Ann. Inst. Statist. Math.
Vol. 44, No. 1, 63–83 (1992)

GEOMETRICAL EXPANSIONS FOR THE DISTRIBUTIONS OF THE SCORE VECTOR AND THE MAXIMUM LIKELIHOOD ESTIMATOR*

MARIANNE MORA

U.F.R. Sciences Economiques, Université de Paris X Nanterre,
200 Avenue de la République, 92001 Nanterre Cedex, France

(Received April 13, 1989; revised June 2, 1990)

Abstract. In the present note, asymptotic expansions for conditional and unconditional distributions of the score vector are derived. Our aim is to consider these expansions in the light of differential geometry, particularly the theory of derivative strings. Expansions for the distributions of the maximum likelihood estimator are obtained from those for the score vector via transformation, with a view to interpreting from the standpoint of differential geometry the various terms entering the expansions.

Key words and phrases: Geometrical expansions, score vector, maximum likelihood estimator, observed and expected geometries.

1. Introduction

Asymptotic expansions for the distributions of the maximum likelihood estimator and the likelihood ratio statistic, with a view to the interpretability from the standpoint of differential geometry of the various terms entering the expansions, have been discussed inter alia by Amari and Kumon (1983), Barndorff-Nielsen (1986*b*, 1988) and McCullagh and Cox (1986). On applied as well as theoretical grounds, the terms of main interest are those of order $O(1)$, $O(n^{-1/2})$ and $O(n^{-1})$ under ordinary repeated sampling with sample size n. However, not all of these terms have been given a differential geometric interpretation nor have all the relevant types of expansion—conditional and unconditional—been considered. The aim of this paper is to complete the picture by considering the expansions in the light of the recently developed theory of derivative strings (Barndorff-Nielsen (1986*a*), Barndorff-Nielsen and Blæsild (1987*a*, 1987*b*)), and by addressing also the closely related question of asymptotic expansions for the distribution of the score vector.

* The present work was carried out at the Department of Theoretical Statistics, University of Aarhus, Denmark, with support from the Danish-French Cultural Exchange Programme.

In fact, the score vector is a natural starting point for a discussion of problems of the present type because of its intrinsic geometric nature, when considered as a differential or as a covariant vector. It turns out that in the expansions for the score vector to be derived below the "variant" aspect of its distribution is wholly subsumed in the leading normal term, more precisely in the determinant of the variance matrix, while all other terms are invariant and involve certain tensors, the nature of which will be discussed in Section 5.

Expansions for the distribution of the maximum likelihood estimator may be obtained from those for the score vector via transformation, and this procedure provides an automatic separation of the terms into variant terms and invariant terms. We shall also comment on the nature of the variant terms, as seen from the theory of derivative strings.

The conditional distributions to be discussed are relative to an exact or approximate ancillary. We use the terms expected and observed geometries in the sense of Barndorff-Nielsen (1986*b*).

Section 2 contains some background material. Sections 3 and 4 present the various expansions for the score vector and the maximum likelihood estimator, respectively, and in Section 5 we discuss the nature of the various variant and invariant terms. Some concluding remarks are collected in Section 6.

2. Some background material

2.1 *Likelihood quantities*

Let $\mathcal{M} = \{\mathcal{X}, p(\omega; x), \Omega\}$ be a statistical model, where $\mathcal{X}$ is the sample space, Ω the parameter space and $p(\omega; x)$ is the probability density function for the data x, with respect to some dominating measure μ on $\mathcal{X}$ and parametrized by a d-dimensional parameter ω. Let $\hat{\omega}$ be the maximum likelihood estimator of ω. Let $\omega = (\omega^1, \dots, \omega^d)$ and $\hat{\omega} = (\hat{\omega}^1, \dots, \hat{\omega}^d)$ be the coordinates of ω and $\hat{\omega}$, respectively, for which arbitrary components are denoted by the letters r, s, t, ...

For a given observed value x, we denote by l:

$$\omega \in \Omega \to l(\omega; x) = \log p(\omega; x) \tag{2.1}$$

the log-likelihood function of the model. It is assumed to be a smooth function in ω.

Partial derivatives of l with respect to the coordinates ω^r of ω are written as:

$$l_{r_1 \cdots r_p}(\omega; x) = \partial_{\omega^{r_1}} \cdots \partial_{\omega^{r_p}} l(\omega; x), \qquad r_1, \dots, r_p = 1, \dots, d, \quad p \geq 1, \tag{2.2}$$

and briefly as $l_{r_1 \cdots r_p}\omega$, or $l_{r_1 \cdots r_p}$. The score vector for the model is then given by

$$l_*(\omega; x) = (l_r(\omega; x))_{r=1,\dots,d}, \tag{2.3}$$

written briefly as $l_*(\omega)$ or l_*.

The joint cumulants of the log-likelihood derivatives are denoted as

$$\begin{aligned} &K_r = E(l_r)(= 0), \\ &K_{r,s} = \mathrm{Cum}(l_r, l_s), \qquad K_{rs} = \mathrm{Cum}(l_{rs}), \\ &K_{r,s,t} = \mathrm{Cum}(l_r, l_s, l_t), \qquad K_{r,st} = \mathrm{Cum}(l_r, l_{st}) \dots . \end{aligned} \tag{2.4}$$

The following is discussed in detail in Barndorff-Nielsen (1986*b*): From the viewpoint of conditional inference we shall assume that a k-dimensional sufficient statistic t and a $(k-d)$-auxiliary statistic a are given such that the correspondence between $(\hat{\omega}, a)$ and t is a smooth bijection. Note, however, that we do not assume that t necessarily constitutes a dimensional reduction of the original data x. In fact, it is sometimes natural to let $t = x$, for instance in the case of the typical location-scale model. In applications, the statistic a is supposed to be ancillary, i.e. exactly or approximately distribution constant, and then inference on $\hat{\omega}$ may be carried out in the conditional model for $\hat{\omega}$ given a. In this framework, without loss of generality, the log-likelihood function may be rewritten as $l(\omega; \hat{\omega}, a)$, and we may then differentiate l partially with respect to the coordinates $\hat{\omega}^s$ of $\hat{\omega}$ as well as with respect to the coordinates ω^r of ω. We define the so-called *mixed derivatives* of l by setting for any $r_1, \ldots, r_p$; $s_1, \ldots, s_q$ in $\{1, \ldots, d\}$:

$$l_{r_1 \cdots r_p; s_1 \cdots s_q}(\omega; \hat{\omega}, a) = \partial_{\omega^{r_1}} \cdots \partial_{\omega^{r_p}} \partial_{\hat{\omega}^{s_1}} \cdots \partial_{\hat{\omega}^{s_q}} l(\omega; \hat{\omega}, a). \tag{2.5}$$

In particular, the *observed information matrix* j is given by

$$j_{rs}(\omega; \hat{\omega}, a) = -l_{rs}(\omega; \hat{\omega}, a), \qquad r, s \in \{1, \ldots, d\}. \tag{2.6}$$

Furthermore, for any symbol indicating a function of ω and of $(\hat{\omega}, a)$ we write a bar / through the symbol to indicate that we make the substitution $\hat{\omega} \to \omega$. For example, we define $\not{l}_{r_1 \cdots r_p; s_1 \cdots s_q}(\omega; a)$, or $\not{l}_{r_1 \cdots r_p; s_1 \cdots s_q}$ for short, by

$$\not{l}_{r_1 \cdots r_p; s_1 \cdots s_q}(\omega; a) = l_{r_1 \cdots r_p; s_1 \cdots s_q}(\omega; \omega, a), \tag{2.7}$$

and the matrix $\not{j}$ by

$$\not{j}(\omega, a) = j(\omega; \omega, a). \tag{2.8}$$

Notice that, by the definition of $\hat{\omega}$, we have

$$\not{l}_r = 0 \tag{2.9}$$

which yields, by differentiation

$$\not{l}_{rs} + \not{l}_{r;s} = 0 \tag{2.10}$$

and thus,

$$\not{j}_{rs} = \not{l}_{r;s}. \tag{2.11}$$

By differentiating (2.10), one gets

$$\not{l}_{rst} + \not{l}_{rs;t} + \not{l}_{rt;s} + \not{l}_{r;st} = 0. \tag{2.12}$$

2.2 *Geometrical structures*

Under a differential geometric viewpoint, the statistical model $\mathcal{M}$ may be set up as a differentiable manifold and, for this purpose, two parallel constructions have been made. The first approach consists in equipping $\mathcal{M}$ with an "expected" geometrical structure and is particularly useful in connection with Edgeworth expansions for the maximum likelihood estimator under curved exponential models (see Amari and Kumon (1983), Amari (1985)). The second approach, based on the conditional inference standpoint, has been developed by Barndorff-Nielsen (1986*b*, 1988). In this framework the model is rigged with an "observed" geometrical structure and this construction requires mixed derivatives of the log-model function, as defined by (2.5) and (2.7). We now recall briefly these two structures.

The expected geometrical structure

Here, the metric tensor is the expected information matrix i, defined by $i_{rs} = E(l_r l_s)$, $r, s = 1, \dots, d$, and a family of α-connections $\overset{\alpha}{\Gamma}$, the *expected α-connections*, is determined by the Christoffel symbols $\overset{\alpha}{\Gamma}{}^t_{rs}$ given by $\overset{\alpha}{\Gamma}{}^t_{rs} = i^{tu}\overset{\alpha}{\Gamma}_{rsu}$ and

$$\overset{\alpha}{\Gamma}_{rst} = E(l_{rs}l_t) + \frac{1-\alpha}{2}T_{rst}, \qquad \alpha \text{ real}, \tag{2.13}$$

where $[i^{rs}]$ denotes the inverse matrix of i and

$$T_{rst} = E(l_r l_s l_t), \tag{2.14}$$

the so-called *expected skewness tensor*, is a covariant tensor of rank 3. Here and in the following, we adopt the Einstein summation convention.

The observed geometrical structure

First the model is equipped with a metric tensor given by the matrix $\not{j}$. Then a collection of connections, the *observed α-connections* on $\mathcal{M}$, is defined by

$$\overset{\alpha}{\not{\Gamma}}{}^t_{rs} = \not{j}^{tu}\overset{\alpha}{\not{\Gamma}}_{rsu}, \qquad \alpha \text{ real}, \tag{2.15}$$

with

$$\overset{\alpha}{\not{\Gamma}}_{rst} = \not{l}_{rs;t} + \frac{1-\alpha}{2}\not{T}_{rst}, \tag{2.16}$$

where

$$\not{T}_{rst} = -(\not{l}_{rst} + \not{l}_{rs;t}[3]) \tag{2.17}$$

is a covariant tensor of rank 3, analogous to the skewness tensor T_{rst} and referred to as the *observed skewness tensor*. The symbol [ ] indicates a sum of similar terms (here 3) corresponding to appropriate permutation of the indices.

We are mostly interested in the particular cases:

$$\overset{1}{\not{\Gamma}}_{rst} = \not{l}_{rs;t}, \qquad \overset{-1}{\not{\Gamma}}_{rst} = \not{l}_{t;rs}, \tag{2.18}$$

which constitute the first terms of the following geometrical objects which appear in the various expansions we are concerned with.

Let

$$\not{\gamma}^{;r}_{s_1\cdots s_t} = \not{j}^{rr'} \not{\gamma}_{s_1\cdots s_t;r'} \qquad t \geq 1, \tag{2.19}$$

$$\not{\gamma}^{r}_{s_1\cdots s_t} = \not{j}^{rr'} \not{\gamma}_{r';s_1\cdots s_t} \qquad t \geq 1. \tag{2.20}$$

It is shown in Barndorff-Nielsen (1986a) that the sets of arrays $\bar{\not{\Gamma}}^{1} = \{\not{\gamma}^{;r}_{s_1\cdots s_t}, t \geq 1\}$ and $\bar{\not{\Gamma}}^{-1} = \{\not{\gamma}^{r}_{s_1\cdots s_t}, t \geq 1\}$ are special instances of geometrical objects referred to as connection strings. Notice that $\{\not{\gamma}^{;r}_{s_1}, \not{\gamma}^{;r}_{s_1 s_2}\}$ and $\{\not{\gamma}^{r}_{s_1}, \not{\gamma}^{r}_{s_1 s_2}\}$ characterize respectively the (1)- and (-1)-observed connections (see also Barndorff-Nielsen and Blæsild (1987a, 1987b, 1988) for general settings).

Now let ω_0 be an arbitrary point of $\mathcal{M}$ and let (ψ^a) be the coordinate system around ω_0 given by

$$\psi^a(\omega) = \not{j}^{aa'}(\omega_0)\not{\gamma}_{a'}(\omega_0;\omega). \tag{2.21}$$

Then, in this particular coordinate system, the observed connection string $\bar{\not{\Gamma}}^{-1}$ reduces to

$$\Gamma^a_b(p) = \delta^a_b \quad \text{and} \quad \Gamma^a_{b_1\cdots b_t}(p) = 0 \quad t > 1$$

(cf. Murray and Rice (1987), Blæsild (1990)).

Using such a parametrization for the statistical model will lead to some simplifications in the expansion for the maximum likelihood estimator, as will appear from the following.

Finally, given any connection Γ, by repeated covariant differentiation of $\Gamma^r_{s_1 s_2}$ relative to Γ, a special connection string referred to as the "canonical connection string generated by Γ" may be defined (cf. Barndorff-Nielsen and Blæsild (1987a)).

The (1)- and (-1)-connection strings $\bar{\not{\Gamma}}^1$ and $\bar{\not{\Gamma}}^{-1}$ and the canonical connection strings generated by the connections $\not{\Gamma}^1$ and $\not{\Gamma}^{-1}$, respectively, will be considered in Section 5.

2.3 *Exponential models*

We shall be concerned with a core, i.e. full and steep, exponential model $\mathcal{S}$ of order $\boldsymbol{k}$, in the sense of Barndorff-Nielsen (1988), with model function $p(\theta;x) = \exp(\theta\cdot x - K(\theta))$, where $K(\theta)$ denotes the cumulant function of a given dominating measure μ and where the parameter θ and the statistic x are $\boldsymbol{k}$-dimensional vectors, $\theta = (\theta^i)_{i=1,\ldots,\boldsymbol{k}}$ and $x = (x_i)_{i=1,\ldots,\boldsymbol{k}}$. Let $m = K'(\theta) = E_\theta(x)$. The domain of values of m is denoted by $\mathcal{T} = K'(\Theta)$, where Θ is the domain of the parameter θ.

The mean value m provides an alternative parametrization of $\mathcal{S}$. We shall denote by H the inverse of K', i.e. $\theta = H(m) \iff m = K'(\theta)$.

By restricting θ to be a smooth function of a d-dimensional parameter ω ($d < \boldsymbol{k}$), that is $\theta = \zeta(\omega)$, such that the domain Ω of values of ω is open and $\zeta'(\omega)$ is of rank d for any ω in Ω, we obtain a curved subfamily $\mathcal{M}$ of $\mathcal{S}$ whose model function is $p(\omega;x) = \exp(\zeta(\omega)\cdot x - K(\zeta(\omega)))$. We denote by $\eta(\omega)$ the mean value m expressed as a function of ω, i.e. $m = \eta(\omega) = K'(\zeta(\omega))$. Geometrically speaking, the model $\mathcal{S}$ may be viewed as a $\boldsymbol{k}$-dimensional manifold and the model $\mathcal{M}$ as a d-dimensional submanifold of $\mathcal{S}$. We shall assume that $\mathcal{M}$ is a regular submanifold in $\mathcal{S}$.

The model $\mathcal{S}$ is equipped with a Riemannian metric tensor g_{ij} given by

$$g_{ij}(\theta) = \partial_{\theta^i}\partial_{\theta^j}K(\theta) = K_{ij}(\theta),$$

or equivalently, in the m-representation of $\mathcal{S}$, by a metric tensor $g^{ij}(m)$ where g^{ij} denotes the matrix H'.

Then, a family of α-connections $\overset{\alpha}{\Gamma}_{ijk}$ is defined by

$$\overset{\alpha}{\Gamma}_{ijk}(\theta) = \frac{1-\alpha}{2}\partial_{\theta^i}\partial_{\theta^j}\partial_{\theta^k}K(\theta) = \frac{1-\alpha}{2}K_{ijk}(\theta).$$

In the submodel $\mathcal{M}$ the log-likelihood function is denoted by

$$l(\omega; x) = \zeta(\omega)\cdot x - K(\zeta(\omega)), \tag{2.22}$$

and the score vector $l_* = (l_r)_{r=1,\dots,d}$ is

$$l_r(\omega; x) = \zeta^i_{/r}(\omega)\{x_i - K_i(\zeta(\omega))\}, \tag{2.23}$$

where $\zeta^i_{/r}(\omega) = \partial_{\omega^r}\zeta^i(\omega)$ and $K_i(\zeta(\omega)) = \partial_{\theta^i}K(\zeta(\omega)) = \eta_i(\omega)$. Moreover, the second order partial derivatives of l with respect to ω are

$$l_{rs}(\omega; x) = \zeta^i_{/rs}(\omega)\{x_i - K_i(\zeta(\omega))\} - \zeta^i_{/r}(\omega)\zeta^j_{/s}(\omega)K_{ij}(\zeta(\omega)). \tag{2.24}$$

The model $\mathcal{M}$ may be equipped with a metric tensor g_{rs} given by

$$g_{rs}(\omega) = \zeta^i_{/r}(\omega)\zeta^j_{/s}(\omega)g_{ij}(\zeta(\omega)) \tag{2.25}$$

and with a family of α-connections $\overset{\alpha}{\Gamma}_{rst}$ given by

$$\overset{\alpha}{\Gamma}_{rst}(\omega) = \zeta^i_{/rs}(\omega)\zeta^j_{/t}(\omega)g_{ij}(\zeta(\omega)) + \zeta^i_{/r}(\omega)\zeta^j_{/s}(\omega)\zeta^k_{/t}(\omega)\overset{\alpha}{\Gamma}_{ijk}(\zeta(\omega)).$$

This geometrical structure is that naturally induced on $\mathcal{M}$ when $\mathcal{S}$ is equipped with the metric g_{ij} and the family $\overset{\alpha}{\Gamma}_{ijk}$.

2.4 *Ancillary statistics for core exponential families*

From the viewpoint of conditionality, the basis for inference on ω is the conditional distribution of the maximum likelihood estimator $\hat{\omega}$ of ω given an ancillary statistic a.

We shall recall briefly how an ancillary statistic a may be introduced (see Amari (1985) and Barndorff-Nielsen (1983, 1987, 1988) for further specifications) for the model $\mathcal{M}$.

First, notice that for the core exponential model $\mathcal{S}$ and for an observed value x, the maximum likelihood estimator $\hat{\theta}$ of θ is given by $\hat{\theta} = K'(x)$ and the maximum likelihood estimator $\hat{m}$ of m is given by $\hat{m} = x$. For the model $\mathcal{M}$, the maximum likelihood estimator $\hat{\omega}$ of ω must satisfy the likelihood equations

$$(2.26) \qquad l_* = 0 \quad \text{i.e.} \quad \zeta^i_{/r}(\omega)\{x_i - K_i(\zeta(\omega))\} = 0, \qquad r = 1, \ldots, d.$$

For any given ω in Ω, the d vectors $\zeta^*_{/1}(\omega), \ldots, \zeta^*_{/d}(\omega)$, where $\zeta^*_{/r}(\omega) = (\zeta^i_{/r}(\omega))_{i=1,\ldots,k}$, $r = 1, \ldots, d$ span the tangent space $T_\omega\mathcal{M}$ of $\mathcal{M}$ at ω.

We may define a $(k-d)$-dimensional submanifold $A(\omega)$ by setting

$$(2.27) \qquad A(\omega) = \{m \in \mathcal{T} : m - \eta(\omega) \in T_\omega\mathcal{M}^\perp\}.$$

Let us consider the family $A = \{A(\omega), \omega \in \Omega\}$.

Then, a coordinate system $v = (v^l)_{l=1,\ldots,k-d}$ is introduced in each $A(\omega)$ such that the origin $v = 0$ locates the point $\eta(\omega)$ in $\mathcal{M}$ and such that (ω, v) may be regarded as a local coordinate system of S around $\mathcal{M}$. Let the smooth transformation from (ω, v) to m be denoted by $m = \Lambda(\omega, v)$ which reduces to $m = \eta(\omega)$ at any point $(\omega, 0)$.

Let us consider for a while the function $\Lambda : (\omega, v) \to m = \Lambda(\omega, v)$, with $m \in A(\omega)$.

We denote by $B_{*1}, \ldots, B_{*k-d}$ the $k-d$ vectors spanning $(T_\omega M^\perp)$.

Then

$$(2.28) \qquad \Lambda_{i/l}(\omega, v) = \partial_{v^l}\Lambda_i(\omega, v) = B_{il}(\omega).$$

Furthermore,

$$(2.29) \qquad \zeta^i_{/r}(\omega)B_{il}(\omega) = 0, \qquad r = 1, \ldots, d; \quad l = 1, \ldots, k-d.$$

In terms of maximum likelihood estimators, first notice that from (2.26) and (2.27) we have $x = \hat{m} \in A(\hat{\omega})$ and then we may represent the point x in terms of the new coordinate system as $x = \hat{m} = \Lambda(\hat{\omega}, a)$ where the auxiliary statistic a is depending on the family A.

The statistic $(\hat{\omega}, a)$ forms a sufficient statistic and the first log-likelihood derivatives may be written as

$$l_r(\omega; \hat{\omega}, a) = \zeta^i_{/r}(\omega)\{\Lambda_i(\hat{\omega}, a) - \eta_i(\omega)\},$$
$$l_{rs}(\omega; \hat{\omega}, a) = \zeta^i_{/rs}(\omega)\{\Lambda_i(\hat{\omega}, a) - \eta_i(\omega)\} - \zeta^i_{/r}(\omega)\Lambda_{i/s}(\omega).$$

If we take partial derivatives of l_* with respect to $\hat{\omega}$ we get

$$l_{r;s}(\omega;\hat{\omega},a) = \zeta^i_{/r}(\omega)\Lambda_{i/s}(\hat{\omega},a). \tag{2.30}$$

When $\mathcal{S}$ is locally parametrized by (ω, v), the corresponding metric tensor $g_{\alpha\beta}$ evaluated at $(\omega, 0)$ is given by

$$g_{\alpha\beta}(\omega) = \Lambda_{i/\alpha}(\omega,0)\Lambda_{i/\beta}(\omega,0)g^{ij}(\eta(\omega)), \tag{2.31}$$

where the indices α, β relate to the coordinates of (ω, v). Formula (2.31) reduces to

$$g_{rs}(\omega) = \eta_{i/r}(\omega)\eta_{j/s}(\omega)g^{ij}(\omega) \tag{2.32}$$

for the $\mathcal{M}$-part, to

$$g_{rl}(\omega) = 0 \tag{2.33}$$

for the mixed part, by the use of (2.29), and to

$$g_{lk}(\omega) = \Lambda_{i/l}(\omega)\Lambda_{j/k}(\omega)g^{ij}(\omega) \tag{2.34}$$

for the A-part. The matrix $g_{lk}(\omega)$ is the metric of $A(\omega)$ at $v = 0$ and it may be assumed that the coordinate system v in any $A(\omega)$ is such that

$$g_{lk}(\omega) = \text{constant} \tag{2.35}$$

is valid for all ω, at $v = 0$.

3. Expansions for the score vector

In this section we derive asymptotic expansions for the distribution of the score vector in the following frameworks:

1) Approximation to the marginal distribution of the score vector.

2) Approximation to the conditional distribution of the score vector for a given ancillary statistic a, under curved exponential models.

3) Approximation to the conditional distribution of the score vector for a given ancillary statistic a, by the use of the p^*-formula.

In the two first cases the model $\mathcal{M}$ is equipped with the expected geometrical structure while the use of the p^*-formula corresponds naturally to the model $\mathcal{M}$ being equipped with the observed geometrical structure.

The three resulting expansions appear as formulas (3.4), (3.15) and (3.22) below.

3.1 *Marginal expansion for the score vector*

In this subsection we apply the Edgeworth approximation of order two to the marginal distribution $p(l_*;\omega)$ of the score vector (see, for example, Skovgaard (1986) for precise settings concerning multivariate Edgeworth expansions).

We need to introduce the Hermite polynomials $H(l_*;i)$ corresponding to the score vector l_*, the covariance matrix of which is i. We shall use a contravariant version of the first polynomials H, i.e.

$$\begin{aligned}
&H^r(l_*;i) = l^r, \qquad H^{rs}(l_*;i) = l^r l^s - i^{rs},\\
&H^{rst}(l_*;i) = l^r l^s l^t - l^r i^{st}[3], \dots,\\
&H^{rstuvw}(l_*;i) = l^r l^s l^t l^u l^v l^w - l^r l^s l^t l^u i^{vw}[15] + l^r l^s i^{tu} i^{vw}[45]\\
&\qquad\qquad - i^{rs} i^{tu} i^{vw}[15]
\end{aligned} \tag{3.1}$$

where

$$l^r(w) = i^{rr'} l_{r'}(w), \qquad r = 1, \dots, d \tag{3.2}$$

denote the components of the contravariant version l^* of the score vector.

The symbol [ ] indicates a sum of similar terms obtained by appropriate permutations (see, for example, Barndorff-Nielsen and Cox (1989) for a general definition and properties of Hermite polynomials).

The Edgeworth approximation of order two to the (unconditional) distribution of the score vector is given by

$$p^{[2]}[l_*;\omega] = \varphi_d(l_*;i)\{1 + Q_1(l_*) + Q_2(l_*)\} \tag{3.3}$$

where

$$\begin{aligned}
Q_1(l_*) &= \frac{1}{6} K_{r,s,t}(l_*) H^{rst}(l_*;i),\\
Q_2(l_*) &= \frac{1}{24} K_{r,s,t,u}(l_*) H^{rstu}(l_*;i) + \frac{1}{72} K_{r,s,t} K_{u,v,w} H^{rstuvw}(l_*;i).
\end{aligned}$$

Under ordinary repeated sampling of size n we have, in wide generality,

$$p(l_*;\omega) = \varphi_d(l_*;i)\{1 + Q_1(l_*) + Q_2(l_*)\} + O(n^{-3/2}) \tag{3.4}$$

where $Q_1(l_*)$ is of order $n^{-1/2}$ and $Q_2(l_*)$ is of order n^{-1}.

The only part of this expansion which is not invariant under reparametrization is contained in the normal density function.

To see this, note first that the Hermite polynomials $H(x;g)$, where x is any quantity, are affine (or Cartesian) tensors. In the particular case where $x = l_*$ it is easy to show that the Hermite polynomials behave as covariant tensors in the ω-argument. Further, the cumulants of l_* behave as contravariant tensors in ω. Thus the correction terms in the Edgeworth expansion are invariant quantities; in particular, $Q_1(l_*)$ and $Q_2(l_*)$ are both invariant under reparametrization of the model $\mathcal{M}$.

3.2 *Conditional expansion for the score vector under curved exponential families, using expected geometry*

Let ω be the true parameter value and let us consider inference on ω under ordinary repeated sampling with sample size n. In the case when repeated observations are allowed, the geometrical structure of the parametric space is quite similar to the geometrical structure we defined above in Subsections 2.3 and 2.4 and then we shall use the same notations.

By using Taylor expansions, Amari (1985) showed that the statistic $T = (\hat{\omega} - \omega, a)$, where the statistic a is defined as an affine ancillary statistic, is asymptotically normally distributed with covariance matrix $g^{\alpha\beta}$.

Our aim is to follow the same procedure as Amari, that is first to obtain expansions for the distributions of the statistics $\chi = (l_*, a)$, and a and finally to derive from these expansions an expansion for the conditional distribution of l_* given a.

Let us first examine the local transformation

$$(\hat{\omega}, a) \xrightarrow{\Phi} \chi = \Phi(\hat{\omega}, a) = (l_*, a) \tag{3.5}$$

where $l_* = l_*(\omega; \hat{\omega}, a)$.

We denote by χ_α the generic component of χ which is rewritten as

$$\chi_r = l_r \qquad \text{or} \qquad \chi_l = a^l. \tag{3.6}$$

Notice that $(\omega, a) \xrightarrow{\Phi} (0, a)$ since $l_*(\omega; \omega, a) = 0$.

The Jacobian matrix of the transformation between χ and $(\hat{\omega}, a)$ is given by

$$J_{\alpha\beta} = \Phi_{\alpha/\beta} \tag{3.7}$$

with

$$\begin{aligned} &J_{rs}(\hat{\omega}, a) = l_{r;s}(\omega; \hat{\omega}, a), \\ &J_{rk}(\hat{\omega}, a) = \partial_{a^k} l_r(\omega; \hat{\omega}, a) = \zeta^i_{/r}(\omega)\Lambda_{i/k}(\hat{\omega}, a), \\ &J_{kl}(\hat{\omega}, a) = \partial_{a^l} a^k = \delta^k_l. \end{aligned} \tag{3.8}$$

The function Λ has been defined in Subsection 2.4. At the point $(\omega, 0)$ these matrices reduce to

$$\begin{aligned} &I_{rs} = J_{rs}(\omega, 0) = g_{rs}(\omega), \\ &I_{rk} = J_{rk}(\omega, 0) = \zeta^i_{/r}(\omega)\Lambda_{i/k}(\omega, 0) = 0, \\ &I_{kl} = J_{kl}(\omega, 0) = \delta^k_l. \end{aligned} \tag{3.9}$$

Thus $J_{\alpha\beta}$ is invertible and hence the transformation Φ is one-to-one around $(\omega, 0)$.

In order to obtain an Edgeworth expansion for the joint distribution of (l_*, a) we expand $\chi = \Phi(\hat{\omega}, a)$ around $(\omega, 0)$.

Then

$$\chi_\alpha = \Phi_\alpha(\omega, 0) + \Phi_{\alpha/\beta}(\omega, 0)T^\beta + \frac{1}{2}\Phi_{\alpha/\beta\gamma}(\omega, 0)T^\beta T^\gamma + \frac{1}{6}\Phi_{\alpha/\beta\gamma\delta}T^\beta T^\gamma T^\delta + \cdots \tag{3.10}$$

i.e.

$$\chi_\alpha = I_{\alpha\beta}T^\beta + \frac{1}{2}\Phi_{\alpha/\beta\gamma}T^\beta T^\gamma + \frac{1}{6}\Phi_{\alpha/\beta\gamma\delta}T^\beta T^\gamma T^\delta + \cdots.$$

From (3.10) it is easily seen that the asymptotic covariance matrix ι of χ given by

$$\iota_{\alpha\beta} = I_{\alpha\alpha'}I_{\beta\beta'}g^{\alpha'\beta'} \tag{3.11}$$

which may be decomposed into

$$\iota_{rs} = g_{rs}, \qquad \iota_{rk} = 0, \qquad \iota_{kl} = g^{kl}. \tag{3.12}$$

Then, under mild regularity conditions, we may write the Edgeworth expansion of the distribution $p(\chi;\omega)$ of χ as

$$p(\chi;\omega) = \varphi_k(\chi;\iota)\{1 + Q_1(\chi) + Q_2(\chi)\} + O(n^{-3/2}) \tag{3.13}$$

where

$$Q_1(\chi) = \frac{1}{6}K_{\alpha,\beta,\gamma}H^{\alpha\beta\gamma}(\chi),$$
$$Q_2(\chi) = \frac{1}{24}K_{\alpha,\beta,\gamma,\delta}H^{\alpha\beta\gamma\delta}(\chi) + \frac{1}{72}K_{\alpha,\beta,\gamma}K_{\delta,\epsilon,\zeta}H^{\alpha\beta\gamma\delta\epsilon\zeta}(\chi)$$

and where the K-quantities denote the joint cumulants of χ and the H-quantities denote the contravariant version of the Hermite polynomials in χ with respect to the matrix ι. The terms Q_1 and Q_2 are respectively of order $n^{-1/2}$ and n^{-1}.

By integrating (3.13) with respect to l_* one obtains

$$p(a;\omega) = \varphi_{k-d}(a;g^{kl})\{1 + \tilde{\varphi}_1(a) + \tilde{\varphi}_2(a)\} + O(n^{-3/2}) \tag{3.14}$$

where

$$\tilde{\varphi}_1(a) = \frac{1}{6}K_{k,l,m}H^{klm}(a),$$
$$\tilde{\varphi}_2(a) = \frac{1}{24}K_{k,l,m,n}H^{klmn}(a) + \frac{1}{72}K_{k,l,m}K_{k',l',m'}H^{klmk'l'm'}(a).$$

Now, since

$$p(l_*;\omega \mid a) = \frac{p(\chi;\omega)}{p(a;\omega)} = \varphi_d(l_*;g_{rs}) \times \frac{1 + Q_1(\chi) + Q_2(\chi)}{1 + \tilde{\varphi}_1(a) + \tilde{\varphi}_2(a)} + \cdots$$

one obtains the following expansion for the conditional distribution of the score vector, using the expected geometry,

$$p(l_*;\omega \mid a) = \varphi_d(l_*;g_{rs})\{1 + C_1(\chi) + C_2(\chi)\} + O(n^{-3/2}) \tag{3.15}$$

where

$$C_1(\chi) = \frac{1}{6}\{K_{r,s,t}H^{rst}(l_*) + 3K_{r,k,l}H^{rkl}(\chi) + 3K_{r,s,k}H^{rsk}(\chi)\}$$

and

$$\begin{aligned} C_2(\chi) = {} & \frac{1}{24}\{K_{\alpha,\beta,\gamma,\delta}H^{\alpha\beta\gamma\delta}(\chi) - K_{k,l,m,n}H^{klmn}(a)\} \\ & + \frac{1}{72}\{K_{\alpha,\beta,\gamma}K_{\delta,\epsilon,\zeta}H^{\alpha\beta\gamma\delta\epsilon\zeta}(\chi) - K_{k,l,m}K_{k',l',m'}H^{klmk'l'm'}(a)\} \\ & + \frac{1}{36}K_{k,l,m}H^{klm}(a)\{K_{k',l',m'}H^{k'l'm'}(a) - K_{\alpha,\beta,\gamma}H^{\alpha\beta\gamma}(\chi)\}. \end{aligned}$$

It may be noticed that alternatively, at least in principle, we could apply the Edgeworth approximation of order two directly to the conditional distribution $p(l^*;\omega \mid a)$ of the score vector l_* given the ancillary statistic a. Thus, if we let

$$K_{r;s}(l_* \mid a), \qquad K_{r,s,t}(l_* \mid a), \qquad K_{r,s,t,u}(l_* \mid a)$$

denote the first conditional cumulants of the score vector, we obtain the following expression

$$\begin{aligned} (3.16)\ p(l_*;\omega \mid a) & \\ = {} & \varphi_d(l_*;K_{r,s}(l_* \mid a)) \\ & \times \left\{1 + \frac{1}{6}K_{r,s,t}(l_* \mid a)H^{rst}(l_*;K_{r,s}(l_* \mid a)) \right. \\ & \quad + \frac{1}{24}K_{r,s,t,u}(l_* \mid a)H^{rstu}(l_*;K_{r,s}(l_* \mid a)) \\ & \quad \left. + \frac{1}{72}K_{r,s,t}(l_* \mid a)K_{u,v,w}(l_* \mid a)H^{rstuvw}(l_*;K_{r,s}(l_* \mid a))\right\} \\ & + O(n^{-3/2}). \end{aligned}$$

Usually the cumulants and conditional cumulants entering (3.15) and (3.16) are not known explicitly and therefore they can only be found by approximation. See for example McCullagh (1987), for more precise settings.

3.3 *Conditional expansion for the score vector, using observed geometry*

In this subsection we shall derive an asymptotic expansion for the conditional distribution of l_* given an ancillary statistic a.

The starting point is an asymptotic expansion for the conditional distribution $p(\hat{\omega};\omega \mid a)$ of the maximum likelihood estimator $\hat{\omega}$, for fixed value of the ancillary a, which was obtained in Barndorff-Nielsen (1986*b*).

This expansion may be written as

$$(3.17)\qquad \begin{aligned} p(\hat\omega;\omega\mid a) &= p^*(\hat\omega;\omega\mid a)\{1+O(n^{-3/2})\}\\ &= \varphi_d(\hat\omega-\omega;\not{j}^{-1})\{1+A_1+A_2+O(n^{-3/2})\}.\end{aligned}$$

The terms A_1 and A_2 are given by

$$\begin{aligned} A_1 = &-\frac{1}{2}\delta^t\not{j}^{rs}(\not{l}_{rs,t}+\not{l}_{rst})+\frac{1}{2}\delta^{rst}(\not{l}_{rs,t}+\frac{2}{3}\not{l}_{rst}),\\ A_2 = &\frac{1}{24}\Big[-3(\delta^{tu}-\not{j}^{tu})\{2\not{j}^{rs}(\not{l}_{rstu}+\not{l}_{rst;u}+\not{l}_{rsu;t}+\not{l}_{rs;tu})\\ &+(2\not{j}^{rv}\not{j}^{sw}-\not{j}^{rs}\not{j}^{vw})(\not{l}_{rs;t}+\not{l}_{rst})(\not{l}_{vw;u}+\not{l}_{vwu})\}\\ &+(\delta^{rstu}-\not{j}^{rs}\not{j}^{tu}[3])\Big\{(3\not{l}_{rstu}+8\not{l}_{rst,u}+6\not{l}_{rs,tu})\\ &-6\not{j}^{vw}(\not{l}_{vw,u}+\not{l}_{vwu})\Big(\not{l}_{rs;t}+\frac{2}{3}\not{l}_{rst}\Big)\Big\}\\ &+3(\delta^{rstuvw}-\not{j}^{rs}\not{j}^{tu}\not{j}^{vw}[15])\Big(\not{l}_{rs,t}+\frac{2}{3}\not{l}_{rst}\Big)\Big(\not{l}_{uv;w}+\frac{2}{3}\not{l}_{uvw}\Big)\Big]\end{aligned}$$

where [3], [15] indicate appropriate permutations of the indices, $\delta^t=(\hat\omega-\omega)^t$, $\delta^{rs}=\delta^r\delta^s,\dots$, A_1 being of order $O(n^{-1/2})$ and A_2 being of order $O(n^{-1})$ under ordinary repeated sampling.

We shall derive from (3.17) a similar expansion for the score vector. In fact, given the ancillary statistic a, the maximum likelihood estimator $\hat\omega$ is in one-to-one correspondence with l_*, at least locally around ω, and the Jacobian of the transformation from $\hat\omega$ to l_* is $|l_{*;*}|$ where $l_{*;*}(\omega;\hat\omega,a)=[l_{r;s}(\omega;\hat\omega,a)]_{r,s=1,\dots,d}$. When $\hat\omega$ is close enough to the true parameter ω, it is quite reasonable to consider $l_{*;*}$ as invertible because for $\hat\omega\to\omega$ we have $l_{*;*}\to\not{j}$ which is invertible in great generality.

Hence we obtain from (3.17):

$$(3.18)\qquad p(l_*;\omega\mid a)=|l_{*;*}|^{-1}\{\varphi_d(\hat\omega-\omega;\not{j}^{-1})(1+A_1+A_2+O(n^{-3/2}))\}.$$

The next step consists in finding an asymptotic expansion for $\delta=\hat\omega-\omega$ in terms of $l^*=\not{j}^{-1}l_*$.

By Taylor expanding $l_r(\omega,\hat\omega)$ in its second argument $\hat\omega$ around ω, we obtain

$$(3.19)\qquad l_r(\omega;\hat\omega)=\not{l}_r(\omega)+\not{l}_{r;s}(\omega)\delta^s+\frac{1}{2}\not{l}_{r;st}(\omega)\delta^{st}+\frac{1}{6}\not{l}_{r;stu}\delta^{stu}+\cdots$$

which may be rewritten as

$$(3.20)\qquad l^r(\omega;\hat\omega)=\delta^r+\frac{1}{2}\not{l}^r_{uv}\delta^{uv}+\frac{1}{6}\not{l}^r_{uvw}\delta^{uvw}+\cdots.$$

On solving for δ^r one gets:

$$(3.21)\qquad \delta^r=l^r-\frac{1}{2}\not{l}^r_{uv}l^{uv}-\frac{1}{6}\{\not{l}^r_{uvw}-3\not{l}^r_{ut}\not{l}^t_{vw}\}l^{uvw}+\cdots$$

with $l^{uv} = l^u l^v$, $l^{uvw} = l^u l^v l^w \cdots$.

By inserting (3.21) in each term of (3.18) and Taylor expanding, we obtain, after some algebra, an expansion of $p(l_*; \omega \mid a)$ in terms of l^*. The last step consists in collecting terms into invariant terms. Details of all these calculations are given in the Appendix. The result is the following:

Under ordinary repeated sampling and with relative error of order $O(n^{-3/2})$

$$p(l_*; \omega \mid a) = \varphi_d(l_*; \not{j})\{1 + B_1 + B_2 + O(n^{-3/2})\} \tag{3.22}$$

with $B_1 = (1/6)\not{H}^{rst}\not{T}_{rst}$ and

$$B_2 = \frac{1}{24}\not{H}^{rstu}\left\{\not{T}_{rstu} - \frac{1}{2}\not{K}_{rs;tu}[6]\right\} - \frac{1}{4}\not{H}^{rs}\not{j}^{tu}\not{K}_{rs;tu} + \frac{1}{72}\not{H}^{rstuvw}\not{T}_{rst}\not{T}_{uvw},$$

B_1 and B_2 being of order $O(n^{-1/2})$ and $O(n^{-1})$ respectively. Here the $\not{H}$-quantities denote the contravariant version of the Hermite polynomials defined in terms of l_* and $\not{j}$. The other quantities are covariant tensors given by

$$\not{T}_{rst} = -(\not{\ell}_{rst} + \not{\ell}_{rs;t}[3]), \tag{3.23}$$

$$\not{K}_{rs;tu} = \not{\ell}_{rs;tu} - \not{\ell}_{rs;v}\not{\ell}_{w;tu}\not{j}^{vw}, \tag{3.24}$$

$$\not{T}_{rstu} = -\left(\not{\ell}_{rstu} + \not{\ell}_{rst;u}[4] + \frac{1}{2}(\not{\ell}_{rs;tu} + \not{\ell}_{rs;w}\not{T}_{tuv}\not{j}^{vw}[6])\right). \tag{3.25}$$

We shall discuss these various quantities in Section 5.

4. Expansions for the maximum likelihood estimator

4.1 *Expansions for the maximum likelihood estimator under curved exponential families*

For completeness, we mention here that a detailed study of the unconditional and conditional expansions for the maximum likelihood estimator, under curved exponential families and using the expected geometrical structure, has been made in Amari (1985) and Amari and Kumon (1983).

4.2 *Conditional expansion for the maximum likelihood estimator, using the expected geometrical structure*

From the viewpoint of invariance and geometrical interpretation, it is relevant to derive the conditional expansion for the maximum likelihood estimator $\hat{\omega}$ via the conditional expansion for the score vector. As mentioned in Subsection 2.5, using the (locally) one-to-one transformation $l_* \to \hat{\omega}$ yields an asymptotic expansion for the distribution of the maximum likelihood estimator, which is divided in two parts, a variant and an invariant.

In Barndorff-Nielsen (1986*b*), $\delta = \hat{\omega} - \omega$ is modified into the bias corrected form as $\delta' = \hat{\omega} - \omega - \mu_1$, where the bias term μ_1 is given by

$$\mu_1^\alpha = \frac{1}{2}\not{j}^{\alpha r}\not{j}^{st}\not{\ell}_{r;st} = \frac{1}{2}\overset{1}{\not{\Gamma}}{}^{\alpha}_{st}\not{j}^{st} \tag{4.1}$$

and is of order n^{-1}.

Now, making the one-to-one transformation $l_* \to \hat{\omega}$, one obtains from (3.22)

$$
\begin{aligned}
(4.2) \quad & p(\delta'; \omega \mid a) \\
&= |l_{*;*}| \varphi_d(l_*; \jmath)\{1 + B_1 + B_2 + O(n^{-3/2})\} \\
&= \varphi_d(\delta'; \jmath^{-1})(\varphi_d(l_*; \jmath)/\varphi_d(\delta'; \jmath^{-1}))|l_{*;*}|\{1 + B_1 + B_2 + \cdots\}
\end{aligned}
$$

where

$$
\varphi_d(l_*; \jmath)/\varphi_d(\delta'; \jmath^{-1}) = |\jmath|^{-1} \exp\left\{\frac{1}{2}(\delta'^{rs} - l^{rs})\jmath_{rs}\right\}.
$$

Next, we derive expansions for $\exp\{(1/2)(\delta'^{rs} - l^{rs})\jmath_{rs}\}$ and $|l_{*;*}||\jmath^{-1}|$ by inserting (3.22) in these two expressions. The explicit forms of these expansions are given in the Appendix. The last step consists again in Taylor expanding and then collecting terms of the same order.

Finally one obtains the following expansion for the bias-corrected maximum likelihood estimator:

$$
(4.3) \qquad p(\delta'; \omega \mid a) = \varphi_d(\delta'; \jmath^{-1})\{1 + B_1 + B_2 + \cdots\}\{1 + C_1 + C_2 + \cdots\}
$$

where $C_1 = -(1/2)\not{\Gamma}_{t;rs}\not{H}^{rst}$ is of order $n^{-1/2}$, and

$$
\begin{aligned}
C_2 = &-\frac{1}{2}\not{\Gamma}_{r;stu}(l^{rstu} - [3]\jmath^{rs}l^{tu}) \\
&+ \frac{1}{8}\not{\Gamma}_{t;rs}\not{\Gamma}_{u;vw}(l^{rstuvw} + \jmath^{rs}\jmath^{tu}\jmath^{vw} + 4l^{sw}\jmath^{ru}\not{H}^{tv} \\
&\qquad - 4l^{tv}\jmath^{uw}\not{H}^{rs} - l^{tu}\jmath^{vw}\not{H}^{rs} + l^{rsvw}\jmath^{tu} \\
&\qquad - 2l^{vw}\jmath^{rs}\jmath^{tu} + 4l^{sv}\jmath^{tr}\jmath^{uw} - 4l^{vw}\jmath^{ts}\jmath^{ru})
\end{aligned}
$$

is of order n^{-1}.

The symbol [ ] indicates here a sum of 3 similar terms obtained by permuting the indices r, s, t.

It may be noted that $l^{rstu} - [3]\jmath^{rs}l^{tu}$ is a contravariant tensor of rank 4 in ω. Also, the factor in C_2 multiplied by the quantity $\not{\Gamma}_{t;rs}\not{\Gamma}_{u;vw}$ is a contravariant tensor of rank 6.

The expansion we have obtained involves the product of two expansions, the first of which (with terms $B_1, B_2, \ldots$) is invariant. The second is related to the observed (-1)-connection.

Thus, if we use the special coordinate system given in (2.21) as a parametrization for the model, the quantities $\overset{-1}{\not{\Gamma}}{}^{i}_{rst}$, $\overset{-1}{\not{\Gamma}}{}^{i}_{rstu}, \ldots$ can be made to vanish. With this choice of parametrization, the expansion for $\hat{\omega}$ reduces to

$$
(4.4) \qquad p(\delta'; \omega \mid a) = \varphi_d(\delta'; \jmath^{-1})\{1 + B_1 + B_2 + \cdots\},
$$

since in such a coordinate system one has

$$
\not{\Gamma}_{t;rs} = \jmath_{tt'}\not{\Gamma}^{t'}_{rs} = j_{tt'}\delta^{t'}_{rs} = 0
$$

and

$$
\not{\Gamma}_{r;stu} = \jmath_{rr'}\not{\Gamma}^{r'}_{stu} = \jmath_{rr'}\delta^{r'}_{stu} = 0.
$$

5. Interpretation of terms

The aim of this section is to discuss the geometrical aspects of the different terms involved in the various expansions we have been considering.

In the expansion (3.4) for the score vector all the correction terms are composed of tensors, as has been already pointed out in Subsection 3.1.

In the case when we consider a curved exponential model, the expansion (3.15) for the score vector obtained by using the expected geometry has not been interpreted yet.

In the expansion (3.22), using the observed geometry, other types of tensors are considered.

First the observed skewness tensor $\not{T}_{rst}$ given by (2.17), which is analogous to the expected skewness tensor T_{rst}.

Two other tensors, $\not{T}_{rstu}$ and $\not{K}_{rs;tu}$, given in formulas (3.25) and (3.24), appear in the expansion.

The tensor $\not{T}_{rstu}$ was introduced by Barndorff-Nielsen (1986*b*). It is obtained from the tensor $\not{T}_{rst}$ by covariant differentiation with respect to the (1)-observed connection $\overset{1}{\not{\Gamma}}$ and by symmetrization, namely $\not{T}_{rstu} = (1/4)(\overset{1}{\not{D}}_u(\not{T}_{rst})[4]$. The $\not{K}$-tensor was introduced by Barndorff-Nielsen (1986*a*) in a paper discussing strings and tensorial combinants. In that paper, the $\not{K}$-tensor is derived from the observed (1)-connection strings by a construction referred to as "intertwining" of strings.

The $\not{K}$-tensor may be related to the observed (1)- and (-1)-connection strings in different ways and in particular to the observed (1) and (-1) Riemannian curvature tensors.

First, consider $\overset{1}{\not{\Gamma}}$ and let us introduce the curvature tensor $\overset{1}{\not{R}}_{rstu}$ associated with $\overset{1}{\not{\Gamma}}$.

By definition,

$$\overset{1}{\not{R}}_{rstu} = (\partial_{\omega^r}\overset{1}{\not{\Gamma}}{}^v_{st} - \partial_{\omega^s}\overset{1}{\not{\Gamma}}{}^v_{rt})\not{j}_{vu} + (\overset{1}{\not{\Gamma}}_{rvu}\overset{1}{\not{\Gamma}}{}^v_{st} - \overset{1}{\not{\Gamma}}_{svu}\overset{1}{\not{\Gamma}}{}^v_{rt}). \tag{5.1}$$

Since

$$\partial_{\omega^r}\overset{1}{\not{\Gamma}}{}^v_{st} = \partial_{\omega^r}\not{j}^{vu}\not{\ell}_{st;u} + \not{j}^{vw}(\not{\ell}_{str;w} + \not{\ell}_{st;rw}),$$

and

$$\overset{1}{\not{\Gamma}}_{rvu}\overset{1}{\not{\Gamma}}{}^v_{st} = \not{\ell}_{rv;u}\not{\ell}_{st;w}\not{j}^{vw},$$

one obtains

$$\begin{aligned}\overset{1}{\not{R}}_{rstu} = {} & \not{j}^{vw}\{(\not{\ell}_{vur} + \not{\ell}_{vu;r} + \not{\ell}_{vr;u})\not{\ell}_{st;w} - (\not{\ell}_{vus} + \not{\ell}_{vu;s} + \not{\ell}_{vs;u})\not{\ell}_{rt;w}\} \\ & + \not{\ell}_{st;ru} - \not{\ell}_{rt;su}.\end{aligned}$$

By using the equation $\not{\ell}_{rst} + \not{\ell}_{rs;t} + \not{\ell}_{rt;s} + \not{\ell}_{r;st} = 0$, one finally gets

$$\overset{1}{\not{R}}_{rstu} = \not{K}_{st;ru} - \not{K}_{rt;su}. \tag{5.2}$$

In the same way, considering $\overset{-1}{\not\Gamma}$ and $\overset{-1}{\not R}_{rstu}$, one finds

$$\overset{-1}{\not R}_{rstu} = \not h_{ru;st} - \not h_{su;rt}. \tag{5.3}$$

There is a similar relationship for corresponding expected quantities from curved exponential families $R^1_{rstu} = H_{st;ru} - H_{rt;su}$ where $H_{st;ru}$ is the inner product of the (1) and (−1) imbedding curvature tensors, i.e. $H_{st;ru} = H^{1k}_{st} H^{-1l}_{ru} q_{kl}$ where H^{1k}_{st} and H^{-1l}_{ru} are the components of the imbedding curvature for the (1)- and (−1)-connection, respectively (see Amari (1985)). This interpretation can be extended beyond curved exponential family models by using Amari's idea of local exponential family, see Amari *et al.* (1987); see also Vos (1989). Because of the structure of the $\not h$-tensor, it is not obvious that a similar interpretation of the $\not h$-tensor may be derived for the observed curved exponential approximation. A clearly geometric meaning of the $\not h$-tensor is still missing.

However, it is possible to relate the $\not h$-tensor to the observed (1)- and (−1)-connection strings also in the following way.

Since

$$\begin{aligned}
\not h_{rs;tu} + \not l_{rst;u} &= \partial_{\omega^t} \overset{1}{\not\Gamma}{}^v_{rs} \not j_{vu} + \overset{1}{\not\Gamma}_{tvu} \overset{1}{\not\Gamma}{}^v_{rs} \\
&= (\partial_{\omega^t} \overset{1}{\not\Gamma}{}^v_{rs} + \overset{1}{\not\Gamma}{}^w_{rs} \overset{1}{\not\Gamma}{}^v_{tw}) \not j_{vu},
\end{aligned}$$

we have, denoting covariant differentiation relative to $\overset{\alpha}{\not\Gamma}$ and with respect to ω^t by $\not D^\alpha_t$,

$$\begin{aligned}
\not h_{rs;tu} &= \overset{1}{\not D}_t(\overset{1}{\not\Gamma}{}^v_{rs}) \not j_{vu} - \not l_{rst;u} \\
&= \{\overset{1}{\not D}_t(\overset{1}{\not\Gamma}{}^v_{rs}) - \not l^{;v}_{rst}\} \not j_{vu} \\
&= \{\overset{1}{\not D}_t(\not l^{;v}_{rs}) - (\not l^{;v}_{rst})\} \not j_{vu}.
\end{aligned} \tag{5.4}$$

We get a similar expression for the observed (−1)-connection, namely

$$\not h_{tu;rs} = \{\not D^1_t(l^v_{rs}) - l^v_{rst}\} \not j_{vu}. \tag{5.5}$$

Therefore formula (5.4) expresses the $\not h$-tensor as the difference between $\not D^1_t(\not l^{;v}_{rs})$ which belongs to the canonical string generated by $\overset{1}{\not\Gamma}$ and the corresponding term in the connection string $\overset{1}{\not\Gamma}$. Except for this, we could not find any satisfying geometrical interpretation of this $\not h$-tensor.

One has an analogous result for the $\not h$-tensor expressed in terms of the observed (−1)-connection.

We may notice that for (k, k) exponential models one has $\not T_{rst} = K_{r,s,t}$, $\not T_{rstu} = K_{r,s,t,u}$ and $\not h_{rs;tu} \equiv 0$. The conditional expansions for the score vector, given in (3.16) and (3.22), are quite similar. We may relate the observed tensors $\not{T}_{rst}$ and $\not{T}_{rstu} - (1/2)\not{K}_{rs;tu}[6]$ to the cumulants of order 3 and 4 in (3.16), respectively (for (k, k) exponential families these observed tensors reduce to the observed cumulants of order 3 and 4). The extra term in (3.22) which is now found in (3.16) should be ascribed to the different geometrical structure involved.

Finally we shall comment on the variant part of the expansion (4.3) obtained for the conditional distribution of the maximum likelihood estimator.

The correction term of order $n^{-1/2}$ contains the Hermite polynomial $\not{H}^{rst}(l_*; \not{j})$ and the term $\not{Y}_{t;rs} = \overset{-1}{\not{\Gamma}}_{rst}$. In the correction term of order n^{-1} we collected terms in order to emphasize the various terms $\not{Y}_{r;st}$, $\not{Y}_{r;stu}$ entering it and which belong to the (-1)-connection string. The remaining terms are not explained yet. In Amari and Kumon (1983), they give an expansion for the conditional distribution of the maximum likelihood estimator with correction terms of order $n^{-1/2}$ and n^{-1}. It may be noted that only the correction term of order $n^{-1/2}$ is given a geometrical interpretation.

6. Concluding remarks

The score vector appears to be a main tool in constructions of expansions that can be expressed in terms of invariant and geometrical quantities. Nevertheless the last expansion obtained in (4.3) is not quite satisfactory, since all the terms contained have not yet been given a clear geometrical interpretation. The question is how to build a procedure, based on the intrinsic geometrical properties of the score vector and the theory of strings, giving a geometrical interpretation.

Another point concerns the link between the order of the correction terms appearing in the expansions obtained above and the corresponding tensors contained in these correction terms. It appears that for approximations with error of order n^{-1} the tensors required are the metric tensor and the skewness tensor, in both the expected case and the observed case. If we consider correction terms of order n^{-1} other tensors appear, namely the $\not{K}$ and $\not{T}_{rstu}$ tensors, which are again related to the observed geometrical structure defined on the model. Thus, these tensors seem to be fundamental geometrical objects but in a sense which has not been fully elucidated yet.

Acknowledgements

I am indebted to the referees for their thorough readings of the paper and very constructive comments.

Appendix

PROOF OF FORMULA (3.22). From (3.18) and (3.21) we obtain the following expansions:

$$1)\ ((|l_{*;*}|)/(|\not{j}|)^{-1}) = 1 - \not{j}^{rs}\not{Y}_{r;st}l^t + \frac{1}{2}\{\not{j}^{rs}\not{Y}_{r;s\tau}\not{Y}^{\tau}_{tu} + \not{j}^{sv}\not{j}^{wr}\not{Y}_{v,wu}\not{Y}_{r;st} - \not{j}^{rs}\not{Y}_{r;stu} + \not{j}^{rs}\not{j}^{vw}\not{Y}_{r;st}\not{Y}_{v,wu}\}l^{tu} + \cdots .$$

Here we use the differentiation formulas: $\partial_r \log|A| = a^{ji}\partial_r a_{ij}$ and $\partial_r a^{ij} = -a^{ik}a^{lj}\partial_r a_{kl}$ where $A = [a_{ij}]$ is any positive definite matrix and $A^{-1} = [a^{ij}]$.

2)
$$\begin{aligned}
\exp\Big\{-\frac{1}{2}\delta^{rs}\jmath_{rs}\Big\} = \exp\Big\{-\frac{1}{2}l^{rs}\jmath_{rs}\Big\} & \\
\times\Big\{1 + \frac{1}{2}\nu_{r;st}l^{rst} &- \frac{1}{8}\nu_{\tau;rs}\nu^{\tau}_{tu}l^{rstu} \\
&+ \frac{1}{6}(\nu_{r;stu} - 3\nu_{r;s\tau}\nu^{\tau}_{tu})l^{rstu} \\
&+ \frac{1}{8}\nu_{r;st}\nu_{u;vw}l^{rstuvw} + O(n^{-3/2})\Big\}.
\end{aligned}$$

3)
$$\begin{aligned}
A_1 = &-\frac{1}{2}\jmath^{rs}(\nu_{rs;t} + \nu_{rst})l^t + \frac{1}{2}l^{rst}\Big(\nu_{rs;t} + \frac{2}{3}\nu_{rst}\Big) \\
&+ \frac{1}{4}\{\jmath^{rs}l^{tu}(\nu_{rs;\tau} + \nu_{rs\tau})\nu^{\tau}_{tu} \\
&\qquad - (2\nu_{rs\tau} + \nu_{rs;\tau}[3])\nu^{\tau}_{tu}l^{rstu} + O(n^{-3/2})\}.
\end{aligned}$$

4) The term A_2, of order n^{-1}, in (3.17) is rewritten in terms of $l^r \dots$, just by making the substitutions $\delta^r \to l^r$, $\delta^{rs} \to l^{rs} \dots$. Finally we obtain

$$p(l_*;\omega \mid a) = \varphi_d(l_*;\jmath)\{1 + B_1 + B_2 + \cdots\}$$

where

$$\begin{aligned}
B_1 = &-\frac{1}{2}\jmath^{rs}l^t(\nu_{rs;t} + \nu_{rst}) + \frac{1}{2}l^{rst}\Big(\nu_{rs;t} + \frac{2}{3}\nu_{rst}\Big) \\
&- \jmath^{rs}l^t\nu_{r;st} + \frac{1}{2}l^{rst}\nu_{rs;t}
\end{aligned}$$

is of order $n^{-1/2}$ under ordinary repeated sampling of size n, and

$$\begin{aligned}
B_2 = &-\frac{1}{8}(l^{tu} - \jmath^{tu})\{2\jmath^{rs}(\nu_{rstu} + \nu_{rst;u} + \nu_{rsu;t} + \nu_{rs;tu}) \\
&\qquad + (2\jmath^{rv}\jmath^{sw} - \jmath^{rs}\jmath^{vw})(\nu_{rs;t} + \nu_{rst})(\nu_{vw;u} + \nu_{vwu})\} \\
&+ \frac{1}{4}\jmath^{rs}l^{tu}(\nu_{rs;\tau} + \nu_{rs\tau})\nu^{\tau}_{tu} + \frac{1}{2}\jmath^{rs}l^{tu}\nu_{r;s\tau}\nu^{\tau}_{tu} \\
&+ \frac{1}{2}\jmath^{sv}\jmath^{wr}l^{tu}\nu_{v;wu}\nu_{r;st} - \frac{1}{2}\jmath^{rs}l^{tu}\nu_{r;stu} \\
&+ \frac{1}{2}\jmath^{rs}l^{tu}\jmath^{vw}\nu_{r;st}\nu_{v;wu} + \frac{1}{2}\jmath^{rs}l^{tu}(\nu_{rs;t} + \nu_{rst})\nu_{v;wu}\jmath^{vw} \\
&+ \frac{1}{24}(l^{rstu} - \jmath^{rs}\jmath^{tu}[3])\Big\{(3\nu_{rstu} + 8\nu_{rst;u} + 6\nu_{rs;tu}) \\
&\qquad - 6\jmath^{vw}(\nu_{vw;u} + \nu_{vwu})\Big(\nu_{rs;t} + \frac{2}{3}\nu_{rst}\Big)\Big\}
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}l^{rstu}\left(\not{Y}_{rs;t}+\frac{2}{3}\not{Y}_{rst}\right)\not{Y}_{v;wu}\not{j}^{vw}\\
&-\frac{1}{2}l^{rstu}\not{Y}_{r;st}\not{Y}_{v;wu}\not{j}^{vw}-\frac{1}{2}l^{rstu}\not{Y}_{r;st}(\not{Y}_{vw;u}+\not{Y}_{vwu})\not{j}^{vw}\\
&-\frac{1}{8}l^{rstu}\not{Y}_{\tau;rs}\not{Y}^{\tau}_{tu}+\frac{1}{6}l^{rstu}(\not{Y}_{r;stu}-3\not{Y}_{r;s\tau}\not{Y}^{\tau}_{tu})\\
&-\frac{1}{4}l^{rstu}(2\not{Y}_{rs\tau}+\not{Y}_{rs;\tau}[3])\not{Y}^{\tau}_{tu}\\
&+\frac{1}{8}(l^{rstuvw}-\not{j}^{rs}\not{j}^{tu}\not{j}^{vw}[15])\left(\not{Y}_{vw;u}+\frac{2}{3}\not{Y}_{vwu}\right)\left(\not{Y}_{rs;t}+\frac{2}{3}\not{Y}_{rst}\right)\\
&+\frac{1}{4}l^{rstuvw}\left(\not{Y}_{rs;t}+\frac{2}{3}\not{Y}_{rst}\right)\not{Y}_{v;wu}+\frac{1}{8}l^{rstuvw}\not{Y}_{r;st}\not{Y}_{u;vw}
\end{aligned}
$$

is of order n^{-1}.

The last step consists in collecting terms such that geometrical quantities appear.

For B_1 it is easily seen that

$$B_1=\frac{1}{6}\not{H}^{rst}(l_*;\not{j})\not{T}_{rst}.$$

For the terms of order n^{-1} the calculations are longer but the expression of B_2 finally reduces to

$$
\begin{aligned}
B_2=&\frac{1}{24}\not{H}^{rstu}\left(\not{T}_{rstu}-\frac{1}{2}\not{h}_{rs;tu}[6]\right)-\frac{1}{2}\not{H}^{rs}\not{j}^{tu}\not{h}_{rs;tu}\\
&+\frac{1}{72}\not{H}^{rstuvw}\not{T}_{rst}\not{T}_{uvw}.
\end{aligned}
$$

PROOF OF FORMULA (4.3).

$$(\varphi_d(l_*;\not{j})/(\varphi_d(\delta';\not{j}^{-1})))=|\not{j}|^{-1}\exp\left\{\frac{1}{2}(\delta'^{rs}-l^{rs})\not{j}_{rs}\right\}.$$

Introducing the expansion (3.21) for δ in this expression and developing leads to

$$(\varphi_d(l_*;\not{j})/(\delta_d(\delta';\not{j}^{-1})))=|\not{j}|^{-1}\{1+A'_1+A'_2+\cdots\}$$

where

$$A'_1=-\frac{1}{2}\not{Y}_{r;st}l^{rst}+\frac{1}{2}\not{Y}_{t;rs}\not{j}^{rs}l^{t}$$

is of order $n^{-1/2}$ and

$$
\begin{aligned}
A'_2=&-\frac{1}{6}\not{Y}_{r;stu}l^{rstu}\\
&+\frac{1}{8}l^{rstu}(4\not{Y}_{r;sv}\not{Y}_{w;tu}\not{j}^{vw}+\not{Y}_{v;rs}\not{Y}_{w;tu}\not{j}^{vw}-\not{Y}_{r;st}\not{Y}_{u;vw}\not{j}^{vw})\\
&+\frac{1}{8}l^{tu}\not{j}^{rs}\not{j}^{vw}(-2\not{Y}_{v;rs}\not{Y}_{w;tu}+\not{Y}_{t;rs}\not{Y}_{u;vw})\\
&+\frac{1}{8}\not{Y}_{r;st}\not{Y}_{u;vw}l^{rstuvw}+\frac{1}{8}\not{Y}_{t;rs}\not{Y}_{u;vw}\not{j}^{rs}\not{j}^{tu}\not{j}^{vw}
\end{aligned}
$$

is of order n^{-1}.

Using the same method for $|l_{*;*}||\jmath|^{-1}$ we get $|l_{*;*}||\jmath|^{-1} = 1 + A_1'' + A_2'' + \cdots$ where $A_1'' = \jmath^{rs} l^t \not{y}_{t;rs}$ is of order $n^{-1/2}$ and

$$\begin{aligned} A_2'' = &\frac{1}{2} l^{tu} \jmath^{rs} \jmath^{vw} (\not{y}_{r;st} \not{y}_{v;wu} - \not{y}_{r;wu} \not{y}_{v;st} - \not{y}_{r;sv} \not{y}_{w;tu}) \\ &+ \frac{1}{2} \jmath^{rs} l^{tu} \not{y}_{r;stu} \end{aligned}$$

is of order n^{-1}.

Inserting these two expressions in formula (4.2), developing and collecting terms yields the final result (4.3).

References

Amari, S. I. (1985). Differential-geometrical methods in statistics, *Lecture Notes in Statistics*, **28**, Springer, Heidelberg.

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

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

Barndorff-Nielsen, O. E. (1980). Conditionality resolution, *Biometrika*, **67**, 293–310.

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika*, **70**, 856–873.

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

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

Barndorff-Nielsen, O. E. (1987). Likelihood, ancillary and strings, *Proceedings of the First World Congress of the Bernoulli Society*, **2**, 205–213.

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

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

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

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

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Blæsild, P. (1990). Yokes: orthogonal and observed normal coordinates, Research Report, 205, Department of Theoretical Statistics, Aarhus University.

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

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

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

Skovgaard, I. M. (1986). On multivariate Edgeworth expansions, *Internat. Statist. Rev.*, **54**, 169–186.

Vos, P. W. (1989). Fundamental equations for statistical submanifolds with applications to the Bartlett correction, *Ann. Inst. Statist. Math.*, **41**, 429–450.