Ann. Inst. Statist. Math.
Vol. 41, No. 2, 289–304 (1989)

THE INFORMATION MATRIX, SKEWNESS TENSOR AND α-CONNECTIONS FOR THE GENERAL MULTIVARIATE ELLIPTIC DISTRIBUTION

ANN F. S. MITCHELL

Department of Mathematics, Imperial College of Science, Technology and Medicine, Huxley Building, 180 Queen's Gate, London SW7 2BZ, U.K.

(Received November 30, 1987; revised May 30, 1988)

Abstract. Expressions for the entries of the information matrix and skewness tensor of a general multivariate elliptic distribution are obtained. From these the coefficients of the α-connections are derived. A general expression for the asymptotic efficiency of the sample mean, when appropriate as an estimator of the location parameter, is obtained. The results are illustrated by examples from the multivariate normal, Cauchy and Student's t-distributions.

Key words and phrases: Multivariate elliptic distributions, information matrix, skewness tensor, α-connections, sample mean, asymptotic efficiency, normal, Cauchy and Student's t-distributions.

1. Introduction

In recent work on multivariate analysis there has been some emphasis on the class of elliptic distributions as a source of alternative models to the normal. Chmielewski (1981) provides a review and bibliography of specific results and areas of application.

Most of the statistical work reviewed involves testing problems, although Maronna (1976) has studied the robust estimation of the location and scale parameters. Maximum likelihood estimation arises mainly in the context of likelihood ratio tests. Some work on finding maximum likelihood estimates is described by Maronna (1976) and Hsu (1985). It would appear that no work has been done on finding general simple expressions for the entries of the information matrix. Its inverse is the asymptotic variance-covariance matrix of the maximum likelihood estimators under suitable regularity conditions and is relevant in determining asymptotic efficiencies of estimators of parameters.

The entries of the information matrix also play an important role in the differential geometry approach to statistical inference, since they define a covariant, symmetric tensor of the second degree. Together with the skewness tensor, they define the α-connections basic to the general approach of Amari (1985) and are fundamental to the study of higher asymptotic properties of estimators, tests and confidence intervals for exponential models. Some results for elliptic distributions in special cases have been given by Mitchell and Krzanowski (1985) and Mitchell (1987, 1988).

It is the purpose of this paper to derive simple, general expressions for the entries of the information matrix and the skewness tensor of a general multivariate elliptic distribution. From these, expressions for the coefficients of the α-connections will be deduced. This provides a basis for future studies of the α-geometry of the class of multivariate elliptic distributions, whose members are not exclusively exponential models, and for the resolution of related statistical inference problems. The results for the information matrix are used to find a general expression for the asymptotic efficiency of the sample mean as an estimator of the location parameter, when the location parameter is the mean. The results will be illustrated by some particular well-known multivariate elliptic distributions. We began with some simplifying probability results.

2. Some simple probability results for elliptic distributions

A p-dimensional random variable X is said to have an elliptic distribution with parameters $\mu^T = (\mu_1, \mu_2, \dots, \mu_p)$ and Ψ, a $p \times p$ positive-definite matrix, if its density is of the form

$$f_h(x|\mu, \Psi) = (\det \Psi)^{-1/2} h\{(x-\mu)^T \Psi^{-1} (x-\mu)\} \tag{2.1}$$

for some function h. We say that X has an $EL_p^h(\mu, \Psi)$ distribution. Its characteristic function is

$$\varphi_h(t) = \exp(it^T\mu)\psi_h(t^T\Psi t)$$

for some function ψ_h. Provided they exist,

$$E(X) = \mu \quad \text{and} \quad \operatorname{cov}(X) = k_h \Psi ,$$

where k_h is a constant given by

$$k_h = -2\left[\frac{d}{du}\psi_h(u)\right]_{u=0} .$$

The class of elliptic distributions includes the normal, Student's t, Cauchy, or, more generally, the Pearson VII, and the generalized Laplace or Bessel distributions. For a $N_p(\mu, \Omega)$ distribution, for example,

$$h(u) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}u\right), \qquad \Psi = \Omega\,,$$

$$\psi_h(u) = \exp\left(-\frac{1}{2}u\right), \qquad k_h = 1\,.$$

Properties of elliptic distributions have been studied by Lord (1954) and Kelker (1970) and are summarized by Muirhead ((1982), pp. 32–40). In particular, any $EL_p^h(\mu, \Psi)$ random variable X can be reduced to an $EL_p^h(0, I)$ random variable Z by the transformation

$$Z = P^{-1}(X-\mu)\,, \tag{2.2}$$

where I is the $p \times p$ identity matrix and P is a non-singular $p \times p$ matrix such that

$$P^T \Psi^{-1} P = I\,. \tag{2.3}$$

Such a random variable Z is said to have a spherical distribution. If $Z^T = (Z_1, Z_2, \ldots, Z_p)$, the Z_i have identical univariate spherical distributions, but are not independent unless Z is normal. Any odd function of the Z_i will have zero expectation, provided the expectation exists.

When $Z^T = (Z_1, Z_2, \ldots, Z_p)$ is an $EL_p^h(0, I)$ random variable, it is well-known that

$$\|Z\| = (Z^T Z)^{1/2} \quad \text{and} \quad V = Z/\|Z\|$$

are independent, V having a uniform distribution on a unit sphere. Moreover, if $V^T = (V_1, V_2, \ldots, V_p)$, it is clear that, for any positive integers r, s and t, $E(V_i^r)$, $E(V_i^r V_j^s)$ and $E(V_i^r V_j^s V_k^t)$ ($i \neq j$; $j \neq k$; $i \neq k$; $i, j, k = 1, 2, \ldots, p$) can be found assuming Z to be $N_p(0, I)$. Using this and the independence of $\|Z\|$ and V, it is easy to show that

$$E(V_i^r) = \begin{cases} \dfrac{(r-1)(r-3)\cdots 1}{p(p+2)\cdots(p+r-2)} & r \text{ even,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.4}$$

$$\begin{aligned} & E(V_i^r V_j^s) \\ &\quad = \begin{cases} \dfrac{(r-1)(r-3)\cdots 1(s-1)(s-3)\cdots 1}{p(p+2)\cdots(p+r+s-2)} & r, s \text{ both even,} \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{2.5}$$

and

(2.6) $$E(V_i^r V_j^s V_k^t) = \begin{cases} \dfrac{(r-1)(r-3)\cdots 1(s-1)(s-3)\cdots 1(t-1)(t-3)\cdots 1}{p(p+2)\cdots(p+r+s+t-2)} & r,\ s,\ t \text{ all even,} \\ 0 & \text{otherwise.} \end{cases}$$

The results obviously do not depend on i, j and k. The proof follows that of Anderson and Stephens (1972) for the cases (2.4) with $r = 4$, and (2.5) with $r = 2$, $s = 2$ and $r = 3$, $s = 1$.

In deriving expressions for the entries of the information matrix of an $EL_p^h(\mu, \Psi)$ distribution and, indeed, in more advanced differential geometric descriptions of such distributions, we come face to face with the random variable

$$W = \frac{d \log h(\|Z\|^2)}{d(\|Z\|^2)} .$$

In particular, moments of the following type are often encountered, viz.,

$$E(Z_i^r W^l), \quad E(Z_i^r Z_j^s W^l) \quad \text{and} \quad E(Z_i^r Z_j^s Z_k^t W^l),$$

for positive integers r, s, t and l and $i, j, k = 1, 2, \ldots, p$; $i \neq j$, $j \neq k$ and $i \neq k$. Using the same type of argument as for (2.4), (2.5) and (2.6), these can be written in terms of (2.4), (2.5) and (2.6) as

(2.7) $$E(Z_i^r W^l) = E(V_i^r) E(\|Z\|^r W^l),$$

(2.8) $$E(Z_i^r Z_j^s W^l) = E(V_i^r V_j^s) E(\|Z\|^{r+s} W^l)$$

and

(2.9) $$E(Z_i^r Z_j^s Z_k^t W^l) = E(V_i^r V_j^s V_k^t) E(\|Z\|^{r+s+t} W^l).$$

In evaluating the non-zero moments for which r, s and t are all even, it can be useful to work with the distribution of $S = \|Z\|^2$, whose density function is well-known and given by

$$f_{h,S}(s) = \begin{cases} \dfrac{\pi^{p/2} s^{p/2-1} h(s)}{\Gamma\left(\frac{1}{2}p\right)} & s > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Moments of the type $E(S^q W^l) = E(\|Z\|^{2q} W^l)$, for positive integer q, can be found very easily in special cases like the normal, Cauchy, Student's t and Pearson VII. For example, in the normal case,

$$E\{\|Z\|^{2q} W^l\} = \frac{(-1)^l 2^{q-l} \Gamma\left(q + \frac{1}{2}p\right)}{\Gamma\left(\frac{1}{2}p\right)} \tag{2.10}$$

and, in the case of Student's t on k d.f. ($k = 1$ giving Cauchy) provided $2l - 2q + k > 0$,

$$E\{\|Z\|^{2q} W^l\} = \frac{(-1)^l (k+p)^l \Gamma\left(\frac{1}{2}k + \frac{1}{2}p\right) B\left(\frac{1}{2}p + q, l + \frac{1}{2}k - q\right)}{2^l k^{l-q} \Gamma\left(\frac{1}{2}p\right) \Gamma\left(\frac{1}{2}k\right)}. \tag{2.11}$$

It is also interesting to note that, whatever the member of the elliptic class under consideration,

$$E(\|Z\|^2 W) = (-p)/2. \tag{2.12}$$

Moreover, from (2.7), (2.8), (2.4) and (2.5), it is easy to see that, for even r, s and t,

$$E(Z_i^{r+s} W^l) = \frac{(r+s-1)(r+s-3)\cdots 1}{(r-1)(r-3)\cdots 1(s-1)(s-3)\cdots 1} E(Z_i^r Z_j^s W^l) \tag{2.13}$$

and, similarly, using (2.9) and (2.6) also,

$$E(Z_i^{r+s+t} W^l) = \frac{(r+s+t-1)(r+s+t-3)\cdots 1}{(r-1)(r-3)\cdots 1(s-1)(s-3)\cdots 1(t-1)(t-3)\cdots 1} \cdot E(Z_i^r Z_j^s Z_k^t W^l). \tag{2.14}$$

There are many identities of the above type, but those in (2.13) and (2.14) are useful in deriving simple expressions for the expectations of $(Z^T A Z) W^l$, $(Z^T A Z)(Z^T B Z) W^l$ and $(Z^T A Z)(Z^T B Z)(Z^T C Z) W^l$, where A, B and C are symmetric matrices with constant entries a_{ij}, b_{ij} and c_{ij}, respectively, (i, $j = 1, 2, \ldots, p$). These expectations arise naturally in the future development. Before deriving them we introduce the following notation,

$$d(A, B) = \sum_{i=1}^{p} a_{ii} b_{ii}$$

and

$$d(A, B, C) = \sum_{i=1}^{p} a_{ii} b_{ii} c_{ii} .$$

From (2.4), (2.5), (2.7) and (2.8), we obtain at once

$$(2.15) \qquad E(Z^T A Z W^l) = \operatorname{tr}(A) E(Z_i^2 W^l) = \frac{1}{p} \operatorname{tr}(A) E(\|Z\|^2 W^l)$$

and, from first principles,

$$\begin{aligned} E\{(Z^T A Z)(Z^T B Z) W^l\} &= d(A, B)\{E(Z_i^4 W^l) - 3E(Z_i^2 Z_j^2 W^l)\} \\ &\quad + \{2 \operatorname{tr}(AB) + \operatorname{tr}(A) \operatorname{tr}(B)\} E(Z_i^2 Z_j^2 W^l) , \end{aligned}$$

which, using (2.13), (2.8) and (2.5), gives

$$\begin{aligned} (2.16) \quad & E\{(Z^T A Z)(Z^T B Z) W^l\} \\ &= \{2 \operatorname{tr}(AB) + \operatorname{tr}(A) \operatorname{tr}(B)\} E(Z_i^2 Z_j^2 W^l) \\ &= \frac{1}{p(p+2)} \{2 \operatorname{tr}(AB) + \operatorname{tr}(A) \operatorname{tr}(B)\} E(\|Z\|^4 W^l) . \end{aligned}$$

Again from first principles, using (2.7) and (2.8), we get

$$\begin{aligned} & E\{(Z^T A Z)(Z^T B Z)(Z^T C Z) W^l\} \\ &\quad = d(A, B, C)\{E(Z_i^6 W^l) - 15E(Z_i^2 Z_j^4 W^l) + 30E(Z_i^2 Z_j^2 Z_k^2 W^l)\} \\ &\quad\quad + \{\operatorname{tr}(A) d(B, C) + \operatorname{tr}(B) d(A, C) + \operatorname{tr}(C) d(A, B) \\ &\quad\quad\quad + 4d(AB, C) + 4d(BC, A) + 4d(AC, B)\} \\ &\quad\quad\quad \cdot \{E(Z_i^2 Z_j^4 W^l) - 3E(Z_i^2 Z_j^2 Z_k^2 W^l)\} \\ &\quad\quad + \{\operatorname{tr}(A) \operatorname{tr}(B) \operatorname{tr}(C) + 2 \operatorname{tr}(A) \operatorname{tr}(BC) + 2 \operatorname{tr}(B) \operatorname{tr}(AC) \\ &\quad\quad\quad + 2 \operatorname{tr}(C) \operatorname{tr}(AB) + 8 \operatorname{tr}(ABC)\} E(Z_i^2 Z_j^2 Z_k^2 W^l) , \end{aligned}$$

which, by (2.13), (2.14), (2.9) and (2.6), simplifies to

$$(2.17)\quad E\{(Z^TAZ)(Z^TBZ)(Z^TCZ)W^l\}$$
$$= \{\mathrm{tr}\,(A)\,\mathrm{tr}\,(B)\,\mathrm{tr}\,(C) + 2\,\mathrm{tr}\,(A)\,\mathrm{tr}\,(BC) + 2\,\mathrm{tr}\,(B)\,\mathrm{tr}\,(AC) + 2\,\mathrm{tr}\,(C)\,\mathrm{tr}\,(AB) + 8\,\mathrm{tr}\,(ABC)\}E(Z_i^2Z_j^2Z_k^2W^l)$$
$$= \frac{1}{p(p+2)(p+4)}\{\mathrm{tr}\,(A)\,\mathrm{tr}\,(B)\,\mathrm{tr}\,(C) + 2\,\mathrm{tr}\,(A)\,\mathrm{tr}\,(BC) + 2\,\mathrm{tr}\,(B)\,\mathrm{tr}\,(AC) + 2\,\mathrm{tr}\,(C)\,\mathrm{tr}\,(AB) + 8\,\mathrm{tr}\,(ABC)\}E(\|Z\|^6W^l)\,.$$

In the refereeing process attention has been drawn to alternative derivations using cumulants, see e.g., McCullagh (1987), and zonal polynomials and invariant polynomials, see James (1964) and Davis (1979, 1981). Except possibly in the case of (2.17), both would seem to use heavier and less immediately accessible machinery than is really warranted. Moreover, they do not reveal identities of the type (2.13) and (2.14) which are useful in more applied contexts.

3. Entries of the information matrix

In determining the entries of the information matrix of X, an $EL_p^h(\mu, \Psi)$ random variable, we shall always use Z to mean the $EL_p^h(0, I)$ random variable defined by (2.2) and (2.3). We shall assume that all expectations encountered exist.

From (2.1), $\log f_h(x\,|\,\mu, \Psi)$ is given by

$$\log f_h = -\frac{1}{2}\log(\det \Psi) + \log h(u)\,,$$

where $u = (x-\mu)^T\Psi^{-1}(x-\mu)$.

On differentiating with respect to μ_i

$$(3.1)\quad \frac{\partial \log f_h}{\partial \mu_i} = -2\,\frac{d\log h(u)}{du}\{(x-\mu)^T\Psi^{-1}e_i\}$$
$$= -2\,\frac{d\log h(\|z\|^2)}{d(\|z\|^2)}\{z^TP^T\Psi^{-1}e_i\}$$
$$= -2wz^TP^T\Psi^{-1}e_i\,,$$

where e_i is the $p\times 1$ vector with 1 in the i-th entry and zeros elsewhere ($i = 1, 2,\ldots,p$). It follows at once, using (2.15), that, for $i, j = 1, 2,\ldots,p$,

$$(3.2)\qquad E\left(\frac{\partial \log f_h}{\partial \mu_i}\frac{\partial \log f_h}{\partial \mu_j}\right) = 4E\{(Z^TP^T\Psi^{-1}e_iZ^TP^T\Psi^{-1}e_j)W^2\}$$

$$= 4E\{(Z^TP^T\Psi^{-1}e_ie_j^T\Psi^{-1}PZ)W^2\}$$

$$= \frac{4}{p}\operatorname{tr}(P^T\Psi^{-1}e_ie_j^T\Psi^{-1}P)E(\|Z\|^2W^2)$$

$$= \frac{4}{p}E(\|Z\|^2W^2)e_i^T\Psi^{-1}e_j$$

$$= 4a_h\sigma^{ij},$$

where σ^{ij} denotes the (i,j)-th entry of Ψ^{-1} and

$$(3.3)\qquad a_h = \frac{1}{p}E(\|Z\|^2W^2).$$

To find $E\{(\partial \log f_h/\partial \mu_i)(\partial \log f_h/\partial \sigma_{kl})\}$ for $l \geq k$, i, k, $l = 1, 2, \ldots, p$, where σ_{kl} denotes the (k,l)-th entry of Ψ, we use the following results due to Dwyer (1967) on matrix derivatives for the symmetric matrix Ψ, viz.,

$$\frac{\partial}{\partial \Psi}\log(\det \Psi) = \Psi^{-1}$$

and

$$(3.4)\qquad \frac{\partial}{\partial \Psi}(x-\mu)^T\Psi^{-1}(x-\mu) = -\Psi^{-1}(x-\mu)(x-\mu)^T\Psi^{-1}.$$

Applying these, we get

$$(3.5)\qquad \frac{\partial \log f_h}{\partial \Psi} = -\frac{1}{2}\Psi^{-1} + \frac{d\log h(u)}{du}\{-\Psi^{-1}(x-\mu)(x-\mu)^T\Psi^{-1}\}$$

$$= -\frac{1}{2}\Psi^{-1} + \frac{d\log h(\|z\|^2)}{d(\|z\|^2)}(-\Psi^{-1}Pzz^TP^T\Psi^{-1})$$

$$= -\frac{1}{2}\Psi^{-1} - w\Psi^{-1}Pzz^TP^T\Psi^{-1},$$

which, combined with (3.1), gives

$$E\left(\frac{\partial \log f_h}{\partial \mu_i}\frac{\partial \log f_h}{\partial \Psi}\right) = E(WZ^TP^T\Psi^{-1}e_i\Psi^{-1}) + 2E(W^2Z^TP^T\Psi^{-1}e_i\Psi^{-1}PZZ^TP^T\Psi^{-1}) = 0\,, \tag{3.6}$$

since each term involves expectations of odd functions of Z_i.

Before finding the entries $E\{(\partial \log f_h/\partial\sigma_{ij})(\partial \log f_h/\partial\sigma_{kl})\}$ ($j \geq i$, $l \geq k$, i, j, k, $l = 1,\ldots,p$), we introduce the following notation,

$$I_{ij} = \begin{cases} I_{(i,i)} & i = j\,, \\ I_{(i,j)} + I_{(j,i)} & i \neq j\,, \end{cases}$$

where $I_{(i,j)}$ denotes the $p \times p$ matrix with (i,j)-th entry 1 and 0 elsewhere. It follows at once, from (3.5), that

$$\frac{\partial \log f_h}{\partial \sigma_{kl}} = -\frac{1}{2}\operatorname{tr}(\Psi^{-1}I_{kl}) - \operatorname{tr}(\Psi^{-1}Pzz^TP^T\Psi^{-1}I_{kl})w = -\frac{1}{2}\operatorname{tr}(\Psi^{-1}I_{kl}) - (z^TP^T\Psi^{-1}I_{kl}\Psi^{-1}Pz)w \tag{3.7}$$

and hence, from (2.12), (2.15) and (2.16), that

$$E\left(\frac{\partial \log f_h}{\partial \sigma_{kl}}\frac{\partial \log f_h}{\partial \sigma_{rs}}\right) = -\frac{1}{4}\operatorname{tr}(\Psi^{-1}I_{kl})\operatorname{tr}(\Psi^{-1}I_{rs}) + E(Z_i^2Z_j^2W^2)\{2\operatorname{tr}(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs}) + \operatorname{tr}(\Psi^{-1}I_{kl})\operatorname{tr}(\Psi^{-1}I_{rs})\}\,, \tag{3.8}$$

which, using (2.8) and (2.5), can be written as

$$E\left(\frac{\partial \log f_h}{\partial \sigma_{kl}}\frac{\partial \log f_h}{\partial \sigma_{rs}}\right) = 2b_h\operatorname{tr}(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs}) + \frac{1}{4}(4b_h - 1)\operatorname{tr}(\Psi^{-1}I_{kl})\operatorname{tr}(\Psi^{-1}I_{rs})\,, \tag{3.9}$$

where

$$b_h = \frac{1}{p(p+2)}E(\|Z\|^4W^2)\,. \tag{3.10}$$

It is a simple, if tedious, matter to now derive the expression in (3.9) in terms of the entries σ^{ij} of Ψ^{-1}. Care must be taken with combinations of different values of k, l, r and s with $l \geq k$, $s \geq r$.

To summarize, the elements of the information matrix are given by

$$E\left(\frac{\partial \log f_h}{\partial \mu_i}\frac{\partial \log f_h}{\partial \mu_j}\right) = 4a_h e_i^T \Psi^{-1} e_j = 4a_h \sigma^{ij},$$

$$(3.11) \quad E\left(\frac{\partial \log f_h}{\partial \mu_i}\frac{\partial \log f_h}{\partial \sigma_{kl}}\right) = 0,$$

$$E\left(\frac{\partial \log f_h}{\partial \sigma_{kl}}\frac{\partial \log f_h}{\partial \sigma_{rs}}\right) = 2b_h \operatorname{tr}(\Psi^{-1} I_{kl} \Psi^{-1} I_{rs}) + \frac{1}{4}(4b_h - 1)\operatorname{tr}(\Psi^{-1} I_{kl}) \operatorname{tr}(\Psi^{-1} I_{rs})$$

$$(l \geq k, s \geq r).$$

In the particular case of multivariate normality, it is easy to show, from (3.3), (3.10) and (2.10) that

$$a_h = b_h = \frac{1}{4}.$$

The elements of the information matrix then reduce to those given by Hayakawa (1980) and derived by Skovgaard (1984).

As a very simple example of one type of use of these results, we consider members of the elliptic class for which $E(X)$ and $\operatorname{cov}(X)$ exist and are given by

$$E(X) = \mu \quad \text{and} \quad \operatorname{cov}(X) = k_h \Psi.$$

In this case an obvious simple estimator of μ would be the sample mean vector, $\bar{X}$. Assuming a random sample of size n and the appropriate Central Limit Theorem, $\bar{X}$ is asymptotically $N_p(\mu, k_h\Psi/n)$. From the inverse of the information matrix in (3.11), the asymptotic efficiency of $\bar{X}$ is given by

$$(3.12) \qquad \left\{\frac{1}{4a_h n}(\det \Psi)^{1/p}\right\} \Big/ \left\{\frac{k_h(\det \Psi)^{1/p}}{n}\right\} = \frac{1}{4a_h k_h},$$

where a_h is defined in (3.3).

To illustrate this result we consider a Student's t-distribution on k ($k > 2$) degrees of freedom. It is well-known that

$$k_h = \frac{k}{k-2},$$

and it is easily verified from (2.11) that

$$a_h = \frac{(k+p)}{4(k+p+2)}.$$

Hence, we find that the asymptotic efficiency of $\bar{X}$ is given by $(1 - 2/k) \cdot (1 + 2/(k+p))$.

Letting $k \to \infty$, irrespective of the dimension p, we get 100% asymptotic efficiency corresponding to the fact that, in the normal case, $\bar{X}$ is the maximum likelihood estimator of μ. For finite $k > 2$, increasing the dimension leads to a fall in the asymptotic efficiency to the lower bound $(1 - 2/k) = 1/k_h$. This is an increasing function of k and is as high as 80% for $k = 10$.

More generally, however, (3.12) provides an interesting interpretation of the factor by which the usual multivariate normal Mahalanobis distance must be multiplied to give the distance between members of a general elliptic family differing only in location. Following Mitchell and Krzanowski (1985), it is easy to see that the factor is the inverse of the asymptotic efficiency of the sample mean, $\bar{X}$, as an estimator of μ.

4. The skewness tensor and α-connections

For simplicity of notation we denote the $r = p(p+3)/2$ distinct parameters by $\theta^T = (\theta_1, \theta_2, \ldots, \theta_r)$, the first p being $\mu_1, \mu_2, \ldots, \mu_p$ and the remainder being σ_{kl} for $l \geq k$. In the differential geometry approach to statistical theory, Amari (1985) uses the entries, $g_{ij}(\theta)$ $(i, j = 1, 2, \ldots, r)$, of the information matrix as the basic covariant symmetric tensor, known as the Riemannian metric tensor. A one-parameter family of affine connections is then defined by the α-connections with coefficients

$$^{\alpha}\Gamma_{ijk} = [ij, k] - \frac{\alpha}{2} T_{ijk},$$

where

$$[ij, k] = \frac{1}{2}\left(\frac{\partial g_{ik}(\theta)}{\partial \theta_j} + \frac{\partial g_{jk}(\theta)}{\partial \theta_i} - \frac{\partial g_{ij}(\theta)}{\partial \theta_k} \right),$$

the Christoffel symbols of the first kind and

$$T_{ijk} = E\left(\frac{\partial \log f_h}{\partial \theta_i} \frac{\partial \log f_h}{\partial \theta_j} \frac{\partial \log f_h}{\partial \theta_k} \right),$$

the skewness tensor ($i, j, k = 1, 2, \ldots, r$).

It follows at once, from (3.1), (3.2), (3.4), (3.5) and (3.6), that

$$T_{ijk} = [ij, k] = {}^{\alpha}T_{ijk} = 0$$

for all $i, j, k \le p$ and for only one of $i, j, k \le p$.

We now consider the case where only one of i, j, k is greater than p. For convenience, we change the notation slightly and in an obvious way. From (3.1) and (3.7), the entries of the skewness tensor corresponding to all permutations of μ_i, μ_j and σ_{kl} ($l \ge k$) are given by

$$\begin{aligned} T_{\mu_i \mu_j \sigma_{kl}} &= E\left(\frac{\partial \log f_h}{\partial \mu_i} \frac{\partial \log f_h}{\partial \mu_j} \frac{\partial \log f_h}{\partial \sigma_{kl}} \right) \\ &= -2 \operatorname{tr}(\Psi^{-1} I_{kl}) E\{(Z^T P^T \Psi^{-1} e_i e_j^T \Psi^{-1} P Z) W^2\} \\ &\quad - 4E\{(Z^T P^T \Psi^{-1} e_i e_j^T \Psi^{-1} P Z)(Z^T P^T \Psi^{-1} I_{kl} \Psi^{-1} P Z) W^3\}, \end{aligned}$$

which, from (2.15) and (2.16), gives

$$\begin{aligned} T_{\mu_i \mu_j \sigma_{kl}} &= -\frac{2}{p} \operatorname{tr}(\Psi^{-1} I_{kl}) \operatorname{tr}(\Psi^{-1} e_i e_j^T) E(\|Z\|^2 W^2) \\ &\quad - \frac{4}{p(p+2)} \{2 \operatorname{tr}(\Psi^{-1} e_i e_j^T \Psi^{-1} I_{kl}) + \operatorname{tr}(\Psi^{-1} e_i e_j^T) \operatorname{tr}(\Psi^{-1} I_{kl})\} \\ &\quad \cdot E(\|Z\|^4 W^3) \\ &= -2(a_h + 2c_h)\sigma^{ij} \operatorname{tr}(\Psi^{-1} I_{kl}) - 8c_h(e_i^T \Psi^{-1} I_{kl} \Psi^{-1} e_j), \end{aligned} \tag{4.1}$$

where a_h is defined in (3.3) and

$$c_h = \frac{1}{p(p+2)} E(\|Z\|^4 W^3). \tag{4.2}$$

Moreover, from (3.2) and (3.4),

$$\begin{aligned} [\mu_i \mu_j, \sigma_{kl}] &= -[\mu_i \sigma_{kl}, \mu_j] = -[\sigma_{kl} \mu_i, \mu_j] \\ &= -2a_h \frac{\partial}{\partial \sigma_{kl}} e_i^T \Psi^{-1} e_j \\ &= -2a_h \frac{\partial}{\partial \sigma_{kl}} \operatorname{tr}(\Psi^{-1} e_i e_j^T) \end{aligned} \tag{4.3}$$

$$= 2a_h \operatorname{tr}(\Psi^{-1}e_ie_j^T\Psi^{-1}I_{kl})$$
$$= 2a_he_i^T\Psi^{-1}I_{kl}\Psi^{-1}e_j\,.$$

Combining (4.1) and (4.3) we get

$$(4.4)\qquad {}^{\alpha}\Gamma_{\mu_i\mu_j\sigma_{kl}} = 2(a_h + 2\alpha c_h)(e_i^T\Psi^{-1}I_{kl}\Psi^{-1}e_j) + \alpha(a_h + 2c_h)\sigma^{ij}\operatorname{tr}(\Psi^{-1}I_{kl})$$

and

$$(4.5)\qquad {}^{\alpha}\Gamma_{\mu_i\sigma_{kl}\mu_j} = {}^{\alpha}\Gamma_{\sigma_{kl}\mu_i\mu_j} = -2(a_h - 2\alpha c_h)(e_i^T\Psi^{-1}I_{kl}\Psi^{-1}e_j) + \alpha(a_h + 2c_h)\sigma^{ij}\operatorname{tr}(\Psi^{-1}I_{kl})\,.$$

All that remains is to consider the final case, viz., $T_{\sigma_{ij}\sigma_{kl}\sigma_{rs}}$ and $[\sigma_{ij}\sigma_{kl}, \sigma_{rs}]$ for $j \geq i$, $l \geq k$, $s \geq r$. In finding $T_{\sigma_{ij}\sigma_{kl}\sigma_{rs}}$ it is clear, from (3.7), that we are required to evaluate expressions of the following type. From (2.15) and (2.12)

$$(4.6)\qquad E\{(Z^TP^T\Psi^{-1}I_{kl}\Psi^{-1}PZ)W\} = -\frac{1}{2}\operatorname{tr}(\Psi^{-1}I_{kl})$$

and, from (2.16) and (3.10),

$$(4.7)\qquad E\{(Z^TP^T\Psi^{-1}I_{ij}\Psi^{-1}PZ)(Z^TP^T\Psi^{-1}I_{kl}\Psi^{-1}PZ)W^2\} = b_h\{2\operatorname{tr}(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}) + \operatorname{tr}(\Psi^{-1}I_{ij})\operatorname{tr}(\Psi^{-1}I_{kl})\}$$

and, finally, from (2.17),

$$(4.8)\qquad \begin{aligned} &E\{(Z^TP^T\Psi^{-1}I_{ij}\Psi^{-1}PZ)(Z^TP^T\Psi^{-1}I_{kl}\Psi^{-1}PZ)(Z^TP^T\Psi^{-1}I_{rs}\Psi^{-1}PZ)W^3\} \\ &= d_h\{\operatorname{tr}(\Psi^{-1}I_{ij})\operatorname{tr}(\Psi^{-1}I_{kl})\operatorname{tr}(\Psi^{-1}I_{rs}) + 2\operatorname{tr}(\Psi^{-1}I_{ij})\operatorname{tr}(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs}) \\ &\quad + 2\operatorname{tr}(\Psi^{-1}I_{kl})\operatorname{tr}(\Psi^{-1}I_{ij}\Psi^{-1}I_{rs}) + 2\operatorname{tr}(\Psi^{-1}I_{rs})\operatorname{tr}(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}) \\ &\quad + 8\operatorname{tr}(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})\}\,, \end{aligned}$$

where

$$(4.9)\qquad d_h = \frac{1}{p(p+2)(p+4)}E(\|Z\|^6W^3)\,.$$

Combining these results we get

$$(4.10)\quad T_{\sigma_{ij}\sigma_{kl}\sigma_{rs}} = \frac{1}{4}(1 - 4d_h - 6b_h)\,\mathrm{tr}\,(\Psi^{-1}I_{ij})\,\mathrm{tr}\,(\Psi^{-1}I_{kl})\,\mathrm{tr}\,(\Psi^{-1}I_{rs})$$
$$- (2d_h + b_h)\{\mathrm{tr}\,(\Psi^{-1}I_{ij})\,\mathrm{tr}\,(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})$$
$$+ \mathrm{tr}\,(\Psi^{-1}I_{kl})\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{rs})$$
$$+ \mathrm{tr}\,(\Psi^{-1}I_{rs})\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl})\}$$
$$- 8d_h\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})\,.$$

To find $[\sigma_{ij}\sigma_{kl}, \sigma_{rs}]$ we note, from (3.9), that we require expressions for terms of the type

$$\frac{\partial}{\partial\sigma_{ij}}\,\mathrm{tr}\,(\Psi^{-1}I_{kl}) \quad \text{and} \quad \frac{\partial}{\partial\sigma_{ij}}\,\mathrm{tr}\,(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})\,.$$

These are given by

$$\frac{\partial}{\partial\sigma_{ij}}\,\mathrm{tr}\,(\Psi^{-1}I_{kl}) = -\,\mathrm{tr}\,(\Psi^{-1}I_{kl}\Psi^{-1}I_{ij})$$

and

$$\frac{\partial}{\partial\sigma_{ij}}\,\mathrm{tr}\,(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs}) = -\,2\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})\,.$$

The expression for $[\sigma_{ij}\sigma_{kl}, \sigma_{rs}]$ is now immediate and given by

$$(4.11)\quad [\sigma_{ij}\sigma_{kl}, \sigma_{rs}] = -\,2b_h\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})$$
$$- \frac{1}{4}(4b_h - 1)\,\mathrm{tr}\,(\Psi^{-1}I_{rs})\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl})\,.$$

Combining (4.10) and (4.11), we get the coefficient of the α-connection

$$(4.12)\quad {}^{\alpha}\Gamma_{\sigma_{ij}\sigma_{kl}\sigma_{rs}} = \frac{\alpha}{8}(4d_h + 6b_h - 1)\,\mathrm{tr}\,(\Psi^{-1}I_{ij})\,\mathrm{tr}\,(\Psi^{-1}I_{kl})\,\mathrm{tr}\,(\Psi^{-1}I_{rs})$$
$$+ \frac{\alpha}{2}(b_h + 2d_h)\{\mathrm{tr}\,(\Psi^{-1}I_{ij})\,\mathrm{tr}\,(\Psi^{-1}I_{kl}\Psi^{-1}I_{rs})$$
$$+ \mathrm{tr}\,(\Psi^{-1}I_{kl})\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{rs})\}$$
$$+ \frac{1}{4}\{\alpha(2b_h + 4d_h) + (1 - 4b_h)\}\{\mathrm{tr}\,(\Psi^{-1}I_{rs})\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}I_{kl})\}$$
$$+ 2(2\alpha d_h - b_h)\,\mathrm{tr}\,(\Psi^{-1}I_{ij}\Psi^{-1}\mathrm{I}_{kl}\Psi^{-1}I_{rs})\,.$$

If we restrict ourselves to members of the $EL_p^h(\mu, \Psi)$ class with the same, known location μ, the only relevant α-connections are those in (4.12). It is easy to see, from (2.10), that, in the normal case, when $\alpha = -1$, these are identically zero and the manifold consequently flat. This is, of course, a well-known result due to Amari (1985) for the exponential family and the expectation parameters.

Finally we note that the affine connections can also be described by Christoffel symbols of the second kind. Skovgaard (1984) has found expressions for these in the normal case when the Riemannian ($\alpha = 0$) connection is considered.

It would have been possible to work in terms of the parameters μ and the distinct entries σ^{kl} ($l \geq k$) of Ψ^{-1}. The results are analogous to those for μ and Ψ and are derived similarly. In particular, in the normal case with μ known, the 1-connections are identically zero and the manifold consequently flat. This is again a well-known result due to Amari (1985) for the exponential family and the natural parameters. The dual coordinate systems ($\sigma_{ij}, j \geq i$) and (σ^{kl}, $l \geq k$) lead to orthogonal parameters. In particular, in the bivariate case, when $X = (X_1, X_2)$, it follows that the variance of the marginal distribution of X_1 and the variance of the conditional distribution of X_2 given X_1 are orthogonal, a result shown by Mitchell (1962) in a direct approach to finding orthogonal parameters.

Acknowledgement

The author wishes to express her gratitude to Dr. L. V. White for a simple expression for the expectation of the product of three quadratic forms.

REFERENCES

Amari, S-I. (1985). *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, **28**, Springer-Verlag, Berlin.

Anderson, T. W. and Stephens, M. A. (1972). Tests for randomness of directions against equatorial and bimodal alternatives, *Biometrika*, **59**, 613–621.

Chmielewski, M. A. (1981). Elliptically symmetric distributions: A review and bibliography, *Internat. Statist. Rev.*, **49**, 67–74.

Davis, A. W. (1979). Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory, *Ann. Inst. Statist. Math.*, **31**, 465–485.

Davis, A. W. (1981). On the construction of a class of invariant polynomials in several matrices, extending the zonal polynomials, *Ann. Inst. Statist. Math.*, **33**, 297–313.

Dwyer, P. S. (1967). Some applications of matrix derivatives in multivariate analysis, *J. Amer. Statist. Assoc.*, **62**, 607–625.

Hayakawa, T. (1980). On the distribution of the likelihood ratio criterion for a covariance matrix, *Recent Developments in Statistical Inference and Data Analysis*, 79–84, North-Holland.

Hsu, H. (1985). Invariant tests and likelihood ratio tests for multivariate elliptically contoured distributions, Tech. Report No. 14, Department of Statistics, Stanford University.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples, *Ann. Math. Statist.*, **35**, 475–501.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhyā Ser. A*, **32**, 419–430.

Lord, R. D. (1954). The use of the Hankel transform in statistics, I. General theory and example, *Biometrika*, **41**, 44–45.

Maronna, R. A. (1976). Robust *M*-estimators of multivariate location and scatter, *Ann. Statist.*, **4**, 51–67.

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

Mitchell, A. F. S. (1962). Sufficient statistics and orthogonal parameters, *Proc. Camb. Philos. Soc.*, **58**, 326–337.

Mitchell, A. F. S. (1987). Discussion of paper by D. R. Cox and N. Reid, *J. Roy. Statist. Soc. Ser. B*, **49**, 26.

Mitchell, A. F. S. (1988). Statistical manifolds of univariate elliptic distributions, *Internat. Statist. Rev.*, **56**, 1–16.

Mitchell, A. F. S. and Krzanowski, W. (1985). The Mahalanobis distance and elliptic distributions, *Biometrika*, **72**, 464–467.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Skovgaard, L. T. (1984). A Riemannian geometry of the multivariate normal model, *Scand. J. Statist.*, **11**, 211–223.