

Construction of multivariate dispersion models

Bent Jørgensen

University of Southern Denmark

Abstract. We consider methods for constructing multivariate dispersion models, illustrated by examples. Such methods are motivated by the need for good regression modelling of multivariate nonnormal correlated data, which requires multivariate distributions with a flexible correlation structure. We first review existing methods for constructing multivariate proper dispersion models, involving quadratic forms of deviance residuals in the style of the multivariate normal density, which we illustrate by a multivariate hyperbola distribution. We develop an extended convolution method for constructing multivariate exponential dispersion models, designed to create a fully flexible covariance structure, which we illustrate by two bivariate gamma distributions. We develop a similar technique for constructing multivariate extreme dispersion models for extremes and survival data, and introduce new bivariate logistic and Gumbel distributions.

1 Introduction

The motivation for this paper comes from the need to develop flexible multivariate distribution families for stochastic modelling of nonnormal data. There is a large variety of such families available, including, for example, multivariate hyperbolic distributions (Barndorff-Nielsen and Blæsild, 1987), elliptically contoured distributions (Fang, 1997), skew-normal and skew-elliptical distributions (Arellano-Valle and Genton, 2010), and multivariate Birnbaum-Saunders distributions (Díaz-García and Domínguez-Molina, 2006; Kundu et al., 2010), to name but a few; see also Jensen (1985). It is not easy to know where to turn, as illustrated by the following comment by Letac (2007):

While the names of distributions in \mathbb{R} are generally unambiguous, at the contrary in the jungle of distributions in \mathbb{R}^k almost nothing is codified outside of the Wishart and Gaussian cases. The scenario is usually as follows: choose a one-dimensional *thingy* type (quite often an exponential dispersion model, namely a natural exponential family and all its real powers of convolution) such as the gamma or negative binomial; then any law in \mathbb{R}^k whose margins are of *thingy* type are said to be a multidimensional *thingy*. Although the study of all distributions with given marginals is rather in the nonparametric domain of study, actually each author who isolates some parametric family will declare that he or she has THE multidimensional *thingy* family.

Key words and phrases. Convolution method, multivariate exponential dispersion model, multivariate extreme dispersion model, multivariate proper dispersion model.

Received February 2011; accepted August 2011.

An alternative strategy is to create multivariate nonnormal models that are aimed at preserving the overall look and feel of a given univariate distribution family, while abandoning the requirement of prespecified marginal distributions. This approach was taken by Jørgensen and Lauritzen (2000), who introduced a class of multivariate proper dispersion models parametrized by a vector parameter $\boldsymbol{\mu}$ and symmetric positive-definite matrix parameter $\boldsymbol{\Sigma}$, with properties that are very similar to those of the multivariate normal distribution (Jørgensen and Rajeswaran, 2005). We review the main methods for constructing multivariate proper dispersion models in Sections 2–3, and introduce some new simulation techniques for such models; see Section 3.2.

When it comes to the other main type of dispersion model, namely, exponential dispersion models, the goal of finding a multivariate generalization with a fully flexible mean and covariance structure corresponding to a mean vector $\boldsymbol{\mu}$ and a symmetric positive-definite dispersion matrix $\boldsymbol{\Sigma}$ has proven elusive so far. For example, Furman and Landsman (2010) introduced a k -variate Tweedie exponential dispersion model with $2(k + 1)$ parameters, which in general falls short of the $k + k(k + 1)/2$ parameters required for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In Section 4 we present a new class of multivariate exponential dispersion models with the desired number of parameters, based on an extension of the convolution method for exponential dispersion models alluded to by Letac (2007) above. As an illustration we discuss two new types of bivariate gamma distributions (Section 4.5). We also introduce a new type of multivariate extreme dispersion model (cf. Jørgensen et al., 2010) for extremes and survival data, obtained by replacing convolution with the minimum operation (Section 5).

2 Multivariate dispersion models

We begin with a brief review of multivariate dispersion models. Our starting point is a univariate dispersion model $DM(\boldsymbol{\mu}, \sigma^2)$ (Sweeting, 1987; Jørgensen, 1987a, 1987b, 1997), which is a family of distributions with probability density functions on \mathbb{R} of the form

$$f(y; \boldsymbol{\mu}, \sigma^2) = a(y; \sigma^2) \exp\left[-\frac{1}{2\sigma^2}d(y; \boldsymbol{\mu})\right] \quad \text{for } y \in \mathbb{R} \quad (2.1)$$

for suitable functions a and d . The parameters $\boldsymbol{\mu} \in \Omega$ (an interval) and $\sigma^2 > 0$ are called the *position* and *dispersion* parameters, respectively. The function d is assumed to be a *unit deviance* satisfying $d(\boldsymbol{\mu}; \boldsymbol{\mu}) = 0$ for $\boldsymbol{\mu} \in \Omega$ and $d(y; \boldsymbol{\mu}) > 0$ for $y \neq \boldsymbol{\mu}$. If the function $a(y; \sigma^2)$ of (2.1) factorizes as $a(\sigma^2)b(y)$, say, we obtain the class of *proper dispersion models*, whose multivariate generalization we consider below.

The multivariate generalization is based on the so-called *deviance residual*, which is defined as $r(y; \boldsymbol{\mu}) = \pm\sqrt{d(y; \boldsymbol{\mu})}$, where $\pm = \text{sgn}(y - \boldsymbol{\mu})$. We assume

from now on that $d(\cdot; \mu)$ is continuous and strictly monotone on each side of μ , which in turn implies that $r(\cdot; \mu)$ is strictly increasing for each $\mu \in \Omega$. Let us consider the vector of deviance residuals,

$$\mathbf{r}(\mathbf{y}; \boldsymbol{\mu}) = [r(y_1; \mu_1), \dots, r(y_k; \mu_k)]^\top, \tag{2.2}$$

where y_j and μ_j denote the elements of the k -vectors \mathbf{y} and $\boldsymbol{\mu}$, respectively. Given a symmetric positive-definite $k \times k$ matrix $\boldsymbol{\Sigma}$, we define the *scaled deviance* as the following quadratic form in the vector of deviance residuals,

$$D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbf{r}^\top(\mathbf{y}; \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} \mathbf{r}(\mathbf{y}; \boldsymbol{\mu}) = \text{tr}[\boldsymbol{\Sigma}^{-1} \mathbf{r}(\mathbf{Y}; \boldsymbol{\mu}) \mathbf{r}^\top(\mathbf{Y}; \boldsymbol{\mu})]. \tag{2.3}$$

The ordinary scaled deviance is obtained as a special case for $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$,

$$D(\mathbf{y}; \boldsymbol{\mu}, \sigma^2 \mathbf{I}) = \frac{1}{\sigma^2} \sum_{j=1}^k d(y_j; \mu_j). \tag{2.4}$$

Following Jørgensen (1999) and Jørgensen and Lauritzen (2000), we define a *multivariate dispersion model* $\text{DM}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as follows:

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = a(\mathbf{y}; \boldsymbol{\Sigma}) \exp[-\frac{1}{2} D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})] \quad \text{for } \mathbf{y} \in \mathbb{R}^k, \tag{2.5}$$

where $a(\mathbf{y}; \boldsymbol{\Sigma})$ is a suitable function such that (2.5) is a probability density functions on \mathbb{R}^k . The *position vector* $\boldsymbol{\mu} \in \Omega^k$ and the *dispersion matrix* $\boldsymbol{\Sigma}$ may be interpreted as analogues of the mean vector and covariance matrix of the multivariate normal distribution, respectively. Since the scaled deviance D is elliptical in terms of the vector of residuals $\mathbf{r}(\mathbf{y}; \boldsymbol{\mu})$, there is a certain resemblance with the class of elliptically contoured distributions, and, in fact, Jørgensen and Lauritzen (2000) considered a generalization of (2.5) that includes elliptically contoured distributions as a special case. The multivariate normal distribution is the special case of (2.5) obtained for $\mathbf{r}(\mathbf{y}; \boldsymbol{\mu}) = \mathbf{y} - \boldsymbol{\mu}$ and $a(\mathbf{y}; \boldsymbol{\Sigma}) = (2\pi)^{-k/2} |\boldsymbol{\Sigma}|^{-1/2}$, where $|\cdot|$ denotes determinant.

Let us now assume that the unit deviance d is twice continuously differentiable with nonzero second derivative. We define the *variance function* $\sigma^2 V(\boldsymbol{\mu})$ on Ω by

$$\sigma^2 V(\boldsymbol{\mu}) = \left[\frac{1}{2\sigma^2} \ddot{d}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}; \boldsymbol{\mu}) \right]^{-1} \quad \text{for } \boldsymbol{\mu} \in \Omega, \tag{2.6}$$

which is finite and nonzero. We call $V(\boldsymbol{\mu})$ the *unit variance function*. Here we use the notation $\dot{d}_y(y; \boldsymbol{\mu})$, $\ddot{d}_{y\boldsymbol{\mu}}(y; \boldsymbol{\mu})$ and $\ddot{d}_{\boldsymbol{\mu}\boldsymbol{\mu}}(y; \boldsymbol{\mu})$, etc. for the first and second derivatives of the function d . The (matrix) variance function corresponding to the scaled deviance D is easily seen to be (Jørgensen and Lauritzen, 2000)

$$\mathbf{V}_{\boldsymbol{\Sigma}}(\boldsymbol{\mu}) = \left[\frac{1}{2} \ddot{D}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \right]^{-1} = \mathbf{V}^{1/2}(\boldsymbol{\mu}) \boldsymbol{\Sigma} \mathbf{V}^{1/2}(\boldsymbol{\mu}) = \boldsymbol{\Sigma} \odot \tilde{\mathbf{V}}(\boldsymbol{\mu}), \tag{2.7}$$

where $\mathbf{V}(\boldsymbol{\mu}) = \text{diag}[V(\mu_1), \dots, V(\mu_k)]$ denotes the diagonal variance function, \odot is the Hadamard (elementwise) product, and $\tilde{\mathbf{V}}(\boldsymbol{\mu})$ denotes the matrix with elements $V^{1/2}(\mu_i) V^{1/2}(\mu_j)$ for $i, j = 1, \dots, k$. This is similar to the variance function (4.20) for multivariate exponential dispersion models obtained below.

In general, there seems to be no simple constructive way of connecting the scaled deviance $D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the function $a(\mathbf{y}; \boldsymbol{\Sigma})$ to make (2.5) a probability density function. However, the p^* -formula (or the saddlepoint approximation) of Barndorff-Nielsen (1983) gives the following asymptotic form for $a(\mathbf{y}; \boldsymbol{\Sigma})$:

$$\begin{aligned} a(\mathbf{y}; \boldsymbol{\Sigma}) &\sim (2\pi)^{-k/2} |\mathbf{V}_{\boldsymbol{\Sigma}}(\mathbf{y})|^{-1/2} \\ &= (2\pi)^{-k/2} |\boldsymbol{\Sigma}|^{-1/2} \mathbf{V}^{-1/2}(\mathbf{y}) \quad \text{for } \mathbf{y} \in \Omega^k, \end{aligned} \tag{2.8}$$

which holds in the *small-dispersion limit* $\|\boldsymbol{\Sigma}\| \rightarrow 0$, where $\|\cdot\|$ denotes Euclidean norm. This is closely related with the class of multivariate proper dispersion models, which we consider next.

3 Multivariate proper dispersion models

3.1 Construction

We call a multivariate dispersion model *proper* if the function $a(\mathbf{y}; \boldsymbol{\Sigma})$ of (2.5) factorizes as $a(\boldsymbol{\Sigma})b(\mathbf{y})$, and we shall now review the main method for constructing multivariate proper dispersion models proposed by Jørgensen and Lauritzen (2000). For this purpose, it is useful to parametrize the univariate dispersion model (2.1) by μ and $\lambda = \sigma^{-2}$ instead of μ and σ^2 . Similarly, we use $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ freely instead of $\boldsymbol{\Sigma}$ in the notation, writing $a(\mathbf{y}; \boldsymbol{\Lambda})$ instead of $a(\mathbf{y}; \boldsymbol{\Sigma})$ and $D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ instead of $D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and so on.

Let us consider a univariate proper dispersion model $\text{PD}_1(\mu, \lambda)$, say, of the form

$$f(y; \mu, \lambda) = a_1(\lambda) V^{-1/2}(y) \exp\left[-\frac{\lambda}{2} d(y; \mu)\right] \quad \text{for } y \in \Omega, \tag{3.1}$$

where the subscript 1 on the normalizing constant a , etc. indicates the dimension. Essentially all univariate proper dispersion models are of the form (3.1) (Jørgensen, 1997, pages 182–183).

A crucial feature of (3.1) is the fact that the normalizing constant $a_1(\lambda)$ depends on (μ, λ) only through λ . We may also express this by saying that $d(Y; \mu)$ is a pivot with respect to the measure $V^{-1/2}(y) dy$. We now make the further assumption that the deviance residual $r(Y; \mu)$, as defined above, is also a pivot. Following Jørgensen and Lauritzen (2000), we shall now see how this assumption may be utilized for constructing a multivariate version of (3.1). The main idea is that if $r(Y; \mu)$ is a pivot, then so is the vector of deviance residuals $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu})$ with respect to the product measure $|\mathbf{V}(\mathbf{y})|^{-1/2} d\mathbf{y} = V^{-1/2}(y_1) dy_1 \otimes \cdots \otimes V^{-1/2}(y_k) dy_k$, where $\mathbf{V}(\boldsymbol{\mu})$ is the diagonal variance function defined above. We define a *multivariate proper dispersion model* $\text{PD}_k(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ by the following probability density function:

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) = a_k(\boldsymbol{\Lambda}) |\mathbf{V}(\mathbf{y})|^{-1/2} \exp\left[-\frac{1}{2} D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})\right] \quad \text{for } \mathbf{y} \in \Omega^k. \tag{3.2}$$

The fact that the normalizing constant $a_k(\mathbf{\Lambda})$ depends on $(\boldsymbol{\mu}, \mathbf{\Lambda})$ only through $\mathbf{\Lambda}$ follows because the scaled deviance $D(\mathbf{Y}; \boldsymbol{\mu}, \mathbf{\Lambda})$ is also a pivot with respect to the product measure $|\mathbf{V}(\mathbf{y})|^{-1/2} d\mathbf{y}$. In the special case where $\mathbf{\Lambda} = \text{diag}\{\lambda_{11}, \dots, \lambda_{kk}\}$, it follows from (2.4) that the marginals of \mathbf{Y} are independent with distributions $\text{PD}_1(\mu_j, \lambda_{jj})$. Jørgensen and Lauritzen (2000) considered a number of examples of (3.2), including multivariate gamma, simplex and von Mises distributions. We also note that the construction is tied to a particular coordinate system, such that $\text{PD}_k(\boldsymbol{\mu}, \mathbf{\Lambda})$ has support on the product space Ω^k .

3.2 Monte Carlo methods

The main mathematical problem in connection with (3.2) is to calculate the normalizing constant $a_k(\mathbf{\Lambda})$, and we shall now see how this can be done by Monte Carlo methods. Let us define the scaled normalizing constant by

$$\bar{a}_k(\mathbf{\Lambda}) = \frac{a_k(\mathbf{\Lambda})}{(2\pi)^{-k/2} |\mathbf{\Lambda}|^{1/2}}, \tag{3.3}$$

so that the saddlepoint approximation (2.8) is equivalent to $\bar{a}_k(\mathbf{\Lambda}) \sim 1$ in the small-dispersion limit. It follows that both \mathbf{Y} and $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu})$ are approximately multivariate normal in the small-dispersion limit (Jørgensen and Rajeswaran, 2005). For $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu})$ we may express this result as the following multivariate normal approximation:

$$\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu}) \sim N_k(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{for } \|\boldsymbol{\Sigma}\| \text{ small.} \tag{3.4}$$

The first of the two Monte Carlo methods that we consider is due to Jørgensen and Lauritzen (2000), who explored the following additive property of the scaled deviance, obtained from (2.3):

$$D(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Lambda}) = D(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Lambda}_0) + D(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Lambda} - \mathbf{\Lambda}_0),$$

where $\mathbf{\Lambda}_0 = \text{diag}\{\lambda_{11}, \dots, \lambda_{kk}\}$ contains the diagonal elements of $\mathbf{\Lambda}$. We may hence calculate the normalizing constant as follows:

$$\bar{a}_k^{-1}(\mathbf{\Lambda}) = \prod_{j=1}^k \bar{a}_1^{-1}(\lambda_{jj}) \mathbb{E}_0 \left(\exp \left[-\frac{1}{2} D(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Lambda} - \mathbf{\Lambda}_0) \right] \right), \tag{3.5}$$

where \mathbb{E}_0 denotes expectation with respect to the distribution $\text{PD}_k(\boldsymbol{\mu}, \mathbf{\Lambda}_0)$ with independent marginals of the form (3.1). It is usually simple to simulate from the univariate distributions $\text{PD}_1(\mu_j, \lambda_{jj})$, and, hence, (3.5) gives a practical way of calculating $a_k(\mathbf{\Lambda})$ by simulation.

The second Monte Carlo method explores the approximate normality (3.4) for $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu})$. However, rather than using the approximation directly, we shall instead define a new distribution for \mathbf{Y} by assuming that $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu}) \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$ holds exactly. Since the function $\mathbf{r}(\cdot; \boldsymbol{\mu})$ is injective, we can find the distribution of \mathbf{Y} by an

easy transformation. Using the form of the multivariate normal probability density function, we find from (3.2) and (3.3) that \mathbf{Y} now has probability density function of the form

$$g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) = J(\mathbf{y}; \boldsymbol{\mu})(2\pi)^{-k/2} |\boldsymbol{\Lambda}|^{1/2} \exp[-\frac{1}{2}D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})] \quad (3.6)$$

$$= \bar{J}(\mathbf{y}; \boldsymbol{\mu}) \bar{a}_k^{-1}(\boldsymbol{\Lambda}) f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}), \quad (3.7)$$

where $J(\mathbf{y}; \boldsymbol{\mu})$ denotes the Jacobian of the transformation $\mathbf{r}(\cdot; \boldsymbol{\mu})$ and

$$\bar{J}(\mathbf{y}; \boldsymbol{\mu}) = J(\mathbf{y}; \boldsymbol{\mu}) |\mathbf{V}(\mathbf{y})|^{1/2},$$

which we call the standardized Jacobian. For reasons that will become clear below (cf. Section 3.3), we shall call (3.6) a *multivariate BS-like distribution*. This distribution is not in general a proper dispersion model, because of the dependence of $J(\mathbf{y}; \boldsymbol{\mu})$ on $\boldsymbol{\mu}$, but the similarity between (3.6) and the $\text{PD}_k(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ density (3.2) turns out to be useful for simulation purposes. It is easy to simulate from $g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ by simulating \mathbf{R} from the multivariate normal distribution $\text{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$ and solving $\mathbf{r}(\mathbf{Y}; \boldsymbol{\mu}) = \mathbf{R}$ to obtain \mathbf{Y} .

To derive the form of the standardized Jacobian $\bar{J}(\mathbf{y}; \boldsymbol{\mu})$, we first note, using the notation for derivatives introduced above, that

$$\dot{r}_y(y; \boldsymbol{\mu}) = \pm \frac{\dot{d}_y(y; \boldsymbol{\mu})}{2\sqrt{d(y; \boldsymbol{\mu})}} = \frac{\dot{d}_y(y; \boldsymbol{\mu})}{2r(y; \boldsymbol{\mu})}. \quad (3.8)$$

Using a result from Jørgensen (1997, page 24) along with l'Hospital's rule, we obtain from (2.6) that, for y near $\boldsymbol{\mu}$,

$$\frac{\dot{d}_y(y; \boldsymbol{\mu})}{2r(y; \boldsymbol{\mu})} \sim \frac{\ddot{d}_{yy}(y; \boldsymbol{\mu})}{2\dot{r}_y(y; \boldsymbol{\mu})} \sim \frac{V^{-1}(y)}{\dot{r}_y(y; \boldsymbol{\mu})}. \quad (3.9)$$

Combining (3.8) and (3.9), we obtain $\dot{r}_y(y; \boldsymbol{\mu}) \doteq V^{-1/2}(y)$, and, hence, the standardized Jacobian satisfies

$$\bar{J}(\mathbf{y}; \boldsymbol{\mu}) = \prod_{j=1}^k [\dot{r}_y(y_j; \boldsymbol{\mu}_j) V^{1/2}(y_i)] \sim 1,$$

so that by (3.7) the densities $f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ and $g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ are asymptotically proportional for \mathbf{y} near $\boldsymbol{\mu}$.

In order to calculate the normalizing constant $\bar{a}_k(\boldsymbol{\Lambda})$ by simulation, we define

$$\begin{aligned} h(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \bar{a}_k^{-1}(\boldsymbol{\Lambda}) f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) \\ &= (2\pi)^{-k/2} |\boldsymbol{\Lambda}|^{1/2} |\mathbf{V}(\mathbf{y})|^{-1/2} \exp[-\frac{1}{2}D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})]. \end{aligned}$$

Then

$$\begin{aligned} \bar{a}_k^{-1}(\boldsymbol{\Lambda}) &= \int h(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) d\mathbf{y} = \int \frac{h(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})}{g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})} g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) d\mathbf{y} \\ &= \mathbb{E}_g[\bar{J}^{-1}(\mathbf{Y}; \boldsymbol{\mu})], \end{aligned} \quad (3.10)$$

where \mathbb{E}_g denotes expectation with respect to $g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$. We may hence calculate $\bar{a}_k^{-1}(\boldsymbol{\Lambda})$ by importance sampling, that is, simulating \mathbf{Y} from the multivariate BS-like distribution $g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ as indicated above, and calculating the simulation average of $\bar{J}^{-1}(\mathbf{Y}; \boldsymbol{\mu})$. It seems likely that the use of antithetic variables or a control variate could improve the efficiency of this simulation somewhat.

In case the standardized Jacobian $\bar{J}(\mathbf{y}; \boldsymbol{\mu})$ is bounded below by a positive constant c , then $h(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) \leq c^{-1}g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$, such that $g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ may be used for rejection sampling from $f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$. The following examples illustrate this method.

3.3 A multivariate hyperbola distribution

As an example of a multivariate proper dispersion model, we consider a multivariate generalization of the hyperbola distribution, defined by the probability density function

$$f(y; \mu, \lambda) = \frac{e^{-\lambda}}{2K_0(\lambda)} y^{-1} \exp\left[-\frac{\lambda}{2} \frac{(y - \mu)^2}{y\mu}\right] \quad \text{for } y > 0,$$

where $\mu, \lambda > 0$ and K_0 is a Bessel function; see Jørgensen (1997, page 192). This is a univariate proper dispersion model with unit variance function $V(\mu) = \mu^2$ for $\mu > 0$. Following Jørgensen and Lauritzen (2000), we consider the corresponding multivariate proper dispersion model with deviance residual:

$$r(y; \mu) = \frac{y - \mu}{\sqrt{y\mu}} = \sqrt{\frac{y}{\mu}} - \sqrt{\frac{\mu}{y}}, \tag{3.11}$$

where clearly $r(Y; \mu)$ is a pivot since μ is a scale parameter and the product measure $|\mathbf{V}(\mathbf{y})|^{-1/2} d\mathbf{y}$ is scale invariant. This gives a multivariate hyperbola distribution $\text{Hy}_k(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ of the form

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \bar{a}_k(\boldsymbol{\Lambda})(2\pi)^{-k/2} |\boldsymbol{\Lambda}|^{1/2} \prod_{j=1}^k y_j^{-1} \exp\left[-\frac{1}{2} D(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Lambda})\right],$$

where D is derived from the vector of deviance residuals corresponding to (3.11).

The normalizing constant $\bar{a}_k(\boldsymbol{\Lambda})$ may be simulated by importance sampling using (3.10). For this purpose we need the standardized Jacobian, which takes the form

$$\bar{J}(\mathbf{y}; \boldsymbol{\mu}) = 2^{-k} \prod_{j=1}^k \left(\sqrt{\frac{y_j}{\mu_j}} + \sqrt{\frac{\mu_j}{y_j}} \right) \geq 1. \tag{3.12}$$

In this case, the form of the multivariate BS-like distribution (3.6) is obtained from the multivariate normal distribution $\text{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$ by transforming coordinatewise by the inverse of the transformation $r(\cdot; \mu)$ defined by (3.11), which yields a multivariate Birnbaum–Saunders distribution (Díaz-García and Domínguez-Molina, 2006; Kundu et al., 2010).

Due to the inequality (3.12), we may simulate from the multivariate hyperbola distribution by rejection sampling using the multivariate Birnbaum–Saunders distribution (3.6). Given a random vector \mathbf{Y} from the multivariate Birnbaum–Saunders distribution, and an independent uniform random variable U , we accept \mathbf{Y} as a multivariate hyperbola random vector provided that $U < \bar{J}^{-1}(\mathbf{Y}; \boldsymbol{\mu})$. This provides a straightforward simulation method for the multivariate hyperbola distribution.

3.4 Statistical inference

To illustrate the analogies between dispersion models and the multivariate normal distribution, we shall now summarize the main results of Jørgensen and Rajeswaran (2005) concerning statistical inference for multivariate proper dispersion models.

Let us consider inference based on a random sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ from the multivariate proper dispersion model $\text{PD}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ defined by (3.2). Jørgensen and Rajeswaran (2005) developed an asymptotic approach based on a combination of conventional large-sample asymptotics and small-dispersion asymptotics. The latter is derived by means of the saddlepoint approximation, which implies convergence to the multivariate normal distribution as follows:

$$\boldsymbol{\Sigma}^{-1/2} \mathbf{V}^{-1/2}(\boldsymbol{\mu})(\mathbf{Y}_i - \boldsymbol{\mu}) \xrightarrow{d} \text{N}_k(\mathbf{0}, \mathbf{I}_k)$$

for $\|\boldsymbol{\Sigma}\|$ small, where \xrightarrow{d} denotes convergence in distribution, which in turn implies the normal approximation (3.4) from above.

Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ denote the maximum likelihood estimators of the two parameters. We obtain the following approximations in the small-dispersion limit:

$$\hat{\boldsymbol{\mu}} \approx \bar{\mathbf{Y}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} \approx \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{Y}_i; \bar{\mathbf{Y}}_n) \mathbf{r}^\top(\mathbf{Y}_i; \bar{\mathbf{Y}}_n).$$

The precise meaning of these and other approximations in the following is made clear by Jørgensen and Rajeswaran (2005). An approximately unbiased estimator \mathbf{S}_n for $\boldsymbol{\Sigma}$ is obtained by correcting the degrees of freedom in the usual way, giving

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^n \mathbf{r}(\mathbf{Y}_i; \bar{\mathbf{Y}}) \mathbf{r}^\top(\mathbf{Y}_i; \bar{\mathbf{Y}}).$$

We may now obtain an asymptotic version of Hotelling's T^2 test for a hypothesis of the form $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ with $\boldsymbol{\Sigma}$ unknown. Let us define

$$\bar{\mathbf{r}}(\boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{Y}_i; \boldsymbol{\mu}).$$

Then under H_0 , the statistics $\bar{\mathbf{r}}(\boldsymbol{\mu}_0)$ and \mathbf{S}_n are asymptotically independent in the small-dispersion limit, with asymptotic distributions

$$\bar{\mathbf{r}}(\boldsymbol{\mu}_0) \approx N_k\left(\mathbf{0}, \frac{1}{n}\boldsymbol{\Sigma}\right), \quad (n-1)\mathbf{S}_n \approx \mathcal{W}_{n-1}(\boldsymbol{\Sigma}),$$

where $\mathcal{W}_{n-1}(\boldsymbol{\Sigma})$ denotes the Wishart distribution with $n-1$ degrees of freedom. We define an analogue of Hotelling's T^2 -statistic by

$$T^2 = n\bar{\mathbf{r}}^\top(\boldsymbol{\mu}_0)\mathbf{S}_n^{-1}\bar{\mathbf{r}}(\boldsymbol{\mu}_0) \stackrel{H_0}{\approx} \frac{k(n-1)}{n-k} F_{k,n-k}. \tag{3.13}$$

The F approximation holds if either $\|\boldsymbol{\Sigma}\|$ is small or if the sample size n is large, the latter in the sense that $F_{k,n-k}$ is asymptotically proportional to a χ_k^2 distribution for n large, which in turn represents the large-sample distribution of T^2 (Jørgensen and Rajeswaran, 2005).

These results indicate that asymptotic theory for multivariate dispersion models have many analogies with exact results from classical multivariate analysis; see also the discussion of multivariate generalized linear models in Section 4.6. Note by comparison that certain results for the multivariate normal distribution hold exactly for elliptically contoured distributions (Anderson and Fang, 1987).

4 Multivariate exponential dispersion models

4.1 General

We shall now develop a new version of what we may call the *convolution method* for constructing multivariate exponential dispersion models, and we begin by discussing the bivariate case. The method is based on the following stochastic representation for the random vector $\mathbf{X} = (X_1, X_2)^\top$:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_{11} \\ U_{12} \end{bmatrix} + \begin{bmatrix} U_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ U_2 \end{bmatrix}, \tag{4.1}$$

where the three vectors on the right-hand side are assumed independent. Let $\text{ED}^*(\mu, \lambda)$ be a given univariate additive exponential dispersion model (cf. Section 4.2) with mean $\lambda\mu$ and variance $\lambda V(\mu)$, say, and assume that $U_j \sim \text{ED}^*(\mu, \lambda_j)$ for $j = 1, 2$. The conventional convolution method, called the variables-in-common method by Balakrishnan and Lai (2009), is obtained by assuming that $U_{11} = U_{12} \sim \text{ED}^*(\mu, \lambda_{12})$. The convolution property (4.7) for ED^* models (see below) implies that $X_j \sim \text{ED}^*(\mu, \lambda_{jj})$, where $\lambda_{jj} = \lambda_j + \lambda_{12}$ for $j = 1, 2$, such that both marginal distributions belong to the given ED^* model. The notation introduced here is slightly more complicated than necessary, but will become useful later on.

The result is a four-parameter bivariate family of distributions, which unfortunately is one parameter short of the goal of five parameters (two means and three

variances/covariances). Nevertheless, the decomposition (4.1) seems like a sensible way to interpolate between independence ($\lambda_{12} = 0$) and complete dependence ($\lambda_1 = \lambda_2 = 0$) between X_1 and X_2 , which embodies the traditional way of creating correlated normal random variables. One drawback of the method is that the correlation between the the two variables is always positive,

$$\text{Corr}(X_1, X_2) = \frac{\lambda_{12}}{\sqrt{\lambda_{11}\lambda_{22}}},$$

although we know from the normal case that it may be possible by analytical methods to extend the domain of the correlation to negative values. It is also required that the ED* model be infinitely divisible, allowing all three λ parameters to vary freely in \mathbb{R}_+ .

The main complication in extending the above method to the k -dimensional case is the combinatorial explosion of the number of terms necessary in order to generalize (4.1). However, in order to construct a family with a fully flexible covariance structure with k means and $k(k+1)/2$ covariance parameters, we need to generalize the convolution technique slightly, which we do by abandoning the requirement $U_{11} = U_{12}$ and instead work with a joint distribution for U_{11} and U_{12} .

4.2 Exponential dispersion models

Before moving on to the extended convolution method in Section 4.3 we need to review some basic facts about natural exponential families and exponential dispersion models. A *natural exponential family* is defined by the probability density functions

$$f(\mathbf{x}; \boldsymbol{\theta}) = a(\mathbf{x}) \exp[\mathbf{x}^\top \boldsymbol{\theta} - \kappa(\boldsymbol{\theta})] \quad \text{for } \mathbf{x} \in \mathbb{R}^k, \quad (4.2)$$

with respect to a suitable measure on \mathbb{R}^k , for some function a , where the domain for $\boldsymbol{\theta}$ is the set

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^k : \int a(\mathbf{x}) e^{\mathbf{x}^\top \boldsymbol{\theta}} d\mathbf{x} < \infty \right\}.$$

We refer to κ as the *cumulant function*. We assume that Θ contains an open subset, and that the distribution (4.2) is not concentrated on any affine subspace of \mathbb{R}^k .

The distribution corresponding to (4.2) has cumulant generating function (CGF) given by

$$\kappa_\theta(\mathbf{s}) = \kappa(\mathbf{s} + \boldsymbol{\theta}) - \kappa(\boldsymbol{\theta}) \quad \text{for } \mathbf{s} \in \Theta - \boldsymbol{\theta}. \quad (4.3)$$

The mean vector and covariance matrix of a random vector \mathbf{X} distributed according to (4.2) may hence be obtained by differentiating (4.3) and setting \mathbf{s} to zero. It follows that the mean vector is

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = \dot{\kappa}(\boldsymbol{\theta}), \quad (4.4)$$

where again dots denote derivatives. Since the covariance matrix $\ddot{\kappa}(\boldsymbol{\theta})$ is positive-definite, the mapping $\dot{\kappa}$ is one-to-one, and we may hence parametrize (4.2) by the mean vector (4.4) for $\boldsymbol{\theta} \in \text{int } \Theta$. We may also express the covariance matrix as a function of the mean vector, which gives rise to the (matrix) variance function

$$\mathbf{V}(\boldsymbol{\mu}) = \ddot{\kappa} \circ \dot{\kappa}^{-1}(\boldsymbol{\mu}) \quad \text{for } \boldsymbol{\mu} \in \Omega, \quad (4.5)$$

where $\Omega = \dot{\kappa}(\text{int } \Theta)$ is the domain of $\boldsymbol{\mu}$.

The *additive exponential dispersion model* generated from (4.2) is defined by the probability density function

$$f^*(\mathbf{x}; \boldsymbol{\theta}, \lambda) = a^*(\mathbf{x}; \lambda) \exp[\mathbf{x}^\top \boldsymbol{\theta} - \lambda \kappa(\boldsymbol{\theta})] \quad \text{for } \mathbf{x} \in \mathbb{R}^k, \quad (4.6)$$

for some function $a^*(\mathbf{x}; \lambda)$, which corresponds to replacing the cumulant function κ by $\lambda \kappa$ in (4.2). We assume that (4.6) is infinitely divisible, such that λ has domain \mathbb{R}_+ . The mean vector of (4.6) is $\lambda \boldsymbol{\mu}$ and the covariance matrix is $\lambda \mathbf{V}(\boldsymbol{\mu})$, where \mathbf{V} , defined by (4.5), is now called the unit variance function. We let $\text{ED}^*(\boldsymbol{\mu}, \lambda)$ denote the distribution corresponding to (4.6), a model with $k + 1$ parameters. This model satisfies the convolution property

$$\text{ED}^*(\boldsymbol{\mu}, \lambda_1) + \text{ED}^*(\boldsymbol{\mu}, \lambda_2) = \text{ED}^*(\boldsymbol{\mu}, \lambda_1 + \lambda_2) \quad \text{for } \lambda_1, \lambda_2 > 0. \quad (4.7)$$

The *reproductive exponential dispersion model* generated from (4.2) is defined by applying the *duality transformation* $\mathbf{Y} = \mathbf{X}/\lambda$ to (4.6), giving

$$f(\mathbf{y}; \boldsymbol{\theta}, \lambda) = a(\mathbf{y}; \lambda) \exp\{\lambda[\mathbf{y}^\top \boldsymbol{\theta} - \kappa(\boldsymbol{\theta})]\} \quad \text{for } \mathbf{y} \in \mathbb{R}^k \quad (4.8)$$

for some function $a(\mathbf{y}; \lambda)$. A random vector \mathbf{Y} distributed according to (4.8) has mean $\boldsymbol{\mu}$ defined by (4.4), and variance

$$\text{Var}(\mathbf{Y}) = \lambda^{-1} \mathbf{V}(\boldsymbol{\mu}). \quad (4.9)$$

In the univariate case ($k = 1$), the reproductive exponential dispersion model (4.8) may be rewritten in the dispersion model form (2.1) (Jørgensen, 1997, page 77). In the multivariate case we see that the variance (4.9) is governed by the single parameter λ . We shall now see how to generalize this to the form $\boldsymbol{\Sigma} \odot \mathbf{V}(\boldsymbol{\mu})$ involving a matrix $\boldsymbol{\Sigma}$ [compare with (2.7)], where \odot is the Hadamard product introduced in Section 2.

4.3 The bivariate case

Continuing with the bivariate case, we shall now develop the extended convolution method in order to obtain a fully flexible covariance structure. Our starting point is a bivariate natural exponential family of the form (4.2), where the cumulant function $\kappa(\theta_1, \theta_2)$ is now considered as a function of the two coordinates of the vector $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ with domain Θ , say. We let s and t be the arguments of

the CGF $\kappa_\theta(s, t)$ defined by (4.3). Hence, the corresponding additive exponential dispersion model has CGF

$$(s, t) \mapsto \lambda_{12}\kappa_\theta(s, t), \tag{4.10}$$

where the parameter λ_{12} is now called the *weight*. Since we assume infinite divisibility, the weight parameter λ_{12} has domain \mathbb{R}_+ .

Compared with the construction of Section 4.1, we now abandon the assumption $U_{11} = U_{12}$ and instead assume that the joint distribution of U_{11} and U_{12} has CGF given by (4.10). We assume once again that the marginal distributions of U_{11} and U_1 belong to the same family, and, similarly, that the marginal distributions of U_{12} and U_2 belong to the same family. We may think of these families as corresponding to the vectors $(U_1, 0)^\top$ and $(0, U_2)^\top$ respectively, so we represent them as (degenerate) bivariate distributions. Hence, let us assume that $(U_1, 0)^\top$ has CGF with positive weight λ_1 , defined by

$$(s, t) \mapsto \lambda_1\kappa_\theta(s, 0). \tag{4.11}$$

Similarly, let $(0, U_2)$ have CGF with positive weight λ_2 , defined by

$$(s, t) \mapsto \lambda_2\kappa_\theta(0, t). \tag{4.12}$$

We now add the three terms (4.10), (4.11) and (4.12), giving the following bivariate CGF for the random vector \mathbf{X} ,

$$K_{\theta,\lambda}(s, t) = \lambda_{12}\kappa_\theta(s, t) + \lambda_1\kappa_\theta(s, 0) + \lambda_2\kappa_\theta(0, t). \tag{4.13}$$

We note that the marginal distribution of X_1 has the same form as (4.11), but with λ_1 replaced by $\lambda_{11} = \lambda_{12} + \lambda_1$, and, similarly, the marginal distribution of X_2 has λ_2 of (4.12) replaced by $\lambda_{22} = \lambda_{12} + \lambda_2$. Both marginal distributions in general depend on both parameters θ_1 and θ_2 . Also note that (4.13), contrary to (4.6), is not a natural exponential family for fixed values of the λ parameter(s).

We shall now calculate the mean vector and covariance matrix for \mathbf{X} by differentiating $K_{\theta,\lambda}$. Let $\dot{\kappa}_1(\theta_1, \theta_2)$ and $\dot{\kappa}_2(\theta_1, \theta_2)$ denote the two components of $\dot{\kappa}(\theta_1, \theta_2)$, and let $\ddot{\kappa}_{ij}(\theta_1, \theta_2)$ for $i, j = 1, 2$ denote the second order derivatives of κ . Then

$$E(\mathbf{X}) = \begin{bmatrix} (\lambda_{12} + \lambda_1)\dot{\kappa}_1(\theta_1, \theta_2) \\ (\lambda_{12} + \lambda_1)\dot{\kappa}_2(\theta_1, \theta_2) \end{bmatrix} = \begin{bmatrix} \lambda_{11}\mu_1 \\ \lambda_{22}\mu_2 \end{bmatrix}, \tag{4.14}$$

say, where the μ_j are the components of $\boldsymbol{\mu}$ defined by (4.4).

We can now express the covariance matrix for \mathbf{X} in terms of the 2×2 unit variance function for (4.6), with entries defined as follows:

$$\mathbf{V}(\boldsymbol{\mu}) = \begin{bmatrix} V_{11}(\boldsymbol{\mu}) & V_{12}(\boldsymbol{\mu}) \\ V_{21}(\boldsymbol{\mu}) & V_{22}(\boldsymbol{\mu}) \end{bmatrix}.$$

The resulting covariance matrix for \mathbf{X} is

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= \begin{bmatrix} \lambda_{11}\ddot{\kappa}_{11}(\theta_1, \theta_2) & \lambda_{12}\ddot{\kappa}_{12}(\theta_1, \theta_2) \\ \lambda_{12}\ddot{\kappa}_{21}(\theta_1, \theta_2) & \lambda_{22}\ddot{\kappa}_{22}(\theta_1, \theta_2) \end{bmatrix} \\ &= \begin{bmatrix} \lambda_{11}V_{11}(\boldsymbol{\mu}) & \lambda_{12}V_{12}(\boldsymbol{\mu}) \\ \lambda_{12}V_{21}(\boldsymbol{\mu}) & \lambda_{22}V_{22}(\boldsymbol{\mu}) \end{bmatrix}, \end{aligned} \tag{4.15}$$

which is of the form $\Lambda \odot \mathbf{V}(\boldsymbol{\mu})$, where Λ is the *weight matrix*, defined by

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}. \tag{4.16}$$

In particular, we find that the correlation between X_1 and X_2 is

$$\text{Corr}(X_1, X_2) = \frac{\lambda_{12}}{\sqrt{\lambda_{11}\lambda_{22}}} \frac{V_{12}(\boldsymbol{\mu})}{\sqrt{V_{11}(\boldsymbol{\mu})V_{22}(\boldsymbol{\mu})}}, \tag{4.17}$$

which has the same sign as the original correlation

$$\frac{V_{12}(\boldsymbol{\mu})}{\sqrt{V_{11}(\boldsymbol{\mu})V_{22}(\boldsymbol{\mu})}} \tag{4.18}$$

for the components of the bivariate natural exponential family that we started from. We also note that the absolute value of the correlation (4.17) is bounded by the absolute value of (4.18).

This construction gives the additive form of the bivariate exponential dispersion model, denoted $\text{ED}_2^*(\boldsymbol{\mu}, \Lambda)$, which satisfies the following convolution property:

$$\text{ED}_2^*(\boldsymbol{\mu}, \Lambda_1) + \text{ED}_2^*(\boldsymbol{\mu}, \Lambda_2) = \text{ED}_2^*(\boldsymbol{\mu}, \Lambda_1 + \Lambda_2), \tag{4.19}$$

generalizing (4.7). We hence call this five-parameter family an *additive bivariate exponential dispersion model*. Just like in the univariate case, this form of the bivariate exponential dispersion model is particularly suited for discrete data (Jørgensen, 1997, page 76), thereby complementing the class of proper dispersion models, which are confined to the continuous case.

As is evident from (4.19), the domain for Λ is an additive semigroup. It seems plausible that the set of values for Λ for which (4.13) is a CGF is bigger than the domain allowed in the stochastic representation (4.1), possibly containing negative values for λ_{12} , but this issue remains to be investigated. In the affirmative case the correlation (4.17) would take both positive and negative values.

As the final step in the construction, we shall now derive the reproductive form of the bivariate exponential dispersion model by means of an analogy of the duality transformation used in connection with (4.8) above. We hence define the random vector $\mathbf{Y} = (Y_1, Y_2)^\top$ as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1/\lambda_{11} \\ X_2/\lambda_{22} \end{bmatrix},$$

with mean vector

$$\text{E}(\mathbf{X}) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and covariance matrix

$$\text{Cov}(\mathbf{Y}) = \begin{bmatrix} \frac{1}{\lambda_{11}} V_{11}(\boldsymbol{\mu}) & \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} V_{12}(\boldsymbol{\mu}) \\ \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} V_{21}(\boldsymbol{\mu}) & \frac{1}{\lambda_{22}} V_{22}(\boldsymbol{\mu}) \end{bmatrix} = \boldsymbol{\Sigma} \odot V(\boldsymbol{\mu}), \tag{4.20}$$

say, where Σ is the symmetric positive-definite matrix defined by

$$\Sigma = \begin{bmatrix} \frac{1}{\lambda_{11}} & \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} \\ \frac{\lambda_{12}}{\lambda_{11}\lambda_{22}} & \frac{1}{\lambda_{22}} \end{bmatrix}. \tag{4.21}$$

We have hence obtained a covariance structure similar to the variance function (2.7) for general multivariate dispersion models. We denote the model corresponding to \mathbf{Y} by $ED_2(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ is the mean vector, and Σ is called the dispersion matrix.

4.4 The multivariate case

In order to work out the general case we shall now extend the above approach by considering a construction based on single variables and pairs of variables. In the trivariate case $k = 3$ we generalize (4.1) as follows:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} U_{11} \\ U_{12} \\ 0 \end{bmatrix} + \begin{bmatrix} U_{21} \\ 0 \\ U_{23} \end{bmatrix} + \begin{bmatrix} 0 \\ U_{32} \\ U_{33} \end{bmatrix} + \begin{bmatrix} U_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ U_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ U_3 \end{bmatrix}, \tag{4.22}$$

where the six terms on the right-hand side of (4.22) are assumed independent. We now start from a trivariate natural exponential family (4.2) with cumulant function $\kappa(\theta_1, \theta_2, \theta_3)$, say, and we let $\kappa_\theta(s, t, u)$ denote the corresponding version of (4.3). Proceeding in a similar fashion as above, we define the CGF of (4.22) to be

$$\begin{aligned} K_{\theta,\lambda}(s, t, u) &= \lambda_{12}\kappa_\theta(s, t, 0) + \lambda_{13}\kappa_\theta(s, 0, u) + \lambda_{23}\kappa_\theta(0, t, u) \\ &\quad + \lambda_1\kappa_\theta(s, 0, 0) + \lambda_2\kappa_\theta(0, t, 0) + \lambda_3\kappa_\theta(0, 0, u), \end{aligned}$$

where the weights $\lambda_{12}, \lambda_{13}, \lambda_{23}, \lambda_1, \lambda_2$ and λ_3 are all positive. This CGF defines the joint distribution of $\mathbf{X} = (X_1, X_2, X_3)^\top$ in such a way that each of the three families of marginal distributions is the same as for the additive exponential dispersion model (4.6).

To complete the construction, we transform from \mathbf{X} to \mathbf{Y} by the duality transformation like above. The result is a trivariate reproductive exponential dispersion model $ED_3(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu}$ obtained from (4.4), and dispersion matrix Σ defined by analogy with (4.21). In particular, the diagonal entries of Σ are the reciprocals of

$$\lambda_{11} = \lambda_1 + \lambda_{12} + \lambda_{13}, \quad \lambda_{22} = \lambda_2 + \lambda_{12} + \lambda_{23} \quad \text{and} \quad \lambda_{33} = \lambda_3 + \lambda_{13} + \lambda_{23},$$

respectively. Again we have achieved a fully flexible covariance structure of the form $\text{Cov}(\mathbf{Y}) = \Sigma \odot V(\boldsymbol{\mu})$.

For general k we proceed in a similar fashion, and define a multivariate reproductive exponential dispersion model $ED_k(\boldsymbol{\mu}, \Sigma)$ by starting from a k -variate natural exponential family, and adding univariate and bivariate terms similar to

(4.22) with a total of $k + k(k - 1)/2$ terms, yielding the desired $k + k(k + 1)/2$ parameters. In principle, one could obtain additional covariance parameters by, for example, adding a further independent term of the form $(U_{41}, U_{42}, U_{43})^\top$ to (4.22), but we avoid this complication in order to obtain models that can be parametrized by their first two moments. Compare with Joe (1996), who set the same type of goal for his construction of multivariate distributions based on recursively mixing conditional distributions.

There are both advantages and disadvantages to the extended convolution method as presented here. The main difficulty is how to construct the initial natural exponential family (4.2) from which the multivariate exponential dispersion model is obtained. Strictly speaking, the extended convolution method merely adds a full covariance structure to a given multivariate natural exponential family, rather than constructing a multivariate exponential dispersion model from scratch. However, the second of the bivariate gamma distributions considered below is constructed from scratch without requiring a bivariate natural exponential family to start with, suggesting that there may be a canonical construction available in certain cases. Further details about multivariate exponential dispersion models are available in Jørgensen (2011b).

The main advantage of the extended convolution method is that we obtain a multivariate family parametrized by the mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$, giving the desired number of parameters and full control over the first two moments of the distribution, subject to the constraints noted above. This, in turn, implies that parameter estimation by quasi-likelihood methods will be straightforward, as discussed in Section 4.6. It is also straightforward to simulate from the distribution, in the sense that if we can simulate from the distribution corresponding to $\lambda\kappa_\theta(s, t, u)$, say, for all values of the parameters λ and $\boldsymbol{\theta}$, then we can simulate from the marginal and bivariate distributions entering (4.22), and add up the terms in order to obtain a simulated value of \mathbf{X} . The fact that the probability density function is difficult to obtain, because of the need to perform the multiple integration implicit from (4.22), is hence less of a practical concern.

4.5 Bivariate gamma distributions

To illustrate some of the issues discussed above, we shall now consider two different bivariate gamma distributions. We note in passing that a bivariate gamma distribution of proper dispersion model form was introduced by Jørgensen and Lauritzen (2000), using the technique described in Section 3.

Let us consider Kibble and Moran's bivariate gamma distribution (Kotz et al., 2000, page 436), following Letac (2007); see also Bernardoff et al. (2008). Actually, Letac attributes the distribution to Wicksell (1933). Let the parameter $\rho > 0$ be fixed, let $\lambda > 0$, and consider the additive exponential dispersion model of the form (4.6) corresponding to

$$a^*(x_1, x_2; \lambda) = \frac{(x_1 x_2)^{\lambda-1}}{\Gamma(\lambda)} \sum_{n=0}^{\infty} \frac{(\rho x_1 x_2)^n}{n! \Gamma(\lambda + n)} \quad \text{for } (x_1, x_2)^\top \in \mathbb{R}_+^2.$$

The corresponding cumulant function is

$$\kappa(\theta_1, \theta_2) = -\log(\theta_1\theta_2 - \rho),$$

defined on the set

$$\Theta = \{(\theta_1, \theta_2) : \theta_1 < 0, \theta_2 < 0, \theta_1\theta_2 - \rho > 0\}.$$

The corresponding reproductive exponential dispersion model (4.8) has mean vector

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \frac{1}{\rho - \theta_1\theta_2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

with domain \mathbb{R}_+^2 and covariance matrix

$$\frac{1}{\lambda} \begin{bmatrix} \mu_1^2 & \phi\mu_1\mu_2 \\ \phi\mu_1\mu_2 & \mu_2^2 \end{bmatrix},$$

where

$$\phi = 1 - \frac{1}{2\rho\mu_1\mu_2}(\sqrt{1 + 4\rho\mu_1\mu_2} - 1).$$

The corresponding bivariate exponential dispersion model $\text{ED}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has gamma marginals, and has covariance matrix of the form (4.20),

$$\begin{bmatrix} \sigma_{11}\mu_1^2 & \sigma_{12}\phi\mu_1\mu_2 \\ \sigma_{21}\phi\mu_1\mu_2 & \sigma_{22}\mu_2^2 \end{bmatrix}. \quad (4.23)$$

The presence of ϕ in (4.23), however, yields a fairly complicated covariance structure, with the correlation restricted to the interval $(0, \phi)$.

We now present a second bivariate gamma distribution, which is a special case of the multivariate gamma distribution of [Mathai and Moschopoulos \(1991\)](#). Let us consider the bivariate CGF defined by

$$\kappa_\mu(s, t) = -\log(1 - \mu_1s - \mu_2t),$$

with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, which corresponds to the distribution of the random vector

$$\begin{bmatrix} \mu_1U \\ \mu_2U \end{bmatrix}, \quad (4.24)$$

where U is a unit exponential random variable. Mimicking (4.13), we define a new bivariate CGF by

$$K_{\mu, \lambda}(s, t) = \lambda_{12}\kappa_\mu(s, t) + \lambda_1\kappa_\mu(s, 0) + \lambda_2\kappa_\mu(0, t), \quad (4.25)$$

whose marginal distributions are both gamma; see also [Balakrishnan and Lai \(2009, page 334\)](#) and references therein. The mean vector is again

$$\begin{bmatrix} \lambda_{11}\mu_1 \\ \lambda_{22}\mu_2 \end{bmatrix},$$

and the covariance matrix is

$$\begin{bmatrix} \lambda_{11}\mu_1^2 & \lambda_{12}\mu_2\mu_1 \\ \lambda_{12}\mu_2\mu_1 & \lambda_{22}\mu_2^2 \end{bmatrix},$$

using the same notation as above. By means of the duality transformation, we obtain the reproductive case $\text{Ga}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu}$ and covariance matrix

$$\begin{bmatrix} \sigma_{11}\mu_1^2 & \sigma_{12}\mu_1\mu_2 \\ \sigma_{21}\mu_1\mu_2 & \sigma_{22}\mu_2^2 \end{bmatrix},$$

where the σ_{ij} denote the entries of the dispersion matrix $\boldsymbol{\Sigma}$ defined by (4.21). The corresponding correlation ranges from 0 to 1. The result is a five-parameter bivariate gamma family with an intuitively appealing form of covariance matrix. It is straightforward to define a multivariate gamma distribution by proceeding along the same lines as for (4.22), using the bivariate gamma distribution $\text{Ga}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for each of the three first terms of (4.22). The resulting multivariate gamma distribution is more general than that of [Mathai and Moschopoulos \(1991\)](#); see [Jørgensen \(2011b\)](#) for further details.

This construction seems to have certain advantages over the construction based on Kibble and Moran's bivariate gamma distribution, not least its simplicity and the fact that it is generated in a canonical way from the univariate gamma distribution. This example highlights the fact that the extended convolution method in effect interpolates between, on the one hand, the bivariate generating distribution, for example, the natural exponential family (4.2), and the distribution with independent marginals. For this reason, the correlation of the starting bivariate distribution limits the range of possible correlations for the corresponding multivariate exponential dispersion model, and from this point of view the distribution (4.24), having completely correlated marginals, is ideal. The only slight disadvantage of the method is that only positive correlations are obtained, but again it seems plausible that (4.25) can be shown also to be a CGF for negative values of λ_{12} .

The key to the success of this construction lies in the scaling property of the gamma distribution, and it hence seems reasonable to speculate that the method can be extended to the whole class of Tweedie exponential dispersion models with power variance functions ([Jørgensen, 1997, Chapter 4](#)), a topic that is explored in [Jørgensen \(2011b\)](#).

4.6 Multivariate generalized linear models

The main motivation for constructing multivariate dispersion models comes from the need to develop flexible regression models for multivariate nonnormal data, and we shall now outline an approach based on multivariate exponential dispersion models.

Let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be independent k -vectors of response variables such that

$$\mathbf{Y}_i \sim \text{ED}_k(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}),$$

and consider a smooth regression model $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is an m -vector of regression parameters. Following Liang and Zeger (1986), we estimate $\boldsymbol{\beta}$ using the quasi-score function

$$\boldsymbol{\psi}_\beta(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{C}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \quad (4.26)$$

where $\mathbf{C}_i = \boldsymbol{\Sigma} \odot \mathbf{V}(\boldsymbol{\mu}_i)$ is the covariance matrix for \mathbf{Y}_i and \mathbf{D}_i is the local model matrix defined by $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}^\top$. The quasi-score estimator $\widehat{\boldsymbol{\beta}}$ obtained by equating (4.26) to zero has asymptotic variance given by the inverse of the Godambe information matrix defined by

$$\mathbf{J}_\beta = \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{C}_i^{-1} \mathbf{D}_i.$$

In general, the quasi-score estimator depends on the value of $\boldsymbol{\Sigma}$, and hence requires an estimate of $\boldsymbol{\Sigma}$. However, the quasi-score function (4.26) is $\boldsymbol{\Sigma}$ -insensitive in the sense that the expected $\boldsymbol{\Sigma}$ -derivative of $\boldsymbol{\psi}_\beta(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is zero. As shown by Jørgensen and Knudsen (2004), this implies that $\widehat{\boldsymbol{\beta}}$ varies slowly with $\boldsymbol{\Sigma}$. The dispersion matrix $\boldsymbol{\Sigma}$ may be estimated by means of a bias-corrected Pearson estimating function involving the cross-product matrix of the residuals $\mathbf{Y}_i - \boldsymbol{\mu}_i$; see Holst and Jørgensen (2010), who also discuss the Newton scoring algorithm for this estimation procedure.

An important special case of this setup is obtained when the regression model $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ is defined as follows:

$$g^\top(\boldsymbol{\mu}_i) = \mathbf{x}_i^\top \mathbf{B},$$

where g is a link function mapping $\boldsymbol{\mu}_i$ coordinatewise, such that $g^\top(\boldsymbol{\mu}_i)$ is a $1 \times k$ vector, \mathbf{x}_i is an m -vector of covariates, and \mathbf{B} is an $m \times k$ matrix of regression coefficients. This model may be called a *multivariate generalized linear model*.

In the special case $\boldsymbol{\Sigma} = \mathbf{I}$, the quasi-score estimator for \mathbf{B} corresponds to fitting k separate univariate generalized linear models, which, because of the slow variation of $\widehat{\boldsymbol{\beta}}$ as a function of $\boldsymbol{\Sigma}$, will not be far from the estimator obtained by combining the quasi-score and Pearson estimating functions to estimate \mathbf{B} and $\boldsymbol{\Sigma}$ jointly. In this sense we have obtained a multivariate regression method that is similar to the conventional multivariate linear regression model with normal errors; see, for example, Johnson and Wichern (2007, Chapter 7).

5 Multivariate extreme dispersion models

5.1 Hazard location families and extreme dispersion models

We shall now turn to the third type of multivariate dispersion models, which is a multivariate extension of the extreme dispersion models introduced by Jørgensen

et al. (2010). This class serves for modelling extremes and survival data, and has many analogies with exponential dispersion models.

We first establish some basic results and notation for extreme dispersion models, following Jørgensen et al. (2010) and Rusch (2009). Let $G(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2)$ denote a bivariate survival function for the random vector $\mathbf{X} = (X_1, X_2)^\top$, where G is assumed to be twice continuously differentiable with support containing the origin of \mathbb{R}^2 . Let $H(x_1, x_2) = -\log G(x_1, x_2)$ denote the integrated hazard function, and define the corresponding (vector) hazard function by

$$\mathbf{h}(x_1, x_2) = \begin{bmatrix} h_1(x_1, x_2) \\ h_2(x_1, x_2) \end{bmatrix} = \dot{H}(x_1, x_2). \tag{5.1}$$

As shown by Rusch (2009), we may think of H as an analogue of the CGF, and the derivatives $H^{(j)}(0, 0)$ are analogues of the cumulants. In particular, we define the *rate vector*

$$\mathbf{r}(X_1, X_2) = \mathbf{h}(0, 0) \tag{5.2}$$

(not to be confounded with the deviance residual above) as an analogue of the mean vector, and the *slope matrix*

$$\mathbf{s}(X_1, X_2) = \dot{\mathbf{h}}(0, 0) = \begin{bmatrix} \ddot{H}_{11}(0, 0) & \ddot{H}_{12}(0, 0) \\ \ddot{H}_{21}(0, 0) & \ddot{H}_{22}(0, 0) \end{bmatrix} \tag{5.3}$$

as an analogue of the covariance matrix. It is important to emphasize that not all properties of the mean vector and covariance matrix carry over to the rate vector and slope matrix. We note, for example, that the rate vector is different from the vector of rates $(r(X_1), r(X_2))^\top$, say. Here $r(X_1)$ is defined from the marginal hazard function, which in turn is obtained from the marginal integrated hazard function $H(x_1, -\infty)$, and similarly for $r(X_2)$. We also note that the slope matrix is not necessarily nonnegative definite.

The rate vector and slope matrix satisfy the following min-additive property:

$$\mathbf{r}(\mathbf{X} \wedge \mathbf{Y}) = \mathbf{r}(\mathbf{X}) + r(\mathbf{Y}) \quad \text{and} \quad \mathbf{s}(\mathbf{X} \wedge \mathbf{Y}) = \mathbf{s}(\mathbf{X}) + s(\mathbf{Y})$$

for \mathbf{X} and \mathbf{Y} independent, where \wedge denotes the componentwise minimum. The rate vector and slope matrix also satisfy the following scaling properties:

$$\mathbf{r}(\mathbf{C}\mathbf{X}) = \mathbf{C}^{-1}\mathbf{r}(\mathbf{X}) \quad \text{and} \quad \mathbf{s}(\mathbf{C}\mathbf{X}) = \mathbf{C}^{-1}\mathbf{s}(\mathbf{X})\mathbf{C}^{-1},$$

where \mathbf{C} is a diagonal matrix with positive entries.

From now on we make the additional assumption that \mathbf{h} is injective and that $\dot{\mathbf{h}}(x_1, x_2)$ is regular for all (x_1, x_2) in the interior of the support of (X_1, X_2) , assumed to be a closed convex subset of \mathbb{R}^2 . Following Rusch (2009), we define a *hazard location family* to be a location family parametrized by its rate vector $\boldsymbol{\mu}$, corresponding to the family of integrated hazard functions of the form

$$\mathbf{x} \mapsto H(\mathbf{x} + \mathbf{h}^{-1}(\boldsymbol{\mu})), \tag{5.4}$$

where the rate vector $\boldsymbol{\mu}$ belongs to the set $\Omega \subseteq \mathbb{R}_+^2$, say. This is analogous to the natural exponential family (4.2). The slope matrix for (5.4) may be expressed as a function of the rate vector by means of the *slope function* (similar to the variance function), defined by

$$v(\boldsymbol{\mu}) = \dot{\mathbf{h}} \circ \mathbf{h}^{-1}(\boldsymbol{\mu}) \quad \text{for } \boldsymbol{\mu} \in \Omega.$$

The slope function characterizes the family (5.4) among all hazard location families.

We now define a *min-additive extreme dispersion model* $\text{XD}^*(\boldsymbol{\mu}, \lambda)$ as corresponding to the following family of integrated hazard functions:

$$\mathbf{x} \mapsto \lambda H(\mathbf{x} + \mathbf{h}^{-1}(\boldsymbol{\mu})) \quad (5.5)$$

for $\lambda > 0$ (assuming min-infinite divisibility), which has rate vector $\lambda\boldsymbol{\mu}$ and slope matrix $\lambda v(\boldsymbol{\mu})$. This model satisfies a property of min-additivity,

$$\mathbf{X}_1 \wedge \cdots \wedge \mathbf{X}_n \sim \text{XD}^*(\boldsymbol{\mu}, n\lambda) \quad (5.6)$$

for $\mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. from $\text{XD}^*(\boldsymbol{\mu}, \lambda)$.

Similarly, we define a *reproductive extreme dispersion model* $\text{XD}(\boldsymbol{\mu}, \lambda)$ by applying the duality transformation $\mathbf{Y} = \lambda\mathbf{X}$ to (5.5), which gives the following family of integrated hazard functions:

$$\mathbf{y} \mapsto \lambda H(\mathbf{y}/\lambda + \mathbf{h}^{-1}(\boldsymbol{\mu})), \quad (5.7)$$

which has rate vector $\boldsymbol{\mu}$ and slope matrix $\lambda^{-1}v(\boldsymbol{\mu})$. This model is often parametrized by $\boldsymbol{\mu}$ and the dispersion parameter $\sigma^2 = 1/\lambda$, and we use the notation $\text{XD}(\boldsymbol{\mu}, \sigma^2)$ for the distribution corresponding to (5.7). In the univariate case, families like the Pareto, logistic and extreme value give rise to extreme dispersion models [see Jørgensen et al. (2010)], whereas Rusch (2009) considered several examples of bivariate extreme dispersion models.

Similar to the case of exponential dispersion models, we see that the form of the slope matrix for extreme dispersion models is governed by a single parameter λ , so there is a need to introduce additional parameters in order to obtain a fully flexible structure of the slope matrix.

5.2 The bivariate case

Extending the results of Rusch (2009), we shall now introduce a new class of multivariate extreme dispersion models, which we do by mimicking the construction of multivariate exponential dispersion models in Section 4.1 above.

In order to develop the bivariate case of extreme dispersion models, let us write

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_{11} \\ U_{12} \end{bmatrix} \wedge \begin{bmatrix} U_1 \\ \infty \end{bmatrix} \wedge \begin{bmatrix} \infty \\ U_2 \end{bmatrix}, \quad (5.8)$$

similar to (4.1). If we assume that the three components on the right-hand side of (5.8) are independent, the joint survival function of X_1 and X_2 has the form

$$(x_1, x_2) \mapsto G(x_1, x_2)G_1(x_1)G_2(x_2), \tag{5.9}$$

where $G_1(x_1)$ and $G_2(x_2)$ are the marginal survival functions of U_1 and U_2 , respectively, and $G(x_1, x_2)$ is the joint survival function of U_{11} and U_{12} .

From now on, let H be a bivariate integrated hazard function of the form considered above, and let G in (5.9) correspond to the integrated hazard function defined by

$$(x_1, x_2) \mapsto \lambda_{12}H(x_1 - \theta_1, x_2 - \theta_2), \tag{5.10}$$

where again λ_{12} is called the weight. This corresponds to the extreme dispersion model (5.5), except that it is parametrized by the location parameter (θ_1, θ_2) instead of the rate vector μ .

We now look at the marginal integrated hazard functions corresponding to $H(x_1 - \theta_1, x_2 - \theta_2)$, namely, $H(x_1 - \theta_1, -\infty)$ and $H(-\infty, x_2 - \theta_2)$, respectively. We hence let $G_1(x_1)$ correspond to the integrated hazard function with positive weight λ_1 given by

$$x_1 \mapsto \lambda_1 H(x_1 - \theta_1, -\infty). \tag{5.11}$$

Similarly, we let $G_2(x_2)$ correspond to the integrated hazard function with positive weight λ_2 given by

$$x_2 \mapsto \lambda_2 H(-\infty, x_2 - \theta_2). \tag{5.12}$$

We now add the three terms (5.10), (5.11) and (5.12), which give the following bivariate integrated hazard function for the random vector (X_1, X_2) ,

$$\begin{aligned} H_{\theta, \lambda}(x_1, x_2) &= \lambda_{12}H(x_1 - \theta_1, x_2 - \theta_2) \\ &+ \lambda_1 H(x_1 - \theta_1, -\infty) + \lambda_2 H(-\infty, x_2 - \theta_2). \end{aligned} \tag{5.13}$$

We note that the marginal distribution of X_1 has the same form as (5.11), but with λ_1 replaced by $\lambda_{11} = \lambda_{12} + \lambda_1$, such that

$$H_{\theta, \lambda}(x, -\infty) = \lambda_{11}H(x - \theta_1, -\infty) \quad \text{and} \quad H_{\theta, \lambda}(-\infty, y) = \lambda_{22}H(-\infty, y - \theta_2),$$

and similarly for the marginal distribution of X_2 , where λ_2 is replaced by $\lambda_{22} = \lambda_{12} + \lambda_2$. We call the model corresponding to (5.13) a *min-additive bivariate extreme dispersion model*; see (5.14) below. In the special case $\lambda_1 = \lambda_2 = 0$, we recover the additive extreme dispersion model (5.5) from above, whereas the special case $\lambda_{12} = 0$ yields components X_1 and X_2 that are independent.

We have hence achieved the goal of defining a five-parameter extension of the min-additive extreme dispersion model (5.5) in such a way that the form of the marginal distributions has been preserved. It is not possible, however, to parametrize the family (5.13) by the slope vector in the same way that we

parametrized the exponential dispersion model above by its mean vector. In fact, the slope vector takes the form

$$\mathbf{r}(X_1, X_2) = \begin{bmatrix} \lambda_{12} \dot{H}_1(-\theta_1, -\theta_2) + \lambda_1 \dot{H}_1(-\theta_1, -\infty) \\ \lambda_{12} \dot{H}_2(-\theta_1, -\theta_2) + \lambda_2 \dot{H}_2(-\infty, -\theta_2) \end{bmatrix},$$

whose components do not factorize in the same way as the mean vector (4.14). Similarly, the form of the slope matrix is more complicated than the bivariate variance function (4.15).

We shall denote the min-additive bivariate extreme dispersion model (5.13) by $\text{XD}_2^*(\boldsymbol{\theta}, \boldsymbol{\Lambda})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ is the location parameter and $\boldsymbol{\Lambda}$ is the matrix defined by (4.16). This model satisfies the following min-reproductive property:

$$\text{XD}_2^*(\boldsymbol{\theta}, \boldsymbol{\Lambda}_1) \wedge \text{XD}_2^*(\boldsymbol{\theta}, \boldsymbol{\Lambda}_2) = \text{XD}_2^*(\boldsymbol{\theta}, \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2), \tag{5.14}$$

similar to (5.6). By the duality transformation

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \lambda_{11} X_1 \\ \lambda_{22} X_2 \end{bmatrix},$$

we obtain the reproductive form $\mathbf{Y} \sim \text{XD}_2(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ of the bivariate extreme dispersion model, where $\boldsymbol{\Sigma}$ is defined by (4.21).

This construction may be generalized to the multivariate case in much the same way as we saw for exponential dispersion models in Section 4.4.

5.3 A bivariate logistic distribution

Following Rusch (2009, page 47), we consider the bivariate logistic distribution with integrated hazard function

$$H(x_1, x_2) = \log(1 + e^{x_1} + e^{x_2}) \quad \text{for } (x_1, x_2) \in \mathbb{R}^2, \tag{5.15}$$

and hazard function

$$\mathbf{h}(x_1, x_2) = (1 + e^{x_1} + e^{x_2})^{-1} \begin{bmatrix} e^{x_1} \\ e^{x_2} \end{bmatrix} \quad \text{for } (x_1, x_2) \in \mathbb{R}^2.$$

We now generate a hazard location model from (5.15), where the rate vector $\boldsymbol{\mu} = \mathbf{h}(x_1, x_2)$ has domain $\Omega = \{\boldsymbol{\mu} \in \mathbb{R}_+^2 : \mu_1 + \mu_2 < 1\}$. The slope matrix has the following form:

$$\mathbf{v}(\boldsymbol{\mu}) = \begin{bmatrix} \mu_1(1 - \mu_1) & -\mu_1\mu_2 \\ -\mu_1\mu_2 & \mu_2(1 - \mu_2) \end{bmatrix},$$

similar to the covariance matrix of the components of a multinomial distribution.

If we now generate a bivariate extreme dispersion model from (5.15), the integrated hazard function $H_{\theta, \lambda}(x_1, x_2)$ becomes

$$\begin{aligned} H_{\theta, \lambda}(x_1, x_2) &= \lambda_{12} \log(1 + e^{x_1 - \theta_1} + e^{x_2 - \theta_2}) \\ &\quad + \lambda_1 \log(1 + e^{x_1 - \theta_1}) + \lambda_2 \log(1 + e^{x_2 - \theta_2}) \end{aligned}$$

for $(x_1, x_2) \in \mathbb{R}^2$. This defines the five-parameter bivariate logistic extreme dispersion model $\text{Lg}_2(\boldsymbol{\theta}, \boldsymbol{\Lambda})$. A multivariate version may be obtained by utilizing the results of Rusch (2009, page 47).

5.4 A bivariate Gumbel distribution

In the construction of a bivariate Gumbel distribution we use a technique similar to the construction of the bivariate gamma distribution via (4.24), except that we use location parameters instead of scale parameters. Hence, let U denote a standard Gumbel variable with integrated hazard function e^x . Consider the random vector

$$\begin{bmatrix} U + \theta_1 \\ U + \theta_2 \end{bmatrix}, \quad (5.16)$$

where $\theta_1, \theta_2 \in \mathbb{R}$. The integrated hazard function for (5.16) is

$$H(x_1, x_2) = e^{\max\{x_1 - \theta_1, x_2 - \theta_2\}} \quad \text{for } (x_1, x_2) \in \mathbb{R}^2.$$

Let us generate a bivariate extreme dispersion model of the form (5.8) with integrated hazard function

$$H_{\theta, \lambda}(x_1, x_2) = \lambda_{12} e^{\max\{x_1 - \theta_1, x_2 - \theta_2\}} + \lambda_1 e^{x_1 - \theta_1} + \lambda_2 e^{x_2 - \theta_2} \quad (5.17)$$

for $(x_1, x_2) \in \mathbb{R}^2$. In this case we interpolate between independence ($\lambda_{12} = 0$) and complete dependence ($\lambda_1 = \lambda_2 = 0$). As it happens, the parameter (θ_1, λ_1) is not identifiable from (5.17), but $\lambda_1 e^{\theta_1}$ is, and, similarly, only $\lambda_2 e^{\theta_2}$ is identifiable from (θ_2, λ_2) .

The problem of identifiability, however, disappears when we transform to the reproductive form by the duality transformation. This gives a bivariate Gumbel extreme dispersion model defined by the survival function

$$\begin{aligned} \overline{H}_{\theta, \lambda}(y_1, y_2) &= \lambda_{12} \exp \max\{y_1/\lambda_{11} - \theta_1, y_2/\lambda_{22} - \theta_2\} \\ &\quad + \lambda_1 e^{y_1/\lambda_{11} - \theta_1} + \lambda_2 e^{y_2/\lambda_{22} - \theta_2}, \end{aligned} \quad (5.18)$$

where all five parameters are identifiable. This bivariate Gumbel distribution is different from the three conventional types of bivariate Gumbel distributions; cf. Balakrishnan and Lai (2009). Similar to the Marshall and Olkin (1967) bivariate exponential distribution, the distribution (5.18) has a singular part, concentrated on the straight line

$$y_2 = \lambda_{22}(\theta_2 - \theta_1 + y_1/\lambda_{11}).$$

A multivariate Gumbel distribution may be obtained by methods similar to those of Section 4.4.

6 Discussion

In order to develop fully flexible multivariate dispersion models, we have reviewed an existing method for constructing multivariate proper dispersion models, and we have introduced new methods for constructing multivariate exponential and extreme dispersion models. These different types of dispersion models may seem

rather disparate at first, but the common form of the variance function in the two first cases suggests that multivariate dispersion models provide a general and flexible framework for constructing new multivariate distributions.

Such a framework can accommodate a wide variety of different types of data, including, for example, multivariate exponential dispersion models for discrete data. The common interpretation of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ leads to a unified methodology for statistical analysis of multivariate data. In addition to the three types of dispersion models discussed here, a fourth type has recently been proposed, namely, the class of geometric dispersion models (Jørgensen and Kokonendji, 2011), but a possible multivariate generalization of this class remains to be explored.

On this background we propose a program for the systematic development of multivariate dispersion models, in order to break a path through the *jungle of distributions* in \mathbb{R}^k , in the words of Letac (2007); see the website Jørgensen (2011a) for details. This program will require a concerted effort on many different fronts, ranging from the practical implementation of simulation and estimation methods to the development of specialized models for longitudinal, spatial and other forms of correlated data. Short of being all-encompassing, such a program holds the promise of providing a good general methodology for modelling multivariate nonnormal data, much like generalized linear models do in the univariate case.

Acknowledgments

This research was supported by the Danish Natural Science Research Council. I am grateful to Yuri Goegebeur for useful comments in relation to the paper.

References

- Anderson, T. W. and Fang, K.-T. (1987). Cochran's theorem for elliptically contoured distributions. *Sankhyā, Ser. A* **49**, 305–315. [MR1056023](#)
- Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate unified skew-elliptical distributions. *Chilean J. Statist.* **1**, 17–33. [MR2756082](#)
- Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distributions: Theory and Methods*, 2nd ed. New York: Springer. [MR2840643](#)
- Barndorff-Nielsen, O. E. (1983). On a formula for the conditional distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365. [MR0712023](#)
- Barndorff-Nielsen, O. E. and Blæsild, P. (1987). Hyperbolic distributions. In *Encyclopedia of Statistical Sciences, Vol. 3* (S. Kotz, C. B. Read and D. L. Banks, eds.) 700–707. New York: Wiley.
- Bernardoff, P., Kokonendji, C. and Puig, B. (2008). Generalized variance estimators in the multivariate gamma models. *Mathematical Methods of Statistics* **17**, 66–73. [MR2400365](#)
- Díaz-García, J. A. and Domínguez-Molina, J. R. (2006). Some generalisations of Birnbaum–Saunders and sinh-normal distributions. *International Mathematical Forum* **35**, 1709–1727. [MR2275956](#)
- Fang, K.-T. (1997). Elliptically contoured distributions. In *Encyclopedia of Statistical Sciences Update, Vol. 1* (S. Kotz, C. B. Read and D. L. Banks, eds.) 212–218. New York: Wiley.

- Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance: Mathematics and Economics* **46**, 351–361. [MR2656337](#)
- Holst, R. and Jørgensen, B. (2010). Efficient and robust estimation for a class of generalized linear longitudinal mixed models. Available at arXiv:1008.2870v1.
- Jensen, D. R. (1985). Multivariate distributions. In *Encyclopedia of Statistical Sciences, Vol. 6* (S. Kotz, C. B. Read and D. L. Banks, eds.) 43–55. New York: Wiley.
- Joe, H. (1996). Families of m -variate distributions with given margins and $m(m - 1)/2$ bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics. Lecture Notes—Monograph Series* **28**, 120–141. Beachwood, OH: IMS. [MR1485527](#)
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, NJ: Pearson Education International. [MR2372475](#)
- Jørgensen, B. (1987a). Exponential dispersion models (with discussion). *J. Roy. Statist. Soc. Ser. B* **49**, 127–162. [MR0905186](#)
- Jørgensen, B. (1987b). Small-dispersion asymptotics. *Braz. J. Probab. Statist.* **1**, 59–90. [MR0932188](#)
- Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall. [MR1462891](#)
- Jørgensen, B. (1999). Dispersion Models. In *Encyclopedia of Statistical Sciences, update, Vol. 3* (S. Kotz, C. B. Read and D. L. Banks, eds.) 172–184. New York: Wiley. [MR1679440](#)
- Jørgensen, B. (2011a). Multivariate dispersion models. Available at <http://imada.sdu.dk/~bentj/>.
- Jørgensen, B. (2011b). Multivariate exponential dispersion models. Unpublished manuscript.
- Jørgensen, B., Goegebeur, Y. and Martínez, J. R. (2010). Dispersion models for extremes. *Extremes* **13**, 399–437. [MR2733940](#)
- Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.* **31**, 93–114. [MR2042601](#)
- Jørgensen, B. and Kokonendji, C. C. (2011). Dispersion models for geometric sums. *Brazilian J. Probab. Statist.* **25**, 263–293.
- Jørgensen, B. and Lauritzen, S. L. (2000). Multivariate dispersion models. *J. Multivariate Anal.* **74**, 267–281. [MR1794004](#)
- Jørgensen, B. and Rajeswaran, J. (2005). A generalization of Hotelling's T^2 . *Communications in Statistics—Theory and Methods* **34**, 2179–2195. [MR2204974](#)
- Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, 2nd ed. New York: Wiley.
- Kundu, D., Balakrishnan, N. and Jamalizadeh, A. (2010). Bivariate Birnbaum–Saunders distribution and associated inference. *J. Multivariate Anal.* **101**, 113–125. [MR2557622](#)
- Letac, G. (2007). The bivariate gamma distribution with Laplace transform $(1 + as + bt + cst)^{-q}$: History, characterizations, estimation. Available at <http://www.math.univ-toulouse.fr/~letac/bivgamma.pdf>.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **542**, 121–130.
- Marshall, A. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30–44. [MR0215400](#)
- Mathai, A. M. and Moschopoulos, P. G. (1991). On a multivariate gamma. *J. Multivariate Anal.* **39**, 135–153. [MR1128677](#)
- Rusch, E. (2009). Multivariate extreme dispersion models. M.Sc. thesis, Univ. Copenhagen.
- Sweeting, T. J. (1987). Approximate Bayesian analysis of censored survival data. *Biometrika* **74**, 809–816. [MR0919849](#)
- Wicksell, S. D. (1933). On correlation functions of type III. *Biometrika* **25**, 121–133.

Department of Mathematics and Computer Science
 University of Southern Denmark
 Campusvej 55
 DK-5230 Odense M
 Denmark
 E-mail: bentj@stat.sdu.dk