Gamma}(m)^{-1}$ at $0$ in $T_m\mathbb{M}$ and the latter are covariant tensors. The dual construction transforms $(0,0)$-contrastings into contravariant tensors. The general construction, given explicitly in (4.6), is the tensor product of the above transformations with the identity transformation of tensors on $\mathbb{M}$.

THEOREM 4.1 (coordinate-free form). — *Let $\mathbf{\Gamma}$ be an invertible connection string field of length* $\max(T, U)$. *Then there is a vector bundle isomorphism between $\mathcal{S}^{rU}_{sT}(\mathbb{M})$ and $\bigoplus_{n=1}^{T} \bigoplus_{p=1}^{U} \mathcal{T}^{r+p}_{s+n}(\mathbb{M})$, determined by*

$$A \otimes B \otimes C \mapsto A \otimes \big(B \,\Box\, \mathbf{\Gamma}(m)^{-1}\big) \otimes \big(\mathbf{\Gamma}(m)_* C\big) \tag{4.6}$$

*for $A$ in $\big(\bigotimes^r T_m\mathbb{M}\big) \otimes \big(\bigotimes^s T^*_m\mathbb{M}\big)$, $B$ in $\mathcal{S}^{0\,0}_{0\,T}(\mathbb{M})_m = \overline{\mathcal{J}}^{T,0}_m(\mathbb{M})$ and $C$ in $\mathcal{S}^{0\,U}_{0\,0}(\mathbb{M})_m = \overline{\mathcal{J}}^{U,0}_m(\mathbb{M})^*$. Here, $\mathbf{\Gamma}(m)_* C$ in $\mathcal{S}^{0\,U}_{0\,0}\big(T_m\mathbb{M}\big)_o = \overline{\mathcal{J}}^{U,0}_o\big(T_m\mathbb{M}\big)^*$ is defined by*

$$\mathbf{\Gamma}(m)_* C(D) = C\big(D \,\Box\, \mathbf{\Gamma}(m)^{-1}\big)$$

*for $D$ in $\overline{\mathcal{J}}^{U,0}_o\big(T_m\mathbb{M}\big)$.*

Thus once we have fixed on *one invertible* connection string field $\mathbf{\Gamma}$ we have isomorphisms of all string fields of lesser or equal lengths with various appropriate collections of tensor fields. This generalizes the well-known result for classical connections, that given one connection we may obtain all other connections by adding tensor fields of a particular type.

To express this result in coordinate terms, we use the coordinate forms of the costrings and contrastrings generated by a connection string. Let $\mathbf{\Gamma}$ be a connection string of length $T$ on $\mathbb{M}$. The $(r,0)$-costring $\{[\mathbf{\Gamma}^I_K] : |K| = 1, 2, \ldots\}$ (with $I = (i_1 \ldots i_r)$) generated by $\mathbf{\Gamma}$ has been defined in (3.21). Now assume that $\mathbf{\Gamma}$ is invertible and let the matrix with elements $\big(\mathbf{\Gamma}_1^{-1}\big)^k_i$ be the inverse of the matrix $[\mathbf{\Gamma}^i_k]$. Then the arrays $G^l_J$ are defined for $|J| \geq 1$ by

$$G^l_J = \big(\mathbf{\Gamma}_1^{-1}\big)^l_i \sum_{\pi=1}^{|J|-1} (-1)^\pi \Delta^i_{J_1} \Delta^{J_1}_{J_2} \ldots \Delta^{J_{\pi-2}}_{J_{\pi-1}} \Delta^{J_{\pi-1}}_{J},$$

where the summation is over all sets $J_1, \ldots, J_{\pi-1}$ with

$$1 < |J_1| < \ldots < |J_{\pi-1}| < |J|$$

and

$$\Delta^i_K = \Delta^i_{k_1 \ldots k_s} = \Gamma^i_{l_1 \ldots l_s} (\Gamma_1^{-1})^{l_1 \ldots l_s}_{k_1 \ldots k_s} = \Gamma^i_L (\Gamma_1^{-1})^L_K ,$$

the summation being over multi-indices $L$ with $|K| = |L|$. The $(0, s)$-contrastring generated by $\Gamma$ has coordinate representation $\{G^L_J\}$ with $|J| = s$ and

$$G^L_J = \sum_{J/u} G^{l_1}_{J_1} \ldots G^{l_u}_{J_u} ,$$

where $L = (l_1 \ldots l_u)$ and the summation runs over standard ordered partitions of $J$ into $u$ subsets $(J_1, \ldots, J_u)$.

THEOREM 4.1 (coordinate-based form).— *Let $\Gamma$ be an invertible connection string field of length* $\max(T, U)$ *and let* $\{\Gamma^I_K : |K| \leq \max(T, U)\}$ *with* $|I| = r$ *and* $\{G^L_J : |L| \leq \max(T, U)\}$ *with* $|J| = s$ *be the* $(r, 0)$*-costring and the* $(0, s)$*-contrastring generated by* $\Gamma$. *Further, let $H$ and $N$ denote sets of arrays* $H^{I\,L}_{J\,K}$ *and* $N^{I\,Q}_{J\,P}$. *Then*

$$H^{I\,L}_{J\,K} = N^{I\,Q}_{J\,P} \Gamma^P_K G^L_Q \tag{4.7}$$

*for* $|K| \leq \max(T, U)$ *and* $|L| \leq \max(T, U)$ *if and only if*

$$N^{I\,Q}_{J\,P} = H^{I\,L}_{J\,K} G^K_P \Gamma^Q_L . \tag{4.8}$$

*Moreover, provided that $H$ and $N$ are related by (4.7) or (4.8), we have*

$$N \in \bigoplus_{n=1}^{T} \bigoplus_{p=1}^{U} \mathcal{T}^{r+p}_{s+n}(\mathbf{M}) \quad \textit{if and only if} \quad H \in \mathcal{S}^{r\,U}_{s\,\infty}(\mathbf{M}) .$$

In the case when $r = s = 0$, Theorem 4.1 gives the construction of tensorial derivatives of a scalar field with respect to an invertible connection string field, as defined in (3.20) in the case $\Gamma^i_k = \delta^i_k$.

### 4.3. Special connection string fields

An important class of invertible connection string fields consists of the *special connection string fields.* These are connection string fields which map to the identity endomorphism of $T\mathbb{M}$ under the projection of $\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})$ to $\mathcal{S}^{1\,0}_{0\,1}(\mathbb{M}) = T\mathbb{M} \otimes T^*\mathbb{M}$; that is to say, they are connection string fields with $\Gamma^i_k = \delta^i_k$. Special connection string fields of length 2 can be identified with affine connections by mapping the string given in coordinate form by $\{[\delta^i_k], [\Gamma^i_{jk}]\}$ to the connection with Christoffel symbols $\Gamma^i_{jk}$. This is the origin of the name "connection string". Note that every invertible connection string field is equivalent to a pair consisting of a special connection string field and a non-singular $(1,1)$-tensor field. In order to state this more precisely, let $\mathrm{Inv}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big)$ and $\mathrm{Spec}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big)$ denote respectively the spaces of invertible and of special connection strings of length $T$ on $\mathbb{M}$. Then there is a bijection

$$\begin{aligned} \mathrm{Inv}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big) &\to \mathrm{Spec}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big) \times_{\mathbb{M}} \mathrm{Aut}(T\mathbb{M}) \\ \boldsymbol{\Gamma} &\mapsto \Big(j^{T,0}_o\big(\pi(\boldsymbol{\Gamma})^{-1}\big) \,\square\, \boldsymbol{\Gamma}\,,\, \pi(\boldsymbol{\Gamma})\Big)\,, \end{aligned} \tag{4.9}$$

where $\pi : \mathrm{Inv}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big) \to \mathrm{Inv}\big(\mathcal{S}^{1\,0}_{0\,1}(\mathbb{M})\big) = \mathrm{Aut}(T\mathbb{M})$ is the projection and, for each $m$ in $\mathbb{M}$, $\pi(\boldsymbol{\Gamma})(m)$ is regarded as a linear function from $T_m\mathbb{M}$ to itself.

Special connection string fields are equivalent to various geometric objects which arise in other contexts. Theorem 4.2 details the equivalence with certain reductions of semi-holonomic frame bundles. The relationship to higher-order connections is given in Theorem 4.3. Theorem 4.4 describes the connection with vector field differentiation strings.

The semi-holonomic $T^{\text{th}}$-order frame bundle $\overline{H}^T(\mathbb{M})$ over $\mathbb{M}$ is the semi-holonomic analogue of the $T^{\text{th}}$-order frame bundle $H^T(\mathbb{M})$. Its fibre over $m$ is the set of invertible semi-holonomic $T$-jets from $(\mathbb{R}^d, o)$ to $(\mathbb{M}, m)$. It is a principal $\overline{\mathcal{P}}_T(d)$-bundle over $\mathbb{M}$, where $\overline{\mathcal{P}}_T(d)$ is the group of invertible semi-holonomic $T$-jets from $(\mathbb{R}^d, o)$ to itself. Note that $\mathrm{GL}(d)$ acts on the right on $\overline{H}^T(\mathbb{M})$ by

$$\begin{aligned} \overline{H}^T(\mathbb{M}) \times \mathrm{GL}(d) &\to \overline{H}^T(\mathbb{M}) \\ (\Delta, X) &\mapsto \Delta \,\square\, j^T_0(X)\,, \end{aligned}$$

which is to say, via the homomorphism

$$\begin{aligned} \iota : \mathrm{GL}(d) &\to \mathcal{P}_T(d) \subset \overline{\mathcal{P}}_T(d) \\ X &\mapsto j^{T,0}_o(X)\,, \end{aligned} \tag{4.10}$$

where $X$ is regarded as a linear function from $\mathbb{R}^d$ to itself. There is a projection $\pi : \overline{H}^T(\mathbb{M}) \to \overline{H}^1(\mathbb{M}) = H^1(\mathbb{M})$ which takes each semi-holonomic $T$-jet to its 1-jet.

THEOREM 4.2.— *There is a fibre bundle isomorphism*

$$\overline{H}^T(\mathbb{M})/\mathrm{GL}(d) \to \mathrm{Spec}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big)$$

*given by*

$$\Delta \mapsto j_o^{T,0}\big(\pi(\Delta)\big) \,\square\, \Delta^{-1}$$

*for $\Delta$ in $\overline{H}^T(\mathbb{M})_m$, where $\pi(\Delta)$ is regarded as a linear mapping from $(\mathbb{R}^d, o)$ to $(T_m\mathbb{M}, o)$. This isomorphism restricts to a fibre bundle isomorphism from $H^T(\mathbb{M})/\mathrm{GL}(d)$ to the space* Sym $\Big(\mathrm{Spec}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big)\Big)$ *of structurally symmetric elements of* $\mathrm{Spec}\big(\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})\big)$. *The corresponding bijections between sections of the bundles identify reductions of $\overline{H}^T(\mathbb{M})$ (respectively $H^T(\mathbb{M})$) to* GL$(d)$ *with special connection string fields (respectively structurally symmetric special connection string fields) of length $T$ on $\mathbb{M}$.*

The reductions of $H^T(\mathbb{M})$ to GL$(d)$ are the *coordinate string fields* on $\mathbb{M}$ considered by Murray [57].

A $T^{\text{th}}$*-order connection* on $\mathbb{M}$ is a connection on the principal bundle $\pi : \overline{H}^T(\mathbb{M}) \to \mathbb{M}$, that is to say a smooth assignment $p \mapsto H_p$ of a subspace $H_p$ of the tangent space $T_p\overline{H}^T(\mathbb{M})$ to each point $p$ of $\overline{H}^T(\mathbb{M})$, which is

(TC1) *horizontal* in that $\pi_* H_p = T_{\pi(p)}\mathbb{M}$,

(TC2) *equivariant* in that $H_{gp} = R_{g*} H_p$ for $g \in \overline{\mathcal{P}}_T(d)$.

Here $R_g$ denotes the right action of $\overline{\mathcal{P}}_T(d)$ on $\overline{H}^T(\mathbb{M})$ given by composition of semi-holonomic jets. Equivalently (Yuen [68]), a $T^{\text{th}}$-order connection on $\mathbb{M}$ is a section $\sigma$ of the projection $\pi : \overline{H}^{T+1}(\mathbb{M}) \to H^1(\mathbb{M})$ such that

$$\sigma\big(j^1(\phi \circ X)\big) = \sigma(j^1\phi) \,\square\, \iota(X)$$

for $\phi$ a local diffeomorphism from $(\mathbb{R}^d, o)$ to $(\mathbb{M}, m)$, $X$ in GL$(d)$, and where $\iota : \mathrm{GL}(d) \to \overline{\mathcal{P}}_{T+1}(d)$ is as defined in (4.10). Since such sections correspond bijectively to sections of $\overline{H}^{T+1}(\mathbb{M})/\mathrm{GL}(d) \to \mathbb{M}$, the following result is immediate from Theorem 4.2.

THEOREM 4.3.— *There is a bijection from $(T-1)^{th}$-order connections on* $\mathbb{M}$ *to special connection string fields of length* $T$ *on* $\mathbb{M}$ *given by* $\sigma \mapsto \mathbf{\Gamma}$, *where*

$$\mathbf{\Gamma}(m) = j_m^{T,0}(j_0^1\phi) \,\square\, \sigma(j_0^1\phi)^{-1}$$

*for* $\phi$ *a local diffeomorphism from* $(\mathbb{R}^d, o)$ *to* $(\mathbb{M}, m)$. *Here* $j_0^1\phi$ *is regarded first as a linear map from* $\mathbb{R}^d = T_o\mathbb{R}^d$ *to* $T_m\mathbb{M}$ *and then as an element of* $H^1(\mathbb{M})$.

Differentiation strings (Barndorff-Nielsen, Blæsild & Mora [14]-[15]) can be considered as ways of using sequences of vector fields to construct linear differential operators which act on vector fields. A *vector field differentiation string of length* $T$ on $\mathbb{M}$ is a sequence $\nabla^0, \ldots, \nabla^T$ of $\mathbb{R}$-multilinear mappings

$$\begin{aligned} \nabla^n &: \mathrm{Sec}(T\mathbb{M})^n \to L(\mathrm{Sec}(T\mathbb{M})\,;\,\mathrm{Sec}(T\mathbb{M})) \\ &(X_1, \ldots, X_n) \mapsto \nabla^n_{(X_1,\ldots,X_n)} \end{aligned}$$

satisfying

$$\nabla^0 Y = Y$$

$$\nabla^n_{(X_1,\ldots,fX_i,\ldots,X_n)} = \sum L_{(X_{i_1},\ldots,X_{i_r})}(f)\nabla^{n-r}_{(X_{j_1},\ldots,X_{j_{n-r}})}\,, \qquad \text{for } n = 1, \ldots, T\,,$$

where $X_1, \ldots, X_n$, $Y$ are vector fields on $\mathbb{M}$, $f$ is a real-valued function on $\mathbb{M}$, the summation runs over all subsets $\{i_1, \ldots, i_r\}$ of $\{i+1, \ldots, n\}$, $i_1 < \cdots < i_r$, the subset $\{j_1, \ldots, j_{n-r}\}$ is the complement in $\{1, \ldots, n\}$ of $\{i_1, \ldots, i_r\}$, $j_1 < \cdots < j_{n-r}$, and $L_{(X_{i_1},\ldots,X_{i_r})} = L_{X_{i_r}} \circ \cdots \circ L_{X_{i_1}}$, with $L_X f = Xf = \mathrm{d}f(X)$ being the usual Lie derivative action of vector fields on real-valued functions.

THEOREM 4.4.— *There is a bijection from special connection string fields of length* $T$ *on* $\mathbb{M}$ *to vector field differentiation strings of length* $T-1$ *on* $\mathbb{M}$, *given by*

$$\mathbf{\Gamma} \mapsto (\nabla^0, \ldots, \nabla^{T-1})$$

*where*

$$\nabla^n_{(X_1,\ldots,X_n)}Y = \mathbf{\Gamma}(L_{(Y,X_1,\ldots,X_n)})\,, \quad n = 0, \ldots, T-1\,,$$

*the differential operator* $L_{(Y,X_1,\dots,X_n)}$ *being regarded as a section of* $\mathcal{S}^{0\,n+1}_{0\,0}(\mathbb{M})$ *and the string field* $\mathbf{\Gamma}$ *in* $\mathrm{Sec}\,\mathcal{S}^{1\,0}_{0\,T}(\mathbb{M})$ *being considered via its projection to* $\mathrm{Sec}\,\mathcal{S}^{1\,0}_{0\,n+1}(\mathbb{M})$ *as an element of* $L\big(\mathrm{Sec}\,\mathcal{S}^{0\,n+1}_{0\,0}(\mathbb{M})\,,\,\mathrm{Sec}(T\mathbb{M})\big)$.

## 4.4. Yokes

We have seen in section 3.2 that one way in which special connection string fields may arise is from yokes. For $g : \mathbb{M} \times \mathbb{M} \longrightarrow \mathbb{R}$ and $m$, $x$ in $\mathbb{M}$, define $g_x : \mathbb{M} \rightarrow \mathbb{R}$, $\gamma_m : \mathbb{M} \rightarrow T^*_m\mathbb{M}$ and $\overset{1}{\mathbf{\Gamma}}{}^{\flat}$ in $\mathrm{Sec}\,\big(\mathrm{Sym}\big(\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M})\big)\big)$ by $g_x(y) = g(x,y)$, $\gamma_m(x) = \mathrm{d}g_x(m)$ and $\overset{1}{\mathbf{\Gamma}}{}^{\flat}(m) = j^{\infty,0}_m\gamma_m$, effectively the connection string construction given in (3.9) with $\alpha = 1$. We have seen in section 3.2 that $g$ is a yoke if, for all $m$ in $\mathbb{M}$, $\gamma_m(m) = 0$ and $j^{1,0}_m(\gamma_m)$ is a non-degenerate element of $J^{1,0}_m(\mathbb{M};T^*_m\mathbb{M}) = T^*_m\mathbb{M} \otimes T^*_m\mathbb{M}$. If $g$ is a yoke, let $\rho$ be the non-degenerate $(0,2)$-tensor field given by $\rho(m) = j^{1,0}_m(\gamma_m)$. As noted after (3.5), $\rho$ is symmetric. Denote by $\mathcal{Y}(\mathbb{M})$ and $\mathcal{R}(\mathbb{M})$ the spaces of yokes and of pseudo-Riemannian metrics on $\mathbb{M}$. Then there is a surjective mapping

$$\begin{aligned}\mathcal{Y}(\mathbb{M}) &\rightarrow \mathrm{Sec}\,\Big(\mathrm{Sym}\,\Big(\mathrm{Inv}\big(\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M})\big)\Big)\Big) \times \mathcal{R}(\mathbb{M}) \\ g &\mapsto \big(\overset{1}{\mathbf{\Gamma}}{}^{\flat}\,,\,\rho\big)\,,\end{aligned}$$

where $\mathrm{Sym}\,\Big(\mathrm{Inv}\big(\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M})\big)\Big)$ denotes the set of structural symmetric elements of $\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M})$ which project to non-degenerate elements of $\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M}) = T^*\mathbb{M} \otimes T^*\mathbb{M}$.

The important feature of a yoke is that the string field $\overset{1}{\mathbf{\Gamma}}{}^{\flat}$ assigns to each point $m$ of $\mathbb{M}$ an $\infty$-jet of a local diffeomorphism from $(\mathbb{M},m)$ to $(T^*_m\mathbb{M}\,,\,o)$, that is to say a jet of a coordinate chart round $m$ taking values in the cotangent space $T^*_m\mathbb{M}$. Using $\rho$ to identify $\mathrm{Sym}\big(\mathcal{S}^{0\,0}_{1\,\infty}(\mathbb{M})\big)$ with $\mathrm{Sym}\big(\mathcal{S}^{1\,0}_{0\,\infty}(\mathbb{M})\big)$ sends $\overset{1}{\mathbf{\Gamma}}{}^{\flat}$ to a structurally symmetric special connection string field $\overset{1}{\mathbf{\Gamma}}$, as in (3.10) with $\alpha = 1$. Then $\overset{1}{\mathbf{\Gamma}} \times \overset{1}{\mathbf{\Gamma}}$ is a structurally symmetric special connection string field on $\mathbb{M} \times \mathbb{M}$ and by Theorem 4.1 it gives a vector bundle isomorphism from $\mathcal{S}^{0\,0}_{0\,\infty}(\mathbb{M} \times \mathbb{M})$ to $\prod_{n=1}^{\infty} \mathcal{T}^0_n(\mathbb{M} \times \mathbb{M})$. Pulling this back to $\mathbb{M}$ by the inclusion of $\mathbb{M}$ as the diagonal of $\mathbb{M} \times \mathbb{M}$ and restricting to structurally symmetric strings gives a vector bundle isomorphism from $\mathrm{Sym}\big(\mathcal{S}^{0\,0}_{0\,\infty}(\mathbb{M} \times \mathbb{M})\big)_{\Delta}$ to $\prod_{r,\,s\geq 1}\big(\odot^r T^*\mathbb{M}\big) \otimes \big(\odot^s T^*\mathbb{M}\big)$. Here $\mathrm{Sym}\big(\mathcal{S}^{0\,0}_{0\,\infty}(\mathbb{M} \times \mathbb{M})\big)_{\Delta}$ denotes the bundle of structurally symmetric $(0,0)$-derivative strings of lenght $\infty$ at the diagonal of $\mathbb{M} \times \mathbb{M}$ and $\odot$ denotes the symmetric tensor product. In particular, the zero-truncated $\infty$-jet of the yoke $g$ at the diagonal of $\mathbb{M} \times \mathbb{M}$ yields a collection $\{T_{(r,s)} \mid r,\, s \geq 1\}$ of $(0\,,\, r+s)$-tensor fields on $\mathbb{M}$. The importance for statistical asymptotics of this construction is that the tensors in (3.33), (3.34) and (3.35) are obtained in this way from the expected or observed likelihood yoke of (3.16) and (3.17).

Also note that we may use the intertwining operation described by Theorem 4.1, together with the "canonical" connection string induced by the normal geodesic coordinates provided by the Christoffel symbols at length 2 of $\overset{1}{\Gamma}{}^{\flat}$, further to represent a yoke as the combination of a sequence of tensors, a connection, and a pseudo-Riemannian metric.

### 4.5. Convolutive multiplication

Derivative costrings can be multiplied together to give new derivative strings by the operation of *convolutive multiplication*. Consider two strings $H_1 \in \mathcal{S}^{r_1}_{s_1}{}^{0}_{T_1}(\mathbb{M})$ and $H_2 \in \mathcal{S}^{r_2}_{s_2}{}^{0}_{T_2}(\mathbb{M})$ at the same point of $\mathbb{M}$. The product $H_1 * H_2$ is the element of $\mathcal{S}^{r_1+r_2}_{s_1+s_2}{}^{0}_{T_1+T_2}(\mathbb{M})$ given in coordinates by

$$(H_1 * H_2)^{IP}_{JQK} = \sum_{K/2} (H_1)^{I}_{JK_1} (H_2)^{P}_{QK_2}\,,$$

where juxtaposition of multi-indices means concatenation: $QK$ denotes the multi-index formed by placing $Q$ before $K$.

From the coordinate-free viewpoint, convolutive multiplication is obtained from the usual tensor convolution in each fibre $\bigoplus_{s=1}^{\infty}(\bigotimes^{s} T^{*}_{m}\mathbb{M})$. Note that this product is noncommutative and that the costrings generated by $\Gamma$ according to formula (3.21) may be obtained by repeated convolutive multiplication of $\Gamma$ with itself. Note also that $B_{t/u}$ as defined in (4.3) is the component in $\left(\bigotimes^{t} E\right)^{*} \otimes \left(\bigotimes^{u} F\right)$ of the $u$-fold convolutive power of $B$.

### 4.6. Differential strings

Differential strings are a generalization of derivative strings. Whereas the building blocks of derivative strings are semi-holonomic jets of real-valued functions, differential strings are based on semi-holonomic jets of sections of tensor powers of the tangent bundle. The coordinate-based definition

(Blæsild & Mora [19]) of differential strings requires the generalization of the arrays $\omega^K_{/C}$ to arrays $[\psi\,,\,\omega]^{EK}_{IC}$, where

$$[\psi\,,\,\omega]^{EK}_{IC} = \sum_{C//2} \omega^K_{/C_1}\,\omega^L_{/C_2}\,(\psi^E_{/I})_{/L}$$

$$= \omega^K_{/C}\,\psi^E_{/I} + \sum_{C/2}\{\omega^K_{/C_1}\,\omega^L_{/C_2} + \omega^K_{/C_2}\,\omega^L_{/C_1}\}(\psi^E_{/I})_{/L}\,,$$

where $C//2$ indicates that the sum is over all ordered partitions $(C_1, C_2)$ of $C$ into 2 subsets, either of which may be empty, such that the order within each $C_i$ is the same as that within $C$. (Here $\omega^K_{/C} = 0$ if exactly one of $K$ and $C$ is empty and $\omega^K_{/C} = 1$ if both $K$ and $C$ are empty.) A *differential string of degree* $(r, s)$, *type* $(p, q)$ *and length* $(T, U)$ *at* $m$ assigns to each local coordinate system $\omega$ round $m$ a multi-array of real-valued arrays $H^{I\,L\,M}_{J\,K\,N}$ indexed by multi-indices $I$, $J$, $K$, $L$, $M$, $N$ with $|I| = r$, $|J| = s$, $|K| \leq T$, $|L| \leq U$, $|M| = p$, $|N| = q$, and transforming under coordinate change from $\omega$ to $\psi$ by the law

$$H^{A\,D\,E}_{B\,C\,F} = \psi^A_{/I}\omega^J_{/B}\,H^{I\,L\,M}_{J\,K\,N}\,[\omega\,,\,\psi]^{N\,D}_{F\,L}\,[\psi\,,\,\omega]^{E\,K}_{M\,C}\,, \tag{4.11}$$

holding the lengths of $A$, $B$, $E$ and $F$ fixed at $|A| = r$, $|B| = s$, $|E| = p$ and $|F| = q$. Because $|A| = |I| = r$ and $|B| = |J| = s$, the multi-indices $I$ and $J$ behave tensorially. On the other hand, the *structural multi-indices* $K$ and $L$ and the *type multi-indices* $M$ and $N$ behave in a more complicated way. Letting $m$ run through the manifold $\mathbb{M}$ yields the space $\mathcal{D}^{rUp}_{sTq}(\mathbb{M})$ of differential strings of degree $(r, s)$, type $(p, q)$ and length $(T, U)$ on $\mathbb{M}$. A coordinate-free description (Jupp [31]) of $\mathcal{D}^{rUp}_{sTq}(\mathbb{M})$ generalizing (4.5) is

$$\mathcal{D}^{rUp}_{sTq}(\mathbb{M}) = \left(\bigotimes{}^r T\mathbb{M}\right) \otimes \left(\bigotimes{}^s T^*\mathbb{M}\right) \otimes \overline{J}^T\left(\bigotimes{}^p T\mathbb{M}\right) \otimes \overline{J}^U\left(\bigotimes{}^q T\mathbb{M}\right)^*,$$

where $\overline{J}^T\left(\bigotimes^p T\mathbb{M}\right)$ denotes the bundle of semi-holonomic $T$-jets of sections of $\bigotimes^p T\mathbb{M}$.

## 4.7. Phyla

The transformation laws (4.2) and (4.11) of derivative strings and differential strings involve higher derivatives $\omega^k_{/C}$ and $\psi^c_{/K}$ of coordinate changes and so are coordinate descriptions of representations of $\mathcal{P}(d)$. Because these transformation laws are polynomial in $\omega^k_{/C}$ and $\det\left(\omega^k_{/c}\right)^{-1}$, the representations are algebraic. By allowing general algebraic representations we obtain objects called *phyla*. Thus, in terms of local coordinates, phyla are represented by arrays $H^{A_1\ldots A_r}_{B_1\ldots B_s}$ which transform under coordinate change from $\omega$ to $\psi$ by

$$H^{A_1\ldots A_r}_{B_1\ldots B_s} = H^{I_1\ldots I_r}_{J_1\ldots J_s} D[\omega\,,\,\psi]^{A_1\ldots A_r\,J_1\ldots J_s}_{B_1\ldots B_s\,I_1\ldots I_r}\,, \tag{4.12}$$

where $D[\omega\,,\,\psi]$ is a block matrix in which the elements of the blocks are polynomials in $\omega^i_{/A}$ and $\psi^a_{/I}$. It follows from Theorem 4.5 below that the arrays $H^{A_1\ldots A_r}_{B_1\ldots B_s}$ can be ordered so that $D[\omega\,,\,\psi]$ is an upper-triangular block matrix. The function which takes a pair $(\omega,\psi)$ of local coordinate systems to the matrix $D[\omega\,,\,\psi]$ is called a *D-matrix*, it satisfies:

(D1) each $D[\omega\,,\,\psi]$ is a nonsingular upper-triangular block matrix in which elements of the blocks are polynomials in $\omega^i_{/A}$ and $\psi^a_{/I}$,

(D2) (the *cocycle condition*)

$$D[\omega\,,\,\psi]D[\psi\,,\,\chi] = D[\omega\,,\,\chi]\,.$$

Phyla arise also in the context of natural bundles. A *natural bundle over n-manifolds* (Palais & Terng [58]) assigns to each $n$-dimensional smooth manifold $\mathbb{M}$ a smooth fibre bundle over $\mathbb{M}$ with total space $F(\mathbb{M})$ such that if $\phi : \mathbb{M} \to \mathbb{N}$ is an embedding then there is a bundle map $F(\phi) : F(\mathbb{M}) \to F(\mathbb{N})$ over the map $\phi$ of base spaces, such that $F$ is continuous in an appropriate sense. Terng [66] showed that natural vector bundles with $d$-dimensional fibre are given by representations of $\mathcal{P}(d)$. It follows that finite-dimensional phyla are precisely the elements of algebraic natural vector bundles (natural vector bundles for which this representation is algebraic). Consequences of more general results of Epstein & Thurston [31] are that for finite-dimensional natural bundles the continuity of $F$ is automatic and that phyla of dimension $d$ have coordinate forms which change by a representation of $\mathcal{P}_T(d)$ with $T \leq 2d+1$.

As well as derivative strings, the class of phyla includes differential strings, the values of the higher-order differential forms of Meyer [55], sectorforms (While [67, Chap. 3]), quasi-jets (Dekrét [24]) from $\mathbb{M}$ to $\mathbb{R}$ and the "new tensors" of Foster [32]-[35]. . Meyer's differential forms of order $T$

are the same as sections of $\tilde{J}^{T-1}(T^*\mathbb{M})$, the bundle of non-holonomic $(T-1)$-jets of cotangent fields on $\mathbb{M}$. Whereas the vector bundle $J^{T-1}(T^*\mathbb{M})$ of (holonomic) $(T-1)$-jets of cotangent fields has fibre dimension

$$d\sum_{i=0}^{T-1}\binom{d+i-1}{i},$$

the bundle $\bar{J}^{T-1}(T^*\mathbb{M})$ of semi-holonomic $(T-1)$-jets has fibre dimension $d(d^T-1)(d-1)^{-1}$ and the bundle $J^{T-1}(T^*\mathbb{M})$ of non-holonomic $(T-1)$-jets has fibre dimension $d(d+1)^{T-1}$. The values of higher-order differential forms are not differential strings but they are quasi-jets. The class of phyla does not include

(i) (for $r \geq 2$) elements of $r^{\text{th}}$ iterated tangent bundles $T^r\mathbb{M}$, nor elements of their subbundles ${}^r\mathbb{M}$ which are invariant under the canonical involutions of $T^r\mathbb{M}$ occurring in the differentiable extensions of Bowman [20], because these transform non-linearly;

(ii) relative tensors of non-integral weight (e.g. Synge & Schild [65, p. 240]), because their transformation laws involve non-polynomial functions of $\det(\omega^i{}_{/a})$.

Extensors (Craig [22]) are based on jets along a curve in $\mathbb{M}$ of tensor fields and so do not fit neatly into the context of phyla, because of the choice of curve.

Before considering representations of $\mathcal{P}_T(d)$, note that $\mathcal{P}_T(d)$ is a semi-direct product of $\mathcal{P}_T^{(1)}(d)$ and $\mathrm{GL}(d)$, where $\mathcal{P}_T^{(1)}(d)$ is the kernel of the group epimorphism

$$\mathcal{P}_T(d) \to \mathcal{P}_1(d) = \mathrm{GL}(d)$$

which maps a polynomial to its linear term and where the homomorphism $\iota$ of (4.10) include $\mathrm{GL}(d)$ in $\mathcal{P}_T(d)$ as the group of $T$-jets of invertible linear functions.

The main property of phyla comes from Theorem 4.5 (Barndorff-Nielsen *et al.* [13]; Terng [66]). This states that every representation of $\mathcal{P}_T(d)$ can be described by a block upper-triangular matrix representation in which the diagonal blocks are representations of $\mathrm{GL}(d)$.

THEOREM 4.5.— *For any algebraic complex representation $\chi : \mathcal{P}_T(d) \to \mathrm{GL}(V)$ of $\mathcal{P}_T(d)$, there is a decomposition $V = V_1 \oplus \cdots \oplus V_k$ of $V$ into $\mathrm{GL}(d)$-irreducibles such that, for $f$ in $\mathcal{P}_T^{(1)}(d)$ and $X$ in $\mathrm{GL}(d)$*

$$\chi(fX) = \begin{pmatrix} \chi_1(X) & \chi_{1\,2}(f)\chi_2(X) & \dots & \chi_{1\,k}(f)\chi_k(X) \\ 0 & \chi_2(X) & \dots & \chi_{2\,k}(f)\chi_k(X) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \chi_k(X) \end{pmatrix} \tag{4.13}$$

*where $\chi_i : \mathrm{GL}(d) \to \mathrm{GL}(V_i)$ is a representation, $\chi_{i\,j} : \mathcal{P}_T^{(1)}(d) \to V_j^* \otimes V_i$ is a map which is $\mathrm{GL}(d)$-equivariant, and*

$$\begin{aligned}\chi_{ij}(fg) = \chi_{ij}(g) + \chi_{i,i+1}(f)\chi_{i+1,j}(g) + \chi_{i,i+2}(f)\chi_{i+2,j}(g) + \cdots \\ \cdots + \chi_{i,j-1}(f)\chi_{j-1,j}(g) + \chi_{ij}(f)\end{aligned}$$

*for $i < j$.*

In the case of the representation of $\mathcal{P}_T(d)$ which underlies example 4.1, a block upper-triangular matrix representation can be deduced from the form of equation (2.23).

An important consequence of the block upper-triangular nature of the matrix in (4.13) is that phyla can be projected onto other phyla which they "contain". More precisely, the arrays $H^{A_1 \dots A_r}_{B_1 \dots B_s}$ in the coordinate expression of a phylon can be ordered into arrays representing $(P_1, \dots, P_k)$ with $P_i \in V_i$, so that, for $1 \le j \le k$, $(P_1, \dots, P_j)$ also represents a phylon. To illustrate this, consider again examples 4.1-4.4.

*Example 4.1.*— Using the notation introduced at the beginning of this section, a stochastic differential $\{[\mathrm{d}x^i], [\mathrm{d}x^{ij}]\}$ projects to its bracket $\{[\mathrm{d}x^{ij}]\}$, as does the drift differential $\{[\mathrm{d}y^i], [\mathrm{d}x^{ij}]\}$. Furthermore, for $1 \le W \le U$, a $(0,0)$-contrastring of length $U$ represented by arrays $H^I$ with $1 \le |I| \le U$ projects to another $(0,0)$-contrastring of length $U$ represented by $K^I$ with

$$K^I = \begin{cases} H^I & \text{for } |I| \ge W, \\ 0 & \text{for } |I| < W. \end{cases}$$

*Example 4.2.*— For $U \le T$, a scalar string of length $T$, represented in local coordinates by arrays $f_K$ with $1 \le |K| \le T$, projects to a scalar string of length $U$, represented by $f_K$ with $1 \le |K| \le U$. A particular case of this is the rather obvious fact that the $T$-jet of a function determines its $U$-jet if $1 \le U \le T$.

*Example 4.3.* — For $U \leq T$, a connection string of length $T$, represented by arrays $\Gamma^i_K$ with $1 \leq |K| \leq T$, projects to a connection string of length $U$, represented by $\Gamma^i_K$ with $1 \leq |K| \leq U$.

*Example 4.4.* — For $U \leq T$, a string in $\mathcal{S}^{0\,0}_{1\,T}(M)$, represented by arrays $\Gamma_{i;K}$ with $1 \leq |K| \leq T$, projects to an element of length $U$, represented by $\Gamma_{i;K}$ with $1 \leq |K| \leq U$.

The class of phyla is closed under taking direct sums, duals and tensor products, as well as under differentiation, which is to say the operations of taking jets, semi-holonomic jets and non-holonomic jets of phylon fields. One might naïvely hope to obtain all phyla by applying these operations to tensor fields. However the class of phyla is much too large for this, containing as it does objects such as those transforming by the following representations of $\mathcal{P}(d)$.

*Example 4.5.* — Let $\kappa_1 \ldots \kappa_{T-1}$ be real numbers. Then the mapping which takes the $\mathcal{P}_T(d)$-element $(A_1, \ldots, A_T)$ to

$$\begin{pmatrix} A_{1/1} & \kappa_1 A_{2/1} & \ldots & \kappa_1 \ldots \kappa_{T-1} A_{T/1} \\ 0 & A_{2/2} & \ldots & \kappa_2 \ldots \kappa_{T-1} A_{T/2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_{T/T} \end{pmatrix}$$

(where the $A_{t/n}$ are defined by (4.3)) is an algebraic representation of $\mathcal{P}_T(d)$ on $\bigoplus_{t=1}^T \bigotimes^t \mathbb{R}^d$ and so defines a class of phyla. Taking $\kappa_1 = \cdots = \kappa_{T-1} = 1$ gives the representation dual to that occurring in (4.4) and the corresponding phyla are the elements of $\mathcal{S}^{0\,T}_{0\,0}(\mathbf{M})$.

Detailed examination of an example related to example 4.5 in the case $T = 2$ and to (4.13) with $k = 2$ (Barndorff-Nielsen *et al.* [13], Terng [66]) shows that there are uncountably many non-isomorphic indecomposable algebraic representations of $\mathcal{P}_T(d)$ and so the class of all phyla is very large indeed. Terng [66] gave a general procedure for classifying phyla in terms of orbits of group actions on Lie algebra cohomology spaces. However, these are very difficult to compute and so a detailed explicit classification of all phyla would appear to be out of reach. This seems to be a clear indication that attention should be focussed on the special cases occurring in stochastic calculus and asymptotic statistical inference, to see what special features present in those cases might generalize to suggest a smaller and hence more amenable subclass of the class of all phyla.

## 5. Conclusion

In this paper we have described the way in which higher-order calculus arises in stochastic calculus and in statistical asymptotics, and have used these instances to motivate a description of the general theory of Taylor strings and phyla, and the relationship of this theory to mathematical considerations of jet bundles, higher order frame bundles, and natural vector bundles. As a consequence of this exposition it can be seen that Taylor strings can be viewed in three parallel ways:

(a) *the coordinate-based approach* using transformation laws, as given for example in equation (4.2) for derivative strings, equation (4.11) for differential strings and equation (4.12) for general phyla;

(b) *the abstract approach* relating strings *etc.* to other mathematical objects such as jet bundles, higher order frame bundles, and natural vector bundles;

(c) *the intuitive approach* treating strings as derived from connection strings which in turn correspond to representing systems of coordinates (as in the discussions of sections 3.3 and 4.2, and related to Murray's coordinate strings).

Each of these approaches has its own advantages: the coordinate-based approach is appropriate for calculations and should prove essential for the computer algebra implementation now under development; the abstract approach is helpful for example in the identification of the tensor $t_{ij;kl}$ of equation (3.35); while the intuitive approach makes it clear what is the fundamental content of Taylor string theory, namely the treatment of Taylor series expansions expressed in terms of a coordinate system whose choice may depend on the point about which the expansion takes place.

Of the two applications we describe in section 2 and 3, that to stochastic calculus is essentially after the fact. Its importance lies in the strong link between the second-order stochastic calculus and Taylor string theory, which allows us to use the formal aspects of second-order stochastic calculus as a gentle introduction to higher-order calculus. The application to asymptotic statistics provided the inspiration for developing Taylor string theory as a context for the fourth-order invariance considerations which then arose. Future developments are likely to centre around topics of computation and approximation in the statistical application: further investigations into invariant approaches to approximations and development of a computer algebra implementation paralleling the computer algebra of symbolic Itô calculus as described in section 2.4. Note however that a recent method of proof of the fundamental Itô formula (as in equation (2.3)), described in Kendall [49], can be viewed in a way reminiscent of Taylor string theory: the length-2 string $\overset{2}{\mathrm{f}}$ derived from a $C^2$ function $f$ is bounded above and below and below by sequences of quadratic approximations, and hence the general Itô formula is deduced from the special case for quadratic $f$. It may be that similar string-theoretic insights into fundamental results in statistical asymptotics await our discovery.

Finally it should be noted that in principle Taylor string theory is in no way confined to statistical applications, being appropriate wherever use is made of high order Taylor series expansions in contexts where invariance considerations should apply.

## Acknowledgment

This research was supported by a twinning grant in the EC Science Programme.

## References

[1] Amari (S.-I.) .— *Differential Geometric Methods in Statistics*, Lecture Notes in Statistics **28**, Springer-Verlag, New York – Heidelberg – Berlin (1985).

[2] Amari (S.-I.), Barndorff-Nielsen (O.E.), Lauritzen (S.L.) and Rao (C.R) . *Differential Geometry in Statistical Inference*, Lecture Notes-Monograph Series **10**, Institute of Mathematical Statistics, Hayward, Califormia – (1987).

[3] Ambrose (W.), Palais (R.S.) and Singer (I.M.) .— *Sprays*, Anais Acad. Bras. de Ciências **32** (1960), pp. 163-178.

[4] Andrews (D.F.) and Stafford (J.E.) .— *Tools for the symbolic computation of asymptotic expansions*, J. Roy. Statist. Soc. **B 55** (1993), pp. 613-627.

[5] Antonelli (P.L.), Chapin (J.) and Voorhees (B.H.) .— *The geometry of random genetic drift VI: a random selection diffusion model*, Adv. in Appl. Probab. **12** (1980), pp. 50-58.

[6] Barndorff-Nielsen (O.E.) .— *On a formula for the distribution of the maximum likelihood estimator*, Biometrika **70** (1983), pp. 343-365.

[7] Barndorff-Nielsen (O.E.) .— *Strings, tensorial combinants, and Bartlett adjustments*, Proc. Roy. Soc. London **A 406** (1986), pp. 127-137.

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

[9] BARNDORFF-NIELSEN (O.E.) .— *Invited contribution to the discussion of R. E. Kass: The Geometry of Asymptotic Inference*, Statist. Science **4** (1989), pp. 222-227; Amendments in Statist. Science **5** (1990), p. 370.

[10] BARNDORFF-NIELSEN (O.E.) and BLÆSILD (P.) .— *Strings: Mathematical theory and statistical examples*, Proc. Roy. Soc. London A **411** (1987), pp. 155-176.

[11] BARNDORFF-NIELSEN (O.E.) and BLÆSILD (P.) .— *Derivative strings: contravariant aspect*, Proc. Roy. Soc. London A **411** (1987), pp. 421-444.

[12] BARNDORFF-NIELSEN (O.E.) and BLÆSILD (P.) .— *Coordinate-free definition of structurally symmetric derivative strings*, Adv. in Appl. Math. **9** (1988), pp. 1-6.

[13] BARNDORFF-NIELSEN (O.E.), BLÆSILD (P.), CAREY (A.L.), JUPP (P.E.), MORA (M.) and MURRAY (M.K.) .— *Finite-dimensional algebraic representations of the infinite phylon group*, Acta Appl. Math. **28** (1992), pp. 219-252.

[14] BARNDORFF-NIELSEN (O.E.), BLÆSILD (P.) and MORA (M.) .— *Derivative strings and higher order differentiation*, Dept. of Theoretical Statistics, Institute of Mathematics, Univ. of Aarhus, Memoir n° **11** (1988).

[15] BARNDORFF-NIELSEN (O.E.), BLÆSILD (P.) and MORA (M.) .— *Generalized higher-order differentiation*, Acta Appl. Math. **16** (1989), pp. 243-259.

[16] BARNDORFF-NIELSEN (O.E.), COX (D.R.) and REID (N.) .— *The role of differential geometry in statistical theory*, Internat. Statist. Rev. **54** (1986), pp. 83-96.

[17] BLÆSILD (P.) .— *Yokes: Orthogonal and extended normal coordinates*, Dept. of Theoretical Statistics, Institute of Mathematics, Univ. of Aarhus, Research Report n° **205** (1990).

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

[19] BLÆSILD (P.) and MORA (M.) .— *Differential strings: basic theory*, Exposition. Math. **10** (1991), pp. 249-273.

[20] BOWMAN (R.H.) .— *On differentiable extensions*, Tensor (N.S.) **21** (1970), pp. 139-150.

[21] CAREY (A.L.) and MURRAY (M.K.) .— *Higher-order tensors, strings and new tensors*, Proc. Roy. Soc. London A **430** (1990), pp. 423-432.

[22] CRAIG (H.V.) .— *Vector and Tensor Analysis*, McGraw-Hill, New York (1943).

[23] CRITCHLEY (F.), MARRIOTT (P.K.) and SALMON (M.H.) .— *Preferred point geometry and statistical manifolds*, Ann. Statist. (to appear).

[24] DEKRÉT (A.) .— *On quasi-jets*, Cas. Pest. Mat. **111** (1986), pp. 345-352.

[25] DELLACHERIE (C.) and MEYER (P.-A.) .— *Probabilities and Potential A*, North-Holland, Amsterdam (1978).

[26] DELLACHERIE (C.) and MEYER (P.-A) .— *Probabilities and Potential B*, North-Holland, Amsterdam (1982).

[27] DELLACHERIE (C.) and MEYER (P.-A) .— *Probabilities and Potential C*, North-Holland, Amsterdam (1988).

[28] EHRESMANN (C.) .— *Les prolongements d'une variété différentiable. I. Calcul des jets, prolongement principal*, C.R. Acad. Sci. Paris **239** (1951), pp. 598-600.

[29] EHRESMANN (C.) .— *Extension du calcul des jets aux jets non holonomes*, C.R. Acad. Sci. Paris **239** (1954), pp. 1762-1764.

[30] EMERY (M.) .— *Stochastic Calculus in Manifolds (with an appendix by P.-A. Meyer)*, Springer-Verlag, New York – Heidelberg – Berlin (1989).

[31] EPSTEIN (D.B.A.) and THURSTON (W.P.) .— *Transformation groups and natural bundles*, Proc. London Math. Soc. **38** (1979), pp. 219-236.

[32] FOSTER (B.L.) .— *Differentiation on manifolds without a connection*, Michigan Math. J. **5** (1958), pp. 183-190.

[33] FOSTER (B.L.) .— *Some remarks on tensor differentiation*, Ann. Mat. Pura Appl. **54** (1961), pp. 143-146.

[34] FOSTER (B.L.) .— *Would Leibniz lie to you? (Three aspects of the affine connection)*, Math. Intelligencer **8** (1986), pp. 34-40, 57.

[35] FOSTER (B.L.) .— *Higher derivatives in geometry and physics*, Proc. Roy. Soc. London A **423** (1989), pp. 443-455.

[36] IKEDA (N.) and WATANABE (S.) .— *Stochastic Differential Equations and Diffusion Processes*, North-Holland and Kodansha, Amsterdam and Tokyo (1981).

[37] ITÔ (K.) .— *Stochastic differentials*, Appl. Math. Optim. **1** (1965), pp. 374-381.

[38] JUPP (P.E.) .— *Derivative strings, differential strings, and semi-holonomic jets*, Proc. Roy. Soc. London A **436** (1992), pp. 89-98.

[39] KASS (R.E.) .— *The geometry of asymptotic inference (with discussion)*, Statist. Science **4** (1989), pp. 188-234.

[40] KENDALL (W.S) .— *Computer algebra, Brownian motion, and the statistics of shape*, Geometrization of Statistical Theory, Proceedings, GST Workshop, Univ. of Lancaster (28-31 Oct. 1987); C.T.J. Dodson ed., Univ. of Lancaster, ULDM Publications, pp. 171-192.

[41] KENDALL (W.S) .— *Stochastic differential geometry: an introduction*, Acta Appl. Math. **9** (1987), pp. 29-60.

[42] KENDALL (W.S) .— *Symbolic computation and the diffusion of shapes of triads*, Adv. in Appl. Probab. **20** (1988), pp. 775-797.

[43] KENDALL (W.S) .— *Martingales on manifolds and harmonic maps*, The Geometry of Random Motion, R. Durrett & M. Pinsky eds., American Mathematical Society, Providence, RI (1988), pp. 121-157 (but see Huang and Kendall, 1991).

[44] KENDALL (W.S) .— *Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence*, Proc. London Math. Soc. **61** (1990), pp. 371-406.

[45] KENDALL (W.S) .— *Computer algebra and stochastic calculus*, Notices Amer. Math. Soc. **37** (1990), pp. 1254-1256.

[46] KENDALL (W.S) .— *Diffusion of Euclidean shape*, Disorder in Physical Systems, G. Grimmett & D. Welsh eds., Oxford Univ. Press, Oxford (1990), pp. 203-217.

[47] KENDALL (W.S) .— *Symbolic Itô calculus: an introduction*, Probabilités Numériques, N. Bouleau & D. Talay eds., INRIA, Rocquencourt, France (1991), pp. 186-192.

[48] KENDALL (W.S) .— *Statistical asymptotics using the computer package "REDUCE"*, Warwick Univ. Dept. of Statistics, Research Report in preparation.

[49] KENDAL (W.S.) .— *A remark on the proof of Itô's formula for $C^2$ functions of continuous semimartingales*, J. Appl. Probab. **29** (1992), pp. 216-221.

[50] KOBAYASHI (S.) .— *Canonical forms on vector bundles of higher order contact*, Proc. Symp. Pure Math. **3** (1961), pp. 186-193.

[51] MCCULLAGH (P.) .— *Tensor Methods in Statistics*, Chapman and Hall, London (1987).

[52] MCCULLAGH (P.) and COX (D.R) .— *Invariants and likelihood ratio statistics*, Ann. Statist. **14** (1986), pp. 1419-1430.

[53] MEYER (P.-A) .— *A differential geometric formalism for the Itô integral*, Stochastic Integrals (Proc. Durham 1980), D. Williams ed., Lect. Notes in Math. **851**, Springer-Verlag, New York – Heidelberg – Berlin (1981), pp. 256-270.

[54] MEYER (P.-A) .— *Géométrie stochastique sans larmes*, Sém. Probab. XV, Lect. Notes in Math **850**, Springer-Verlag, New York – Heidelberg – Berlin (1981), pp. 44-102.

[55] MEYER (P.-A) .— *Qu'est-ce qu'une différentielle d'ordre $n$?* Exposition. Math. **7** (1989), pp. 249-264.

[56] MORA (M.) .— *Geometrical expansions for the distributions of the score vector and the maximum likelihood estimator*, Ann. Inst. Statist. Math. **44** (1992), pp. 63-83.

[57] MURRAY (M.K.) .— *Coordinate systems and Taylor series in statistics*, Proc. Roy. Soc. London A **415** (1988), pp. 445-452.

[58] PALAIS (R.S.) and TERNG (C.L.) .— *Natural vector bundles have finite order*, Topology **16** (1977), pp. 271-277.

[59] PROTTER (P.) .— *Stochastic Integration and Differential Equations, a new approach*, Springer-Verlag, New York – Heidelberg – Berlin (1990).

[60] REVUZ (D.) and YOR (M.) .— *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York – Heidelberg – Berlin (1991).

[61] ROGERS (L.C.G.) and WILLIAMS (D.) .— *Diffusions, Markov Processes, and Martingales Vol. II: Itô calculus*, Wiley, Chichester (1987).

[62] SAUNDERS (D.J.) .— *The Geometry of Jet Bundles*, London Mathematical Society Lecture Notes **142**, Cambridge University Press, Cambridge (1989).

[63] SCHWARTZ (L.) .— *Géométrie différentielle du deuxième ordre, semi-martingales et équations différentielles stochastiques sur une variété différentielle*, Sém. Probab. XVI (supplément : géométrie différentielle stochastique), Lect. Notes in Math. **921**, Springer-Verlag, New York – Heidelberg – Berlin, (1982), pp. 1-150.

[64] SCHWARTZ (L.) .— *Semimartingales and their Stochastic Calculus on Manifolds*, Les Presses de l'Université de Montréal, Montréal (1984).

[65] SYNGE (J.-L.) and SCHILD (A.) .— *Tensor Calculus*, Mathematical Expositions **5**, University of Toronto Press, Toronto (1949).

[66] TERNG (C.-L.) .— *Natural vector bundles and natural diffential operators*, Amer. J. Math. **100** (1978), pp. 775-828.

[67] WHITE (J.E.) .— *The Method of Iterated Tangents with Applications in Local Riemannian Geometry*, Monographs and Studies in Mathematics **13**, Pitman, Boston, (1982).

[68] YUEN (P.C.) .— *Higher order frames and linear connections*, Cahiers Topologie Géom. Différentielle **12** (1971), pp. 333-374.