Journal of Multivariate Analysis **74**, 267–281 (2000)
doi:10.1006/jmva.1999.1885, available online at http://www.idealibrary.com on IDEAL®

Multivariate Dispersion Models

Bent Jørgensen

Department of Statistics and Demography, University of Southern Denmark, Odense, Denmark

and

Steffen L. Lauritzen

Department of Mathematics, Aalborg University, Aalborg, Denmark

Received September 25, 1998

models preserve some of the main properties of the multivariate normal distribution, and include the elliptically contoured distributions and certain other known distributions as special cases. We give explicit methods for constructing multivariate proper dispersion models. This is exemplified by constructing multivariate gamma, Laplace, hyperbola, and von Mises distributions. © 2000 Academic Press

AMS subject classifications: 62H05, 62H11.

Key words and phrases: elliptically contoured distribution, multivariate gamma distribution, multivariate proper dispersion model, generalized yoke.

1. INTRODUCTION

The multivariate normal distribution occupies a central position in multivariate analysis of continuous data, and much multivariate distribution theory is directed towards creating alternatives to the multivariate normal for skewed or otherwise non-normal data, while preserving some of its main properties. Some early work on bivariate distributions was reviewed by Mardia (1970), and the multivariate case was reviewed by Johnson and Kotz (1972) and Jensen (1985). Other examples of non-normal multivariate distributions include the elliptically contoured distributions (Fang, 1997) and the multivariate hyperbolic distributions (Barndorff-Nielsen and Blæsild, 1987). An extensive introduction to the construction of multivariate models with various types of dependence structures was given by Joe (1997).

Our main concern here is with suitable distributions for generalized linear models with a multivariate response vector. A step in this direction

0047-259X/00 $35.00

was taken by Liang and Zeger (1986), who introduced the generalized estimating equation approach, suitable for multivariate or longitudinal data with exponential dispersion model marginals. In this approach the joint distribution is not specified, except for second moments, and Liang and Zeger de-emphasize the role of the joint distribution. In fact, it is not clear wether a full multivariate distribution with the given moments and marginals exists. The generalized estimating equation method was extended by Artes and Jørgensen (1998) to the case of dispersion model margins (Jørgensen, 1997b).

We introduce a new class of multivariate dispersion models suitable as error distributions for multivariate generalized linear models. Rather than requiring specific moments or marginals, our definition is geared towards good statistical properties. Our ultimate goal is to generalize the many classical multivariate techniques such as Hotelling's T^2, multivariate regression, MANOVA, and so on. In the present paper we introduce the multivariate dispersion models and consider some examples, whereas details regarding the statistical properties will be published elsewhere.

Multivariate dispersion models are defined in Section 2 and a geometric method for their construction is introduced in Section 3. In Section 4 we construct multivariate proper dispersion models from univariate ones; in particular, we introduce a multivariate gamma distribution. Statistical properties are discussed briefly in Section 5. We introduce further examples in Section 6 and end with some further discussion in Section 7.

2. DEFINITION OF MULTIVARIATE DISPERSION MODELS

Our starting point is a univariate *dispersion model* (Jørgensen, 1997b), which is a family of distributions with probability density functions on **R** (with respect to a given measure) of the form

$$f(y; \mu, \sigma^2) = a(y; \sigma^2) \exp\left\{ -\frac{1}{2\sigma^2} d(y; \mu) \right\}, \tag{1.1}$$

for suitable functions a and d, where the position parameter μ varies in Ω (an open interval), and the dispersion parameter σ^2 is positive. We denote (1.1) by $\mathrm{DM}(\mu, \sigma^2)$.

The function d is assumed to be a *unit deviance*, that is, it satisfies $d(y; \mu) > 0$ for $y \neq \mu$ and $d(\mu; \mu) = 0$ for $\mu \in \Omega$. In this sense d represents a kind of squared distance, and more specifically, d is a generalization of the Kullback–Leibler information divergence. Note that $-d$ is a *yoke* in the sense of Blæsild (1987). If the function $a(y; \sigma^2)$ factorizes as $(\sigma^2)\, b(y)$, say, we call (1.1) a *proper dispersion model*. The gamma, inverse Gaussian, and normal distributions are proper dispersion models, the normal distribution corresponding to the case where d is squared Euclidean distance.

The natural setting for many types of continuous multivariate data is a general smooth manifold, although there is little loss of generality in considering the sample space to be an open subset of $\mathbf{R}^p$, since most of our considerations are local. In the simplest case a *multivariate dispersion model* is defined by the following probability density function on $\mathbf{R}^p$,

$$
\begin{aligned}
f(y; \mu, \Sigma) &= a(y; \Sigma) \exp\{-\tfrac{1}{2} t^\top(y; \mu)\, \Sigma^{-1} t(y; \mu)\} \\
&= a(y; \Sigma) \exp[-\tfrac{1}{2} \operatorname{tr}\{\Sigma^{-1} t(y; \mu)\, t^\top(y; \mu)\}], \qquad (1.2)
\end{aligned}
$$

where $\mu \in \Omega$ (an open region in $\mathbf{R}^p$), Σ is a symmetric positive-definite $p \times p$ matrix and $t(y; \mu)$ is a suitably defined vector of residuals satisfying $t(\mu; \mu) = 0$ for $\mu \in \Omega$. The parameters μ, called the *position vector*, and Σ, called the *dispersion matrix*, may be interpreted as an analogues of respectively the mean vector and the variance-covariance matrix of the multivariate normal distribution. If $a(y; \Sigma)$ factorizes as $a(y; \Sigma)\, b(y)$, we call (1.2) a *multivariate proper dispersion model*. The multivariate normal distribution is obtained for $t(y; \mu) = y - \mu$ and $a(y; \Sigma) = a(\Sigma) = (2\pi)^{-p/2} \{\det(\Sigma)\}^{-1/2}$. As another example, we define the *multivariate Laplace distribution* by the following density with respect to Lebesgue measure on $\mathbf{R}^p$,

$$
f(y; \mu, \Sigma) = a(\Sigma) \exp\{-\tfrac{1}{2} t^\top(y - \mu)\, \Sigma^{-1} t(y - \mu)\},
$$

where $a(\Sigma)$ is a normalizing constant, t is defined by

$$
t(y - \mu) = (\pm\sqrt{|y_1 - m_1|}, \ldots, \pm\sqrt{|y_p - \mu_p|})^\top, \qquad (1.3)
$$

and $\pm = \operatorname{sgn}(y_j - \mu_j)$ denotes the sign of $y_j - \mu_j$ for each j. Note that when the dispersion matrix Σ is diagonal, the components of Y follow independent univariate Laplace distributions. The multivariate Laplace distribution may be useful for robust analysis of multivariate data. Many other similar examples may be constructed by using functions of the form $t(y; \mu) = t(y - \mu)$.

To generalize (1.2) further, we proceed as follows. Let $\mathscr{Y}$ denote a smooth C^∞ manifold of dimension p; cf. doCarmo (1979). Consider the probability density function

$$
f(y; \mu, \delta, \Lambda) = a(y; \delta, \Lambda) \exp[-\tfrac{1}{2} \delta c\{\langle \Lambda, T(y; \mu) \rangle\}] \qquad (1.4)
$$

for $y, \mu \in \mathscr{Y}$, $\delta \in \mathbf{R}_+$, where a and c are suitable functions, with c increasing and $c(0) = 0$, $c'(0) > 0$. We assume that $T\colon \mathscr{Y} \times \mathscr{Y} \mapsto V$ is a (generalized positive) *yoke*, in the sense that $(V, <, +)$ is an ordered real vector space,

$\langle \cdot, \cdot \rangle$ is a suitable inner product on V, $\Lambda \in V^*$, the dual of V, and T satisfies

$$T(y; y) = 0, \qquad T(y; \mu) \geqslant 0 \tag{1.5}$$

and $T(y; \mu) \neq 0$ for $\mu \neq y$. The parameter Λ is the analogue of the precision (inverse variance) of the multivariate normal distribution. Here and in the following, we use Λ rather than Σ^{-1} in the notation, and a is now a function of Λ rather than Σ and so on. We still refer to (1.4) as a multivariate dispersion model, and the model is called proper if a factorizes as $a(y; \delta; \Lambda) = a(\delta, \Lambda)\, b(y)$ for suitable a and b.

Of special interest is the case where

$$\langle \Lambda, T(y; \mu) \rangle = \operatorname{tr}\{\Lambda T(y; \mu)\}, \tag{1.6}$$

and $V = S_p$, the space of $p \times p$ symmetric matrices, with the Löwner ordering $A \geqslant B$ if and only if $A - B$ is non-negative definite. In particular, we may consider the case

$$T(y; \mu) = t(y; \mu)\, t^{\top}(y; \mu), \tag{1.7}$$

where $t\colon \mathcal{Y} \times \mathcal{Y} \mapsto \mathbf{R}^p$ is a suitable vector function. When c is the identity function, the choices (1.6) and (1.7) bring us back to (1.2). In particular, the choice $t(y; \mu) = y - \mu$ gives the multivariate normal yoke in (1.7), whereas for c not the identity this yoke gives the class of elliptically contoured distributions (Fang, 1997).

2.1. *Multivariate t Distribution*

To exemplify the definition, we consider the multivariate t distribution introduced by Cornish (1954). It is defined as the distribution of a random p-dimensional vector

$$T = \frac{\sqrt{\alpha}}{S} X,$$

where S is chi-squared with degrees of freedom $\alpha > 0$ and X is distributed independently thereof as $\mathcal{N}_p(0, \Lambda^{-1})$, Λ being a symmetric positive-definite $p \times p$ matrix. The density is equal to

$$f_p(t; \Lambda, \alpha) = \frac{\Gamma\{(\alpha + p)/2\}}{(\pi\alpha)^{p/2}\, \Gamma(\alpha/2)} \{\det(\Lambda)\}^{1/2} (1 + \alpha^{-1} t^{\top} \Lambda t)^{-(\alpha + p)/2}.$$

If we let $\delta = \alpha + p$, and transform to $Y = \mu + T/\sqrt{\alpha} = \mu + X/S$, we obtain a multivariate proper dispersion model of the form (1.4),

$$f(y;\mu,\delta,\Lambda)=a(y;\delta,\Lambda)\exp\left[-\frac{\delta}{2}c\{\langle\Lambda,T(y;\mu)\rangle\}\right]$$
$$=\frac{\Gamma(\delta/2)\{\det(\Lambda)\}^{1/2}}{(\pi)^{p/2}\,\Gamma\{(\delta-p)/2\}}$$
$$\times\exp\left(-\frac{\delta}{2}\log[1+\mathrm{tr}\{\Lambda(y-\mu)(y-\mu)^\top\}]\right),$$

where we use the normal yoke

$$T(y;\mu)=(y-\mu)(y-\mu)^\top,$$

and

$$c(x)=\log(1+x).$$

Being a scale mixture of the multivariate normal distribution, the construction of this distribution is to some extent analogous to the t-construction discussed in Jørgensen (1997a), but we abstain from further details here.

3. A GEOMETRIC CONSTRUCTION

A particularly appealing class of multivariate proper dispersion models is obtained by a geometric argument, corresponding to the p^*-construction in Jørgensen (1997a). We consider a yoke of the type (1.6), and take c to be the identity and $\delta=1$. Thus here and in the following, the statement $T\geqslant 0$ means that T is non-negative definite and $\mu\neq y$ then implies that at least one eigenvalue is positive. For the construction to work we also assume that the yoke is *locally quadratic* in the sense that it is infinitely often differentiable and that the matrix $G_\Lambda(y)=\{g_\Lambda^{jk}(y)\}$ with entries

$$g_\Lambda^{jk}(y)=\frac{1}{2}\frac{\partial^2}{\partial\mu_j\,\partial\mu_k}\mathrm{tr}\{\Lambda T(y;\mu)\}\bigg|_{\mu=y}$$
$$=\frac{1}{2}\sum_{uv}\lambda^{uv}H_{uv}^{jk}(y)$$

is positive definite if Λ is, where

$$H_{uv}^{jk}(y)=\frac{\partial^2 T_{uv}(y;\mu)}{\partial\mu_j\,\partial\mu_k}\bigg|_{\mu=y}.$$

That it is non-negative definite is a direct consequence of (1.5). Note that the Laplace yoke (1.3) is not locally quadratic because it is not differentiable at $y = \mu$.

When the symmetric tensor $G_\Lambda(y)$ is positive definite, it determines a Riemannian metric on $\mathcal{Y}$ which is the observed information metric for μ when Λ is considered fixed (Amari *et al.*, 1987). This in turn determines a uniform measure v_Λ on $\mathcal{Y}$. When $\mathcal{Y} = \mathbf{R}^n$, this measure is given as

$$v_\Lambda(dy) = \{\det G_\Lambda(y)\}^{1/2}\, dy.$$

We say that the yoke is *proper* if all these measures are proportional for different Λ, in which case we let $v = v_I$. For a singly connected manifold, the yoke is proper if the determinant factorizes as

$$\det G_\Lambda(y) = h_1(\Lambda)\, h_2(y)$$

for some functions h_1 and h_2. This is clearly equivalent to the logarithmic derivative being independent of Λ for all y.

Since

$$\frac{\partial}{\partial y_j} \log \det G_\Lambda(y) = \operatorname{tr}\{G_\Lambda(y)^{-1}\, G^j_\Lambda(y)\},$$

where

$$G^j_\Lambda(y) = \frac{\partial}{\partial y_j} G_\Lambda(y),$$

it can readily be checked, in any concrete case, whether a yoke is proper. For example, differentiating once more with respect to Λ and exploiting

$$\begin{aligned} \frac{\partial}{\partial \lambda_{uv}} G_\Lambda(y)^{-1} &= -G_\Lambda(y)^{-1} \left\{\frac{\partial}{\partial \lambda_{uv}} G_\Lambda(y)\right\} G_\Lambda(y)^{-1} \\ &= -G_\Lambda(y)^{-1}\, H_{uv}(y)\, G_\Lambda(y)^{-1}, \end{aligned}$$

we obtain that the necessary and sufficient condition for a yoke on a singly connected manifold to be proper is that we for all j, u, v have

$$\operatorname{tr}\{G_\Lambda(y)^{-1}\, G^j_\Lambda(y)\, G_\Lambda(y)^{-1}\, H_{uv}(y)\} = \operatorname{tr}\left\{G_\Lambda(y)^{-1} \frac{\partial}{\partial \lambda_{uv}} G^j_\Lambda(y)\right\}. \tag{1.8}$$

A yoke is always proper when $\mathcal{Y}$ is one-dimensional. The multivariate normal yoke is proper because $G_\Lambda(y) = \Lambda$ is constant in y and therefore both sides of (1.8) are zero.

For a proper yoke with v the common uniform measure, following Jørgensen (1997a) we next consider the normalizing constant

$$a(\mu, \Lambda)^{-1} = \int_{\mathscr{Y}} \exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda T(y; \mu)\}]\, v(dy).$$

If we now have

$$a(\mu, \Lambda) \equiv a(\Lambda), \tag{1.9}$$

a multivariate proper dispersion model is given by the family of densities

$$f(y; \mu, \Lambda) = a(\Lambda) \exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda T(y; \mu)\}] \tag{1.10}$$

with respect to the measure v.

The reciprocal of the normalizer $a(\mu, \Lambda)$ is equal to the Laplace transform of the measure v when transformed with $T(y; \mu)$. Thus the condition (1.9) expresses that $T(y; \mu)$ is pivotal with respect to the geometric measure v. This will be a valuable guide for the subsequent constructions. We are looking for proper yokes that are pivots for their induced measures. As an aside, let us note that the p^*-formula of Barndorff-Nielsen (1983) for the distribution of the maximum likelihood estimator given an ancillary statistic is exact in proper dispersion models constructed as above when Λ is held fixed. This follows precisely as in Jørgensen (1997a), and we omit the details.

4. FROM UNIVARIATE TO MULTIVARIATE

There is a standard construction that takes any univariate proper dispersion model and gives a multivariate proper dispersion model as a result. Consider a univariate model

$$f(y; \mu, \lambda) = a_1(\lambda) \exp\left\{-\frac{\lambda}{2} d(y; \mu)\right\},$$

where the unit deviance $d(y; \mu)$ is a univariate proper yoke and the density is with respect to its geometric measure v on $\mathbf{R}$. The notation a_1 is to distinguish the normalizer from those in higher dimensions introduced below. Note that the geometric measure is $v(dy) = V^{-1/2}(y)\, dy$, where

$$V(y) = \frac{2}{\left.\dfrac{\partial^2 d(y; \mu)}{\partial \mu^2}\right|_{\mu = y}}$$

is the *unit variance function* (Jørgensen, 1997b, p. 4). Since the normalizer $a_1(\lambda)$ does not depend on μ, $d(Y; \mu)$ is a pivot.

Next, let $r(y; \mu) = \pm\sqrt{d(y; \mu)}$ denote the deviance residual, where $\pm =$ sgn$(y - \mu)$. Assume that also $r(Y; \mu)$ is a pivot. Let

$$t(y; \mu) = \{r(y_1; \mu_1), ..., r(y_p; \mu_p)\}^\top$$

denote the vector of deviance residuals for a p-vector of data. Then the yoke

$$T(Y; \mu) = t(Y; \mu)\, t^\top(Y; \mu) = \begin{Bmatrix} r(Y_1; \mu_1) \\ \vdots \\ r(Y_p; \mu_p) \end{Bmatrix} \{r(Y_1; \mu_1), ..., r(Y_p; \mu_p)\} \tag{1.11}$$

is also a pivot with respect to the product measure

$$\begin{aligned} v(dy_1, ..., dy_p) &= v(dy_1) \otimes \cdots \otimes v(dy_p) \\ &= \prod_{j=1}^{p} \{V^{-1/2}(y_j)\, dy_j\}. \end{aligned} \tag{1.12}$$

A multivariate version of the univariate model therefore appears as

$$f(y; \mu, \Lambda\} = a_p(\Lambda) \exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda T(y; \mu)\}], \tag{1.13}$$

where the density is taken with respect to the product measure v. We denote this model $\mathrm{DM}_p(\mu; \Sigma)$, where $\Sigma = \Lambda^{-1}$.

The new model (1.13) is clearly a multivariate proper dispersion model and in the case when Λ is diagonal, it has marginals distributed independently as in the univariate model. The normalizing constant may be calculated as

$$a_p^{-1}(\Lambda) = \prod_{j=1}^{p} a_1^{-1}(\lambda_{jj})\, \mathbf{E}(\exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda_0 T(y; \mu)\}]),$$

where Λ_0 has zeros on the diagonal and off-diagonal elements equal to Λ, and the expectation is with respect to the distribution with independent marginals corresponding to the diagonal of Λ. This expectation may be calculated by means of Monte Carlo methods such as those described and exploited, for example, in Geyer and Thompson (1992).

To investigate whether the family so constructed is of the type described in the previous section, we must determine whether the product measure

(1.12) is proportional to the geometric measure induced by the second derivative of the yoke. We first differentiate to find the quantity $H_{uv}^{jk}(y)$ and get

$$\frac{\partial^2 T_{uv}(y;\mu)}{\partial\mu_j\,\partial\mu_k} = r_u^{jk} r_v + r_u^j r_v^k + r_u^k r_v^j + r_u r_v^{jk},$$

where the superscript j denotes differentiation with respect to μ_j and so on. Since $r(y;y)=0$ and the terms r_u^j vanish unless $j=u$ we obtain

$$H_{uv}^{jk}(y) = (\delta_u^j \delta_v^k + \delta_u^k \delta_v^j)\, V^{-1/2}(y_j)\, V^{-1/2}(y_k),$$

where δ_u^j is the indicator function of the set $\{j=u\}$. The metric is hence given by the matrix

$$\begin{aligned} g_\Lambda^{jk}(y) &= \tfrac{1}{2}\operatorname{tr}(\Lambda H_{uv}^{jk}(y)) \\ &= \tfrac{1}{2} V^{-1/2}(y_j)\, V^{-1/2}(y_k)(\lambda_{jk}+\lambda_{kj}) \\ &= V^{-1/2}(y_j)\, \lambda_{jk}\, V^{-1/2}(y_k). \end{aligned} \tag{1.14}$$

The determinant of this matrix is easily calculated, and we obtain the geometric measure

$$\{\det g_\Lambda(y)\}^{1/2}\, dy = \{\det(\Lambda)\}^{1/2} \left\{ \prod_{j=1}^{p} V^{-1/2}(y_j)\, dy_j \right\}. \tag{1.15}$$

This measure is proportional to the measure ν, with proportionality factor equal to $\{\det(\Lambda)\}^{1/2}$. The model is hence of the form considered in Section 3. The p^*-formula gives the following asymptotic value for a_p:

$$a_p(\Lambda) \sim (2\pi)^{-p/2}\, \{\det(\Lambda)\}^{1/2}. \tag{1.16}$$

This holds for $\det(\Lambda)$ large, the *small-dispersion* limit, where Y becomes close to μ.

An interesting feature of this standard construction is that it is possible to use it to construct graphical models, extending the notion of covariance selection models (Dempster, 1972) to more general graphical association models; cf. Whittaker (1990) and Lauritzen (1996). This is a consequence of the general form of the density. If $\lambda_{jk}=0$, the density factorizes into a product of a function that does not depend on y_j, and one that does not depend on y_k. Hence Y_j and Y_k are conditionally independent given the remaining variables. Also we want to be able to deal with conditional independence restrictions in non-normal cases, We abstain from pursuing this aspect further here.

The above construction generally does not lead to conditional and marginal distributions of the exact same form as the distribution itself, but this is approximately so in the small-dispersion limit.

We note that the construction is tied to a particular coordinate system because of the form of the product measure (1.12), although this may often be natural in practice. It is, in fact, possible to construct multivariate distributions of mixed types, starting from different distributions for each coordinate, which will be useful for setting up multivariate regression models for response vectors of mixed types.

As an example of a multivariate proper dispersion model which is constructed as above, we consider the gamma distribution with density

$$f(y; \mu, \lambda) = a_1(\lambda)\, y^{-1} \exp\left\{ -\frac{\lambda}{2} d(y; \mu) \right\}, \qquad y > 0, \tag{1.17}$$

where the unit deviance d is

$$d(y; \mu) = 2\left(\frac{y}{\mu} - \log \frac{y}{\mu} - 1 \right),$$

the unit variance function is $V(\mu) = \mu^2$, and

$$a_1(\lambda) = \frac{\lambda^\lambda e^{-\lambda}}{\Gamma(\lambda)}.$$

The deviance residual is

$$r(y; \mu) = \pm \sqrt{d(y; \mu)},$$

where the sign is $\pm = \operatorname{sgn}(y/\mu - 1)$. Clearly $r(Y; \mu)$ is a pivot. Hence the standard construction can be applied. The corresponding multivariate gamma distribution takes the form

$$f(y; \mu, \Lambda\} = a_p(\Lambda) \prod_{j=1}^{p} y_j^{-1} \exp[-\tfrac{1}{2} \operatorname{tr}\{ \Lambda t(y; \mu)\, t^\top(y; \mu)\}],$$

corresponding to the yoke obtained from r via (1.11), where the density is relative to the usual Lebesgue measure on $\mathbf{R}_+^p$. This multivariate gamma distribution seems to be new. In particular, its margins are not gamma, except when Λ is diagonal, in contrast to most other multivariate gamma distributions proposed in the literature; cf. Krishnaiah (1985). For $p=2$, the reciprocal of the normalizing constant can be expressed as

$$a_2^{-1}(\Lambda)=a_1^{-1}(\lambda_{11})\,a_1^{-1}(\lambda_{22})\,\mathbf{E}[\exp\{-\lambda_{12}r(Y_1;1)\,r(Y_2;1)\}],$$

where Y_1 and Y_2 are independent and $Y_j \sim \mathrm{Ga}(1, \lambda_{jj}^{-1})$ for $j=1, 2$. Consider the inverse Gaussian distribution with density

$$f(y;\lambda,\mu)=\left(\frac{\lambda}{2\pi}\right)^{1/2} y^{-3/2}\exp\left\{-\lambda\frac{(y-\mu)^2}{2y\mu^2}\right\}, \qquad y>0.$$

This is a univariate proper dispersion model with unit deviance

$$d(y;\mu)=\frac{(y-\mu)^2}{y\mu^2}.$$

The yoke $d(Y;\mu)$ is pivotal and the unit variance function is $V(y)=y^3$, so this univariate dispersion model is of the geometric type. We would like to use the signed square root

$$r(y;\mu)=\frac{y-\mu}{y^{1/2}\mu}$$

for the multivariate construction, but $r(Y;\mu)$ is not a pivot. Hence the standard construction does not always work, and unfortunately it is not clear how to proceed in such cases.

5. STATISTICAL PROPERTIES

We consider maximum likelihood estimation based on a random sample $Y_1, ..., Y_n$ from the multivariate dispersion model $\mathrm{DM}_p(\mu;\Sigma)$, the latter assumed to be generated from the univariate model $\mathrm{DM}(\mu,\sigma^2)$. Let $\hat{\mu}$ denote the maximum likelihood estimate based on the independence model (Σ diagonal), defined by the score equations

$$\sum_{i=1}^{n} d'(Y_{ij};\hat{\mu}_j)=0 \qquad \text{for} \quad j=1, ..., p, \tag{1.18}$$

where $d'(y;\mu)$ denotes the derivative of d with respect to μ. For exponential dispersion models such as the gamma distribution the estimator is $\hat{\mu}=\bar{Y}$, the sample mean vector.

For arbitrary Σ the score function for μ_j based on the full sample is

$$\begin{aligned}\frac{\partial l}{\partial \mu_j} &= -\frac{1}{2}\left\{\sum_{k=1}^{p} \lambda_{jk} \sum_{i=1}^{n} \frac{d'(Y_{ij};\mu_j)}{r(Y_{ij};\mu_j)} r(Y_{ik};\mu_k)\right\} \\ &= -\frac{1}{2}\left\{\lambda_{jj} \sum_{i=1}^{n} d'(Y_{ij};\mu_j) + \sum_{k\neq j} \lambda_{jk} \sum_{i=1}^{n} \frac{d'(Y_{ij};\mu_j)}{r(Y_{ij};\mu_j)} r(Y_{ik};\mu_k)\right\}.\end{aligned} \tag{1.19}$$

On plugging $\hat{\mu}$ from (1.18) into (1.19), the first term becomes zero, and in the small-dispersion limit, the second term is also zero. This follows from the small-dispersion approximation (Jørgensen, 1997b, p. 25)

$$d'(Y;\mu) \approx r(Y;\mu)\, V^{-1/2}(\mu).$$

Hence, $\hat{\mu}$ is the small-dispersion approximation to the maximum likelihood estimator for μ. Inserting the estimator $\hat{\mu}$ into the likelihood and maximizing with respect to Σ, the saddlepoint approximation (1.16) gives a maximization problem similar to that for the multivariate normal, and we obtain the small-dispersion approximation

$$\hat{\Sigma} \approx \frac{1}{n} \sum_{i=1}^{n} t(Y_i;\hat{\mu})\, t^{\top}(Y_i\ \hat{\mu}).$$

The above results have been developed by Rajeswaran (1998), who also showed that a generalization of Hotelling's T^2 test is given by

$$T^2 = n\bar{t}^{\top}\hat{\Sigma}^{-1}\bar{t},$$

where

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t(Y_i;\mu).$$

Rajeswaran (1998) also showed that the distribution of T^2 is approximately F, both for large samples and in the small-dispersion limit,

$$T^2 \approx \frac{p(n-1)}{n-p} F_{p,\,n-p},$$

showing an approximate version of the classical result for the multivariate normal.

These results indicate that asymptotic theory for multivariate dispersion models have many analogies with exact results from classical multivariate analysis. Note by comparison that certain results from the multivariate normal distribution hold exactly for elliptically contoured distributions (Anderson and Fang, 1987).

6. FURTHER EXAMPLES

6.1. *Multivariate Hyperbola and Simplex Distributions*

The univariate hyperbola distribution is a proper dispersion model given by the density

$$f(y; \mu, \lambda) = \frac{e^{-\lambda}}{2K_0(\lambda)}\, y^{-1} \exp\left\{-\frac{\lambda}{2}\frac{(y-\mu)^2}{y\mu}\right\}, \tag{1.20}$$

for y, $\mu > 0$, where K_0 is a Bessel function; see Jørgensen (1997b, p. 192). The corresponding deviance residual is

$$r(y; \mu) = \frac{y-\mu}{\sqrt{y\mu}}$$

and clearly $r(Y; \mu)$ is a pivot since μ is a scale parameter. The standard construction now gives the multivariate hyperbola distribution

$$f(y; \mu, \Lambda\} = a_p(\Lambda) \prod_{j=1}^{p} y_j^{-1} \exp[-\tfrac{1}{2}\operatorname{tr}\{\Lambda t(y; \mu)\, t^{\top}(y; \mu)\}], \tag{1.21}$$

with respect to Lebesgue measure on $\mathbf{R}_+^p$, where $t(y; \mu)$ is the vector of deviance residuals.

Following Jørgensen (1997b, p. 200), we find that the transformation $z = y/(1+y)$, along with the reparametrization $\mu_0 = \mu(1+\mu)$, transforms the hyperbola distribution (1.20) into the simplex distribution on the unit interval (0, 1) given by

$$f(z; \mu_0, \lambda) = \frac{e^{-\lambda}}{2K_0(\lambda)}\, z^{-1}(1-z)^{-1} \exp\left\{-\frac{\lambda}{2}\, d(z; \mu_0)\right\},$$

with unit deviance

$$d(y; \mu_0) = \frac{(z-\mu_0)^2}{z(1-z)\,\mu_0(1-\mu_0)}.$$

By applying the same transformation coordinatewise to (1.21), we obtain the following multivariate simplex distribution on $(0, 1)^p$:

$$f(z; \mu_0, \Lambda\} = a_p(\Lambda) \prod_{j=1}^{p} \{z_j^{-1}(1-z_j)^{-1}\} \exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda t(z; \mu_0)\, t^\top(z; \mu_0)\}];$$

here $t(z; \mu_0)$ is the vector of deviance residuals with components

$$\frac{z_j - \mu_{j0}}{\sqrt{z_j(1-z_j)\, \mu_{j0}(1-\mu_{j0})}}.$$

6.2. *Multivariate von Mises Distribution*

Consider the univariate von Mises distribution of the form

$$f(y; \mu, \lambda) = a_1(\lambda)\, e^{-\lambda\{1-\cos(y-\mu)\}}, \qquad y, \mu \in [0, 2).$$

We define the deviance residual by

$$r(y; \mu) = \pm\sqrt{2\{1-\cos(y-\mu)\}},$$

where the sign is $\pm = \operatorname{sgn} \sin(y-\mu)$. Since $r(Y; \mu)$ is a pivot, the standard construction leads to a multivariate von Mises distribution of the form

$$f(y; \mu, \Lambda\} = a_p(\Lambda) \exp[-\tfrac{1}{2} \operatorname{tr}\{\Lambda t(y; \mu)\, t^\top(y; \mu)\}],$$

on $[0, 2\pi)^p$, where $t(y; \mu)$ is the vector of deviance residuals.

7. DISCUSSION AND PERSPECTIVES

Many questions remain to be solved in order to construct a full-fledged theory of multivariate analysis based on dispersion models. An interesting perspective is to extend some of the many classical normal models for time series, longitudinal data, variance components, and other forms of correlated data to non-normal data by imposing the same kind of structures on the dispersion matrix Σ as on the variance–covariance matrix in the normal case. It would be interesting to compare such an approach with the many other types of modelling of correlated non-normal data currently in use, in particular for the extension of graphical models to multivariate proper dispersion models mentioned in Section 4.

ACKNOWLEDGMENTS

We are grateful to J. Rajeswaran, V. Seshadri, and P. Song for useful discussions in relation to this article. The first author is grateful to the Department of Theoretical Statistics, Aarhus University, and the Biometry Research Unit, Danish Institute of Agricultural Science, for support during the preparation of this paper.

REFERENCES

S.-I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, and C. R. Rao, "Differential Geometry in Statistical Inference," Institute of Mathematical Statistics, Hayward, CA, 1987.

T. W. Anderson and K.-T. Fang, Cochran's theorem for elliptically contoured distributions, *Sankhyā Ser. A* **49** (1987), 305–315.

R. Artes and B. Jørgensen, "Longitudinal Data Estimating Equations for Dispersion Models," Technical Report 9802, Department of Statistics, University of São Paulo, 1998.

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

O. E. Barndorff-Nielsen and P. Blæsild, Hyperbolic distributions, *in* "Encyclopedia of Statistical Sciences" (S. Kotz, C. B. Read, and D. L. Banks, Eds.), Vol. 3, pp. 700–707, Wiley, New York, 1987.

P. Blæsild, Yokes: Elemental properties with statistical applications, *in* "Geometrization of Statistical Theory" (C. T. J. Dodson, Ed.), pp. 193–196, University of Lancaster, Lancaster, UK, 1987.

E. A. Cornish, The multivariate t distribution, *Austral. J. Phys.* **7** (1954), 531–542.

A. P. Dempster, Covariance selection, *Biometrics* **28** (1972), 157–175.

M. P. doCarmo, "Differential Geometry of Curves and Surfaces," Prentice–Hall, Englewood Cliffs, 1976.

K.-T. Fang, Elliptically contoured distributions, *in* "Encyclopedia of Statistical Sciences" (S. Kotz, C. B. Read, and D. L. Banks, Eds.), Update Vol. 1, pp. 212–218, Wiley, New York, 1997.

C. G. Geyer and E. A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data (with discussion), *J. Roy. Statist. Soc. Ser. B* **54** (1992), 657–699.

D. R. Jensen, Multivariate distributions, *in* "Encyclopedia of Statistical Sciences" (S. Kotz, C. B. Read, and D. L. Banks, Eds.), Vol. 6, pp. 43–55, Wiley, New York, 1985.

H. Joe, "Multivariate Models and Dependence Concepts," Chapman & Hall, London, 1997.

N. L. Johnson and S. Kotz, "Distributions in Statistics: Continuous Multivariate Distributions," Wiley, New York, 1972.

B. Jørgensen, Proper dispersion models (with discussion), *Brazilian J. Probab. Statist.* **11** (1997a), 89–140.

B. Jørgensen, "The Theory of Dispersion Models," Chapman & Hall, London, 1997b.

P. R. Krishnaiah, Multivariate gamma distributions, *in* "Encyclopedia of Statistical Sciences" (S. Kotz, C. B. Read, and D. L. Banks, Eds.), Vol. 6, pp. 63–66, Wiley, New York, 1985.

S. L. Lauritzen, "Graphical Models," Oxford Univ. Press, Oxford, UK, 1996.

K.-Y. Liang and S. L. Zeger, Longitudinal analysis using generalized linear models, *Biometrika* **73** (1986), 13–22.

K. V. Mardia, "Families of Bivariate Distributions," Griffin, London, 1970.

J. Rajeswaran, M.Sc. thesis University of British Columbia.

J. Whittaker, "Graphical Models in Applied Multivariate Analysis," Wiley, Chichester, 1990.