Modelling Directional Dispersion Through Hyperspherical Log-Splines

José T.A.S. Ferreira and Mark F.J. Steel*

Department of Statistics
University of Warwick, UK

Abstract

We introduce the directionally dispersed class of multivariate distributions, a generalisation of the elliptical class. By allowing dispersion of multivariate random variables to vary with direction it is possible to generate a very wide and flexible class of distributions. Directionally dispersed distributions have a simple form for their density, which extends a spherically symmetric density function by including a function D modelling directional dispersion. Under a mild condition, the class of distributions is shown to preserve both unimodality and moment existence. By adequately defining D, it is possible to generate skewed distributions. Using spline models on hyperspheres, we suggest a very flexible, yet practical, implementation for modelling directional dispersion in any dimension. Finally, we use the new class of distributions in a Bayesian regression setup and analyse the distributions of a set of biomedical measurements and a sample of U.S. manufacturing firms.

Keywords: Bayesian regression model, directional dispersion, elliptical distributions, existence of moments, modality, skewed distributions.

1 Introduction

Spherical distributions, in the sense of Fang *et al.* (1990), can be characterised by having constant dispersion along every direction. In this article we introduce distributions that reflect the situation where dispersion can vary with direction.

A continuous random variable $x \in \Re^m$ is said to follow a spherical distribution around zero with density f if the latter is of the form

$$f(x) = g^{(m)}\left(x'x\right), \tag{1}$$

where $g^{(m)}$ is a m-dimensional density generator or labelling function, which satisfies (2.19) in Fang *et al.* (1990). From (1) we see that the density depends on x only through its l_2 norm, denoted by $\|x\|$. The distributions that we introduce here are generated by densities that depend on both norm and direction of x through the ratio $\|x\|^2/D(d_x)$, where d_x denotes the direction of x and D is a function taking positive values. Function D determines directional dispersion, *i.e* the dispersion of the distribution along each direction and, as will be seen in the sequel, by controlling D it is possible to generate a very rich and flexible class of distributions, which we call the directionally dispersed class, abbreviated to D-class.

* *Address for correspondence*: Mark Steel, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. Tel.: +44-24-7652 3369; Fax: +44-24-7652 4532; Email M.F.Steel@stats.warwick.ac.uk.

The simplicity of directionally dispersed distributions, in combination with their great flexibility makes them quite appealing for use in applications. Each member of the D-class can be defined by choosing three elements with unequivocal roles: the density generator of a spherical distribution, the centre of the distribution, and D. The density generator determines how the distribution tails off, the centre of the distribution determines location, and D determines directional dispersion. In addition, each one of these elements can be chosen independently of the others.

Among the characteristics of directionally dispersed distributions, we highlight the following facts: if the density generator is decreasing in its argument, then, under rather general conditions on D, the directionally dispersed distribution is unimodal; moment existence is unaffected by passing from the corresponding spherical distribution to our more general setup; the D-class is closed under linear transformations.

Recent years have seen an increased interest in flexible classes of distributions that can model distributional asymmetry (see Genton 2004 for a review). By choosing D such that there is a set of directions of positive measure on the unit sphere for which $D(d_x) \neq D(d_{-x})$, we obtain a skewed distribution. By choosing a flexible form for D, we generate a wide class of unimodal skewed distributions. Nevertheless, by imposing that $D(d_x) = D(d_{-x})$ for all $x \in \Re$, we can still obtain symmetric distributions in the sense that x and $-x$ have the same distribution. In particular, Section 2 shows that the D-class generalises both the spherical and elliptical (Kelker 1970) classes of continuous distributions.

The class of distributions that we introduce here covers the very general class of v-spherical distributions of Fernández *et al.* (1995). The latter paper defined distributions generated by densities with fixed shapes of the isodensity sets. In our notation, the isodensity sets will be parameterised by the function D and the choice of the function v in Fernández *et al.* (1995) is here replaced by the choice of D. In addition, the focus of the present paper is quite different. Fernández *et al.* (1995) did not provide any operational guidelines for exploiting the generality of their class as a modelling tool. Here we are primarily interested in conducting inference on the shape of the isodensity sets and provide a very flexible modelling framework for multivariate data.

In this article we do not focus on particular forms for the directional dispersion function. Instead, we introduce a general methodology that allows us to model dispersion in an almost unrestricted manner. The set of all directions in $\Re^m$ can be parameterised through its unit hypersphere $\mathcal{S}^{m-1}$. This suggests constructing D using functions defined on hyperspheres. In particular, we use hyperspherical splines (Wahba 1990, Taijeron *et al.* 1994) to model the logarithm of D. These have the advantage of being intuitive and well-defined for any dimension m, and their smoothness can be controlled directly.

We then propose a Bayesian regression model using the D-class where D is modelled with splines. In particular, we derive a prior structure on D which is invariant under linear orthogonal transformations. Inference under these models must be carried out using numerical methods and we introduce a reversible jump Markov chain Monte Carlo (MCMC) algorithm (Green 1995) to implement this. Finally, we provide two illustrations. In the first, the distribution of biomedical data is analysed, comparing both symmetric and skewed directionally dispersed distributions with their elliptical counterpart and a skewed distribution defined as in Ferreira and Steel (2004a). In the second application, we study the size distribution of a set of firms, using the D-class with two different density generators.

Section 2 introduces the D-class of distributions. Section 3 deals with modelling function D, in particular using hyperspherical log-splines. In Section 4, we introduce the Bayesian regression model. Section 5 presents some details on conducting inference and Section 6 presents the applications.

Finally, Section 7 provides some concluding remarks. All proofs are deferred to the Appendix.

2 The D-Class of Distributions

The most common generalisation of the class of spherical distributions is the elliptical class, generated by affine linear transformations. Let Σ be an $m \times m$ covariance matrix, x have a spherical distribution centred at zero with density (1) and $\mu_y \in \Re^m$. Then, $y = \Sigma^{1/2}x + \mu_y$ has an elliptical distribution with density given by

$$f_m(y|\mu_y, \Sigma) = |\Sigma|^{-1/2} g^{(m)} \left[(y-\mu_y)'\Sigma^{-1}(y-\mu_y)\right]. \tag{2}$$

Now consider $d_{y-\mu_y} = \frac{y-\mu_y}{\|y-\mu_y\|}$. The direction of $y - \mu_y$ is given by $d_{y-\mu_y} \in \mathcal{S}^{m-1}$, the unit sphere in $\Re^m$. It is immediate that (2) can be written as

$$f_m(y|\mu_y, \Sigma) = |\Sigma|^{-1/2} g^{(m)} \left[\frac{(y-\mu_y)'(y-\mu_y)}{D(d_{y-\mu_y})}\right], \tag{3}$$

where

$$D(d_{y-\mu_y}) = \left(d'_{y-\mu_y}\Sigma^{-1}d_{y-\mu_y}\right)^{-1}. \tag{4}$$

The formulation of $f_m(y|\mu_y, \Sigma)$ given in (3) allows a clearer comparison between the spherical and the elliptical classes. For the former, the l_2 norm of $y - \mu_y$ is all that is needed to calculate the density values. In contrast, for the elliptical case, the direction of $y - \mu_y$ also plays a part. The function D, with its directional argument can be seen as modelling the dispersion of $y-\mu_y$ along $d_{y-\mu_y}$, with larger values of D implying larger dispersion. Henceforth, we will call $D(d_{y-\mu_y})$ the "directional dispersion function" of $y - \mu_y$.

Note that for $\Sigma = \sigma^2 I_m$, with $\sigma^2 > 0$ and I_m the m-dimensional identity matrix, we obtain a spherical distribution for y centred at μ_y and $D(d_{y-\mu_y}) = \sigma^2$. Thus, spherical densities are characterised by having a constant dispersion function, and they provide a natural benchmark for our class.

Despite the increased flexibility provided by the elliptical class, imposing a specific, rigid form for D does not make use of the full potential of directional dispersion. The class of distributions that we present here (the D-class) is based on dropping the restriction that the dispersion function D is of the form given in (4), thus generating a much broader class of distributions, with densities of the form

$$f_m(y|\mu_y, g^{(m)}, D) = \frac{1}{K_D} g^{(m)} \left[\frac{(y-\mu_y)'(y-\mu_y)}{D(d_{y-\mu_y})}\right], \tag{5}$$

where D is a function from $\mathcal{S}^{m-1}$ to $\Re_+$ and K_D is a constant. Alternatively, we shall denote this as $y \sim DC_m(\mu_y, g^{(m)}, D)$. The distributions in the D-class are then indexed by three elements: the location μ_y, the density generator $g^{(m)}$ and the directional dispersion function D.

The class defined here is exactly as rich as the flexible v-spherical class of Fernández *et al.* (1995), but has a different and more operational parameterisation. The v-spherical class is defined through the density function $f_m(y|\mu_y, \tau) = \tau^m g[v\{\tau(y-\mu_y)\}]$, where v is a scalar function such that $v > 0$ (with the exception of measure zero sets) and $v(ky) = kv(y)$ for any $k \geq 0$, $g(r) \propto g^{(m)}(r^2), r \geq 0$ and τ is a scale parameter. The function v determines the shape of the isodensity sets. Thus, in this v-spherical framework, (5) corresponds to choosing $v(x) = ||x||/\{\tau D^{\frac{1}{2}}(d_x)\}$. In our parameterisation τ is superfluous. However, there are important differences in that Fernández *et al.* (1995) focuses on fixing v, while varying the density generator $g^{(m)}$ (see *e.g.* their Theorem 1 on inference robustness).

In contrast, the present paper is mainly about conducting inference on the shape of the isodensity sets (*i.e.* on D) and is particularly focused on providing an operational methodology that allows us to use the full flexibility of the D-class in practical modelling of data. The difference in parameterisation adds interpretability and facilitates implementation, as we shall see in the sequel.

Using a transformation to polar coordinates, the following result states the necessary and sufficient condition on D so that (5) is a probability density function.

Theorem 1. Let $y - \mu_y = \rho\, t(\omega)$, where $\rho \in \Re^+$, $\omega = (\omega_1, \ldots, \omega_{m-1}) \in \Omega = [0, \pi)^{m-2} \times [0, 2\pi)$ and $t(\omega) = [t_1(\omega), \ldots, t_m(\omega)]'$ with

$$t_j(\omega) = \left(\prod_{i=1}^{j-1} \sin \omega_i \right) \cos \omega_j, \quad 1 \leq j \leq m-1 \quad \text{and} \quad t_m(\omega) = \prod_{i=1}^{m-1} \sin \omega_i.$$

Then, $D[t(\omega)]$ is a suitable directional dispersion function, for any $\mu_y \in \Re^m$ and density generator $g^{(m)}$, if and only if the integral $\int_\Omega \left(\prod_{i=1}^{m-2} \sin^{m-1-i} \omega_i \right) D[t(\omega)]^{m/2} d\omega$ exists.

Given the form of the integral in Theorem 1, the following result holds.

Corollary 1. If there is a finite constant C such that $D < C$ for any $d_{y-\mu_y} \in \mathcal{S}^{m-1}$, then $f_m(y|\mu_y, g^{(m)}, D)$ defined by (5) is a density function in $\Re^m$.

The restrictions in Corollary 1 do not exclude many situations of practical interest. Bounding D away from infinity implies that the distance between any two isodensity sets is always finite.

The definition of the density function in (5) is completed by calculating the integrating constant.

Corollary 2. If D is a directional dispersion function satisfying the condition of Theorem 1, then

$$K_D = \frac{\Gamma(m/2)}{2\pi^{m/2}} \int_\Omega \left(\prod_{i=1}^{m-2} \sin^{m-1-i} \omega_i \right) D[t(\omega)]^{m/2} d\omega.$$

One feature of the results in Theorem 1 and Corollaries 1 and 2 is that they do not depend on either location or density generator. As such, if the directional dispersion function D is valid for a particular choice of location and, more importantly, density generator then it is also valid for any other choice of location and density generator. Further, the integrating constant K_D only needs to be calculated once.

By appropriately selecting D, it is possible to obtain very different shapes for the directionally dispersed distributions. Contour plots of four bivariate densities are presented in Figure 1. Plot (a) represents an elliptical distribution centred at zero and with Σ having diagonal elements equal to two and off diagonal equal to one, obtained by selecting D as in (4). The departure of the densities depicted in plots (b)-(d) from elliptical densities is evident. In Figure 1(b) the contours are not concentric, in contrast to elliptical distributions. The continuous density represented in Figure 1(c) is clearly made up of the distinct parts. For $y_2 > 0$ the contours are l_1-spherical (as defined in Osiewalski and Steel, 1993), while for negative y_2 the contours are spherical. Finally, Figure 1(d) depicts a rather complicated density which is relatively simple to obtain in our setup.

The directional dispersion functions of the densities presented in Figure 1 are plotted in Figure 2, using the parameterisation $d_{y-\mu_y} = [\cos \omega, \sin \omega]'$, for $\omega = [0, 2\pi)$, spanning the complete set of

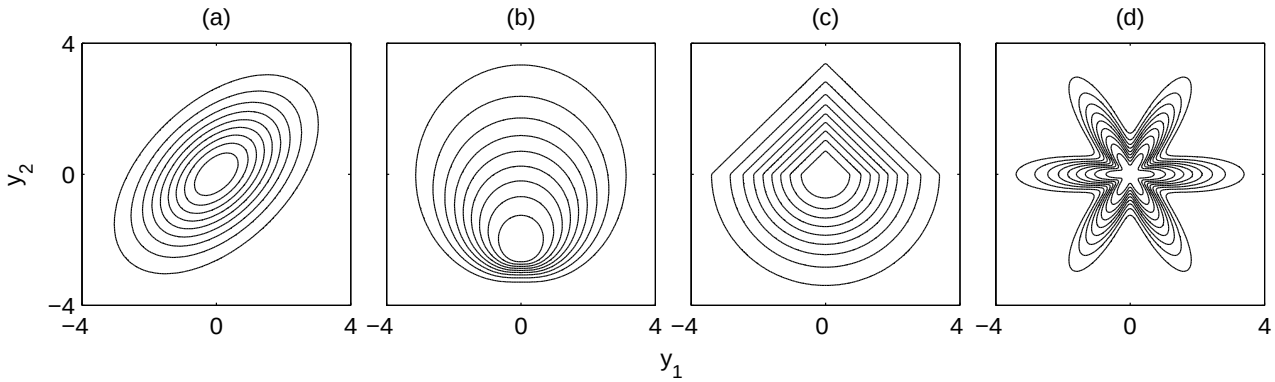

Figure 1: Examples of four two dimensional members of the directional dispersed class.

directions in $\Re^2$. The symmetric pattern of the densities in Figures 1(a) and (d) is also present in the directional dispersion functions, represented with solid and dot-dashed lines respectively. While symmetry is characteristic of elliptical densities, the contours in Figures 1(d) illustrate that it can also be carried over to a more general set of directionally dispersed distributions. Generally, symmetry is induced by choosing D such that for all $y \in \Re^m$, $D(d_{y-\mu_y}) = D(d_{\mu_y-y})$. The dotted and dashed lines in Figure 2 represent D for the densities with contours in Figures 1(b) and (c), respectively. These densities generate skewed distributions, and illustrate the variety of forms that can be generated by modelling D. The function plotted by the dotted curve in Figure 2 has the property that $D(d_{\mu_y-y}) = D(d_{y-\mu_y})^{-1}$, which causes uncentred contours. Finally, the piecewise form of the dashed curve in Figure 2 determines the split in shape represented in Figure 1(c).

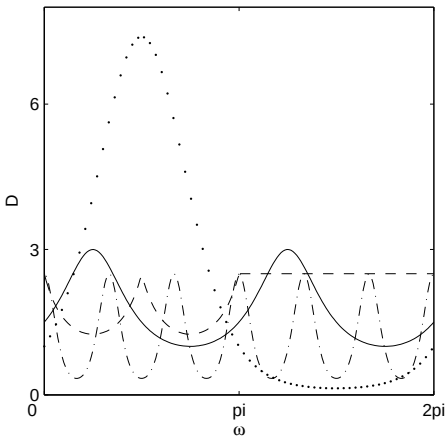

Figure 2: Directional dispersion as a function of $d_{y-\mu_y} = [\cos\omega, \sin\omega]'$ for the densities presented in Figure 1(a) (solid), (b) (dotted), (c) (dashed) and (d) (dot-dashed).

Parameter μ_y determines the location or centre of directionally dispersed distributions. The centre of a distribution is particularly relevant when the distribution is unimodal. The following result holds for members of the D-class.

Theorem 2. Let $g^{(m)}$ be a decreasing function. Then, for any D meeting the condition of Corollary 1, the distribution with density $f_m(y|\mu_y, g^{(m)}, D)$ as in (5) is unimodal and the mode is at μ_y.

Thus, if the density generator $g^{(m)}$ corresponds to a unimodal spherical distribution in (1) with mode at zero, then it will also generate a unimodal distribution through (5) with mode at μ_y. Ensuring unimodality is often of primary importance in the context of applications and Theorem 2 states that the general framework provided by varying (bounded) D does not affect this property.

The density generator also determines the rate at which the tails of the distribution taper off. For the D-class, the role played by $g^{(m)}$ is the same as for spherical distributions. The only difference is that scale is now a function of direction. As $D(d)$ gets larger the distance between the isodensity sets is increased with respect to the one for the spherical distribution defined by constant D. By controlling tail behaviour, the density generator also controls moment existence. Within the entire D-class, $g^{(m)}$ alone determines the moment properties, as the following result states.

Theorem 3. Let $y \sim DC_m(\mu_y, g^{(m)}, D)$ defined by (5) with any D meeting the condition of Corollary 1. Then, existence of moments of y will only depend on $g^{(m)}$.

The result of Theorem 3 ensures us that the results on moment existence that hold for spherical and elliptical distributions are immediately extended to the whole D-class with bounded dispersion functions.

Let us finally explore the properties of the D-class under linear transformations.

Theorem 4. Let $y \sim DC_m(0, g^{(m)}, D)$ and consider the linear transformation $z = \mu + A'y$, where $\mu \in \Re^m$ and A is a nonsingular matrix, and define $A'A = \Omega$. Then, $z \sim DC_m(\mu, g^{(m)}, D_z)$ with

$$D_z(d) = (d'\Omega^{-1}d)^{-1} D\left[(d'\Omega^{-1}d)^{-\frac{1}{2}} A'^{-1} d\right] \tag{6}$$

for any $d \in \mathcal{S}^{m-1}$.

Thus, the D-class is closed under affine linear transformations and such transformations do not affect the tail behaviour or unimodality (from Theorem 2 if we assume that D is bounded). Theorem 4 immediately generalizes the well-known result mentioned in the beginning of this Subsection that whenever y is spherical (*i.e.* $D(d)$ is constant), we generate an elliptical z by linear transformations. It also illustrates that if y is elliptical with scale matrix B^{-1}, z will also be elliptical, but with scale matrix $A'B^{-1}A$. In addition, it leads to the following corollary:

Corollary 3. Let $y \sim DC_m(0, g^{(m)}, D)$. Only if D takes the form $D(d) = (d'Bd)^{-1}$ for some nonsingular matrix B, can linear transformations of y be spherically distributed around zero. These linear transformations are given by $z = A'y$ where $AA' = B$.

Several proposals for multivariate skewed distributions have recently been proposed in the literature. However, up to our knowledge none of the suggested methods explicitly considers the directional nature of skewed data. Due to its direct modelling of directional dispersion, the D-class has great potential in the modelling of skewed data. Ferreira and Steel (2004b) proposed methods for describing multivariate skewness with directional skewness measures, where directional skewness is defined along a direction. Quantifying such directional skewness is straightforward in the D-class.

3 Modelling Directional Dispersion

In order to take advantage of the D-class it is necessary to develop forms for the directional distribution function D that can adapt to the particular application. There are two potential avenues for tackling this problem. The first is to develop forms for D to suit specific situations. This is for example the case of D in (4), corresponding to elliptical distributions or the D implicit in Example 1 of Fernández *et al.* (1995). The second is to develop flexible forms that can adapt to a large number of circumstances. Even though the first approach gives us more control over D, flexible families of directional distribution functions are perhaps more appealing in practical modelling situations. Thus, this paper will only focus on the latter alternative, where we wish to conduct inference on D within a wide class, rather than prespecify it in a much more rigid framework.

In the sequel, we always assume that D is bounded (as in Corollary 1). This is likely to be true for cases of practical relevance. As D is strictly positive, we model its logarithm, denoted by $\log D$.

The definition of functions on hyperspheres has a large tradition, especially for the two- and three-dimensional cases. There are several alternatives present in the literature of which we mention hyperspherical harmonics (Müller 1966) and hyperspherical splines (Taijeron *et al.* 1994).

We model $\log D$ using spline models or, equivalently, we model D with hyperspherical log-splines. The simplicity of hyperspherical splines is very appealing. As will be seen below, hyperspherical splines are constructed similarly to splines in real spaces. Despite their simplicity, hyperspherical splines are very flexible and can well approximate any smooth function. As with splines on real spaces, it is possible to control the smoothness of a hyperspherical spline. This characteristic is important to us as we would generally like D to be smooth. Taijeron *et al.* (1994) compare hyperspherical splines with hyperspherical harmonics and find the former numerically more stable. Hyperspherical harmonics also exhibit a wiggly behaviour that we do not expect to be present in most applications.

Applying hyperspherical spline models in our context, we assume that $\log D$ is given by

$$\log D(d) = c_0 + \sum_{k=1}^{K} c_i R_m^l(d|d_k), \tag{7}$$

where K is a positive integer, $c = (c_0, c_1, \ldots, c_K)' \in \Re^{K+1}$, $\mathcal{B}_D = \{d_1, \ldots, d_K\}$ is a set of vectors in $\mathcal{S}^{m-1}$ and $R_m^l(d|d_k) \in \mathcal{C}(l)$, the set of functions with l continuous derivatives, are real valued spline basis functions defined on $\mathcal{S}^{m-1}$ for $k = 1, \ldots, K$. Vectors $d_k,\ k = 1, \ldots, K$, are usually called knots. As the function in (7) is a linear combination of $\mathcal{C}(l)$ basis functions, it is also in $\mathcal{C}(l)$. By choosing an appropriate l, it is therefore possible to control the smoothness of D.

Basis functions for spline interpolation and smoothing on the circle ($m = 2$) and the sphere ($m = 3$) were determined in Wahba (1975) and Wahba (1981), respectively. Taijeron *et al.* (1994) generalised the above results and provided the solution for the problems in any dimension $m > 1$. In the form of (7), the solution is provided when $R_m^l(d|d_k),\ k = 1, \ldots, K$, represent a family of reproducing kernels related to Green's function for the Laplacian on the hyperspheres. For $m \neq 2$, closed forms for $R_m^l(d|d_k)$ are not known, leading to the use of approximating reproducing kernels for which a closed form is available. For our modelling of D the use of the approximating reproducing kernels does not lead to any modelling restrictions. Thus, we will also use the notation $R_m^l(d|d_k)$ for the approximating kernel.

For the case in S^1 ($m = 2$), the reproducing kernel is given by

$$R_2^l(d|d_k) = \frac{(-1)^{l/2}(2\pi)^{l-1}}{(2\pi)!} B_{l+2}\left(\frac{\theta}{2\pi}\right),$$

with θ the angle between d and d_k and B_{l+2} the Bernoulli polynomial of degree $l+2$ (Abramowitz and Stegun 1972).

For $m > 2$, Taijeron *et al.* (1994) indicates that the approximating reproducing kernels can be defined by,

$$R_m^l(d|d_k) = \frac{\Gamma(m/2)}{m-2}\left[\frac{q(l, m-2, d'd_k)}{l!} - \frac{1}{(l+1)!}\right],$$

with

$$q(r, s, z) = \int_0^1 (1-h)^r (1-2hz+h^2)^{-s/2} dh.$$

For $l = 0, 1, \ldots$ and $m > 2$, closed forms for R_m^l are available and can be calculated easily with a symbolic mathematics package. Examples can be found in Wahba (1981) and Taijeron *et al.* (1994).

Once l corresponding to the smoothness of $\log D$ has been specified, an interpolating spline (7) can then be characterised by $K, \mathcal{B}_D$ and c. Given a vector of values $v = (v_1, \ldots, v_K)' \in \Re^K$ associated with the knots, c can be obtained as the solution of the linear system of equations

$$\begin{pmatrix} 1_K & R \\ 0 & 1'_K \end{pmatrix} c = \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{8}$$

where 1_K is the K-dimensional vector with all components equal to one and R is a $K \times K$ matrix with entry (i, j) equal to $R_m^l(d_j|d_i)$. Vector c is such that its last K elements add up to zero and ensures that $\log D(d_k) = v_k,\ k = 1, \ldots, K$. This is a suitable tool for use in numerical computations as the matrix on the left-hand side of (8) is non-singular as long as $d_k,\ k = 1, \ldots, K$ are distinct.

The flexibility of spline models is well documented in the literature (see *e.g.* Wahba, 1990). By allowing an appropriate number of knots, positioned at suitable locations, it is possible to approximate any function in $\mathcal{C}(l)$.

4 Regression Modelling

In the remainder we assume that we have n observations from an underlying process, given by pairs $(x_i, y_i),\ i = 1, \ldots, n$, where $x_i \in \Re^k$ is a vector of explanatory variables and $y_i \in \Re^m$ is the variable of interest. Throughout, we condition on x_i without explicit mention. We assume that the observables $y_i \in \Re^m,\ i = 1, \ldots, n$, are independently generated from

$$y_i = h(B, x_i) + \epsilon_i, \tag{9}$$

where $h(B, x_i)$ is a function in $\Re^m$, B parameterises the location and ϵ_i has a directionally dispersed distribution centred at the origin, with density generator $g^{(m)}$ and directional dispersion function D. Equivalently, from (9) and Theorem 4, $y_i \sim DC_m(h(B, x_i), g^{(m)}, D)$. In some cases, the labelling function $g^{(m)}$ can depend on extra parameters, such as the degrees of freedom for the Student-t. Inference on these parameters, denoted by γ, can also be conducted.

We adopt a prior structure given by

$$P_{B,D,\gamma} = P_B \times P_D \times P_\gamma,$$

assuming prior independence between the location parameters, the dispersion function and γ. Prior distributions for B and γ will be discussed in the context of the applications, and we will focus here on the somewhat more challenging specification of P_D. A sensible prior on D is crucial as it provides an automatic protection against overfitting (see also DiMatteo *et al.*, 2001).

Since the logarithm of D is given by an interpolating hyperspherical spline form, with unknown number of knots K, knot locations $\mathcal{B}_D$ and knot values v, the prior on D can be defined through a prior on K, $\mathcal{B}_D$ and v. The elicitation of a prior distribution on these parameters is facilitated if we write

$$P_{K,\mathcal{B}_D,v} = P_{\mathcal{B}_D,v|K} P_K.$$

In Bayesian curve fitting through splines, two prior distributions on K have often been suggested: a Uniform prior on the set of integers between one and some $K_{\max}$, and a Poisson prior with parameter λ_K, restricted to $K > 0$ (see DiMatteo *et al.*, 2001 and references therein). In this article we use the latter, thus having the possibility to penalise models with a number of knots that we believe inadequate without restricting the support.

We also assume independence between locations and values of the knots. This is an almost inevitable choice in the absence of further information. The prior on $\mathcal{B}_D$ is derived by assuming that each of its elements has a Uniform distribution on $\mathcal{S}^{m-1}$, independent of all other elements.

The prior distribution on D is completed by setting the prior on v. For each component we use a Normal prior with mean μ_v and variance σ_v^2, implying that at each knot, the distribution of D is Log-Normal. Thus, we centre the prior over the benchmark case of sphericity.

By assigning a Uniform prior on the knot locations and a jointly independent prior on the values at the knots, we define a prior that is invariant to linear orthogonal transformations and, therefore, is uninformative about the direction of skewness. Prior knowledge about directional dispersion could be incorporated by making the priors on $\mathcal{B}_D$ and v dependent.

5 Inference

Inference for the regression model introduced in the previous section requires numerical methods. Further, as the number of knots K is left unspecified, inference must be conducted on a space with variable dimension. Here we briefly describe some details of a reversible jump MCMC sampler (Green 1995) that can be used in this context. Further details can be obtained from the authors upon request. General references on the theory and application of MCMC methods can be found in *e.g.* Tierney (1994) and Gilks *et al.* (1996).

Like in the previous section, we leave the regression function and the indexing parameter of the density generator unspecified for now and focus solely on sampling the parameters of D. Due to the independent roles of location, D and γ, sampling B, D and γ independently is likely to work well.

In order to parameterise the knot location d_k we use polar coordinates parameterised by $\omega^k = (\omega_1^k, \ldots, \omega_{m-1}^k)$ as in Theorem 1. The Uniform prior of d_k on $\mathcal{S}^{m-1}$ implies the prior is given by

$$p(\omega^k) \propto \prod_{j=1}^{m-2} \sin^{m-j-1} \omega_j^k, \quad k = 1, \ldots, K.$$

The sampler has four move types: (a) change values, (b) change locations, (c) introduce a new knot and (d) remove an existing knot. Moves (a) and (b) are fixed dimension moves and are defined

as random walk Metropolis-Hastings moves (see *e.g.* Chib and Greenberg, 1995), while moves (c) and (d) change the dimension of the parameter space.

For move (a), we select one existing knot $k \in \{1, \ldots, K\}$ and then propose a new knot value v_k^* sampled from a Normal distribution with mean v_k, the current value of the knot, and variance tuned so as to obtain an acceptance rate close to 30% (as discussed in *e.g.* Chib and Greenberg, 1995).

Sampling location is again started by selecting one existing knot $k \in \{1, \ldots, K\}$. Then, component $j \in 1, \ldots, m-1$ of ω^k is selected and a new value for ω_j^k* is proposed from a Normal distribution centred at ω_j^k, variance tuned to obtain the same acceptance rate as for move (a), and truncated to $[0, \pi)$ if $j \leq m-2$ or $[0, 2\pi)$ if $j = m-1$. When $m = 2$, the location of each knot is parameterised using only one angle. In this case, ordering the locations of the knots is possible and a sampler similar to the change-point locations sampler suggested in Section 4 of Green (1995) could be used. We did not find any significant differences between these samplers and therefore we always use the former sampling scheme which is valid for any dimension.

For the introduction of a new knot, we set $K^* = K + 1$, sample a location d_{K^*} in $\mathcal{S}^{m-1}$ from the Uniform prior distribution and then sample a knot value v_{K^*} from a Normal distribution centred on the current value of $\log D$ at d_{K^*} and variance tuned as previously. The proposed new set of locations and values are then $\mathcal{B}_D^* = \mathcal{B}_D \cup d_{K^*}$ and $v^* = (v', v_{K^*})'$.

Finally, removing a knot can only be done when the number of existing knots is larger than one and is performed by selecting $k \in \{1, \ldots, K\}$ from the corresponding discrete Uniform distribution and setting $K^* = K - 1$, $\mathcal{B}_D^* = \mathcal{B}_D \setminus d_k$ and $v^* = (v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_K)'$.

6 Examples

In the sequel, we always make use of cubic hyperspherical splines, *i.e* we use R_m^l with $l = 2$ as the spline basis. This generates dispersion functions that we feel are sufficiently smooth for most applications. Nevertheless, we remind the reader that choosing a different l does not imply any methodological changes. We always normalise the data so that each component has mean zero and variance one. This simplifies the choice of the prior hyperparameters. For the first application, we use $g^{(m)}$ to be the density generator of the multivariate standard Normal distribution, so that we do not need γ. In the second application, we also use a multivariate Student density generator, with $\gamma > 0$ the degrees of freedom parameter.

The description of the model requires the specification of the hyperparameters λ_K, μ_v and σ_v^2. We chose $\lambda_K = 5m$ reflecting the fact that we expect a larger number of knots to be necessary in higher dimensions. We set μ_v to zero and $\sigma_v = 1.4$ corresponding to a prior that puts 90% of the mass of the directional dispersion along the direction set by each knot between one tenth and ten.

Inference is conducted using MCMC chains of 120,000 iterations, retaining every 10th sample after a burn-in period of 20,000 draws. All results reported in the sequel are based on the output from these samplers.

We assess the relative adequacy of different models using Bayes factors (see Kass and Raftery, 1995). Estimates of marginal likelihoods are obtained using the p_4 measure in Newton and Raftery (1994), with their $\delta = 0.1$.

6.1 Australian Institute of Sport Data

First, we illustrate D-class distributions in the context of fitting multivariate distributions. In the setup introduced in Section 4, this corresponds to choosing $h(B, x_i) = B$, with B an m-dimensional vector of real coefficients, indicating the mode. For B we assume a multivariate Normal prior with mean zero and covariance matrix $100I_m$, corresponding to a diffuse proper prior. We use random walk Metropolis-Hastings moves to sample, in turn, each element of B.

We compare the D-class model, henceforth D-Normal, with three other models: the D-class model where symmetry is imposed (symmetric D-Normal), the elliptical (Normal) and the model using the skewed Normal distribution as defined in Ferreira and Steel (2004a) (FS-Normal). The latter distribution is based on a general linear transformation of a multivariate random variable with independent skewed components and has a parameterisation of fixed dimension. In the framework introduced in Section 3, dispersion functions that generate symmetric distributions are easily generated by imposing that if a spline knot d is in $\mathcal{B}_D$ so is knot $-d$, both with the same knot value. For both Normal and FS-Normal models, we adopt the prior defined in Ferreira and Steel (2004b). These priors are designed to be somewhat vague and are invariant with respect to orthogonal transformations. Results with other relatively noninformative priors are similar.

A dataset from the Australian Institute of Sport, measuring four biomedical variables: body mass index (BMI), percentage of body fat (PBF), sum of skin folds (SSF), and lean body mass (LBM), is used. The data were collected for n =202 athletes at the Australian Institute of Sport and are described in Cook and Weisberg (1994). For simplicity of illustration, here we consider the distribution of the six different pairs of variables.

We begin by formally comparing the suitability of the different classes of distributions using Bayes factors. Table 1 presents the logarithm of the Bayes factors with respect to the Normal model for each of the six pairs. Larger positive values indicate more support for the model in question, while negative values denote that the data favour the Normal model instead. The first conclusion that can be drawn from the results in Table 1 is that both skewed alternatives are strongly favoured by the data. Restricting our attention to symmetric distributions we realise that neither of these models performs uniformly better than the other. In addition, differences between the symmetric alternatives are small. A distinct situation occurs for the skewed models. Without exception, the D-class finds more support, with differences in logarithm of Bayes factors up to 15.2 in favour of the D-class alternative, corresponding to a Bayes factor close to four million (which is equal to the posterior odds under unitary prior odds). Only for the distribution of BMI and LBM do the models perform similarly.

Table 1: Biomedical data: Log of Bayes factors for the different models and different pairs of variables with respect to the Normal alternative.

Model	BMI/PBF	BMI/SSF	BMI/LBM	PBF/SSF	PBF/LBM	SSF/LBM
Symmetric D-Normal	0.82	-0.87	-3.74	-3.25	2.62	2.58
FS-Normal	48.02	44.42	21.77	46.99	41.78	38.20
D-Normal	53.42	50.23	22.26	62.13	56.98	50.65

Plotting the posterior predictive D-Normal densities for the six problems illustrates the advantages

of this class. The plots below the diagonal in Figure 3 display predictive contours of the D-class densities superimposed on the data. The shapes of the contours exhibit a marked departure from symmetry. This can be seen from the fact that, in all cases, the mode of the distribution is not at the centre of the contours. Even for the case of the posterior predictive density of variables PBF and SSF, where each contour appears to be close to an ellipse, the mode of the distribution is very near the lower left side of the contours, implying a substantial amount of skewness. The flexibility of the directionally dispersed distributions, demonstrated by the diversity of contour shapes in Figure 3, is generally much more suitable to model the skewness of these data than the more rigid nature of the skewed distributions in Ferreira and Steel (2004a), underlying the FS-Normal model.

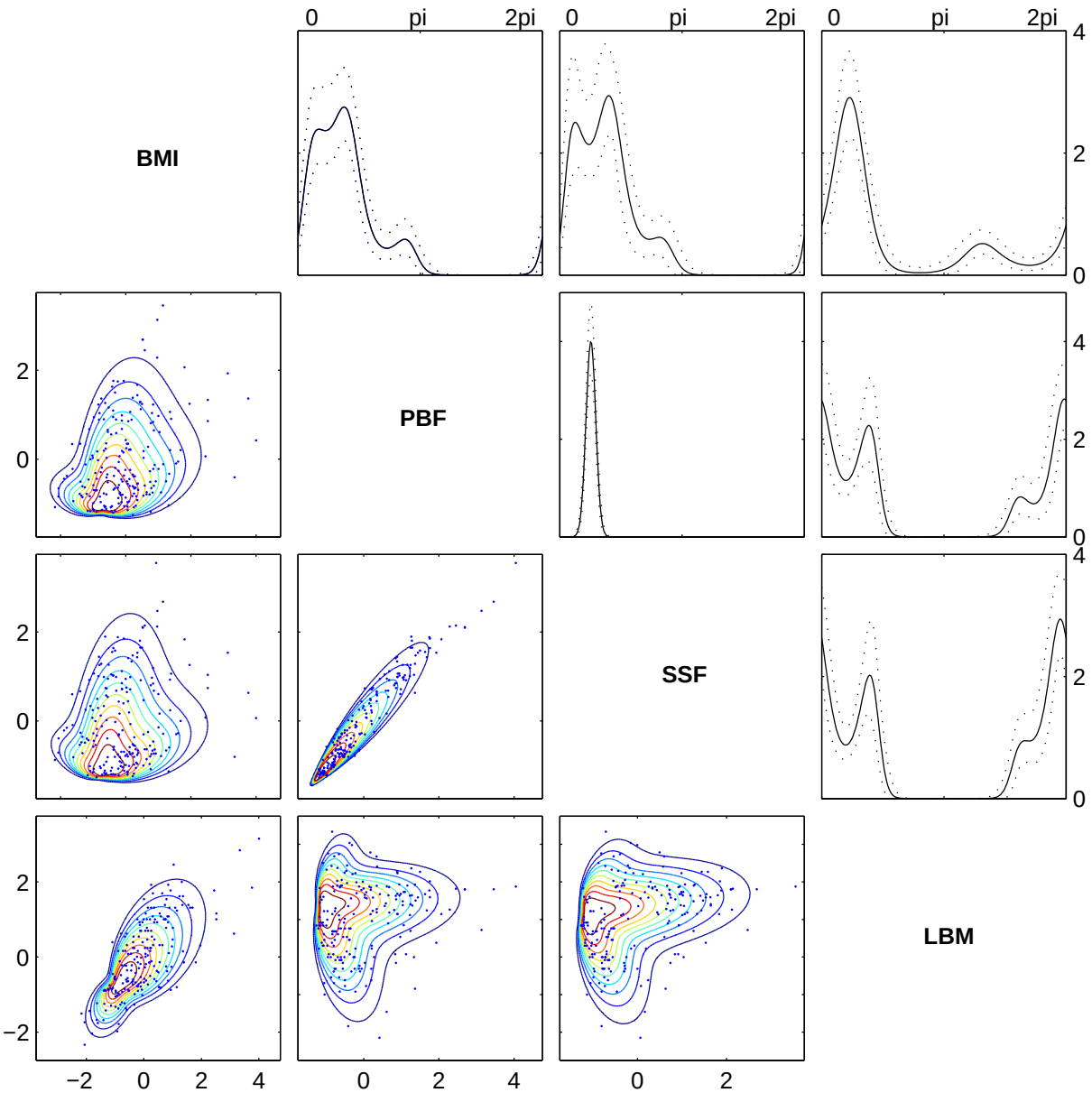

Figure 3: Biomedical data. Below the diagonal: contour plots of the posterior predictive D-Normal densities of the pairs of variables superimposed on the data, denoted by dots. Above the diagonal: mean posterior directional dispersion functions as a function of direction (solid lines) and 95% credible intervals (dotted lines).

The plots above the diagonal in Figure 3 present the posterior means and 95% credible intervals for the directional dispersion functions, obtained from the MCMC output. These plots are quite useful in determining the orientation of the distributions. Areas with small directional dispersion values are areas where the data are more concentrated. Large D indicates greater spread. These plots also highlight the smoothness of the direction dispersion functions modelled with the spline model in $\mathcal{C}(2)$.

As mentioned at the end of Section 2, one of the appealing characteristics of the D-class is that the description of multivariate skewness, as in Ferreira and Steel (2004b), is straightforward. Figure 4 presents posterior mean and 95% credible intervals of directional skewness for the distributions of BMI/PBF, BMI/LBM, PBF/SSF, and PBF/LBM. The cases of BMI/SSF and SSF/LBM were excluded as they are very similar to BMI/PBF and PBF/LBM, respectively. We followed Ferreira and Steel (2004b) and quantified directional skewness using the univariate measure of skewness introduced in Arnold and Groeneveld (1995). We only present the range $[0, \pi)$ because adding π merely reverses the sign of the measure of skewness. Skewness for the different distributions again highlights the flexibility of the D-class. In all cases but the one presented in Figure 4(b), skewness of the distribution can be quite extreme (values close to the limits $\pm$ 1). For BMI/PBF and PBF/LBM the transition between extreme positive and negative values of directional skewness was more gradual than for PBF/SSF, illustrating the sharpness of the directional change in the joint distribution of these last two variables.

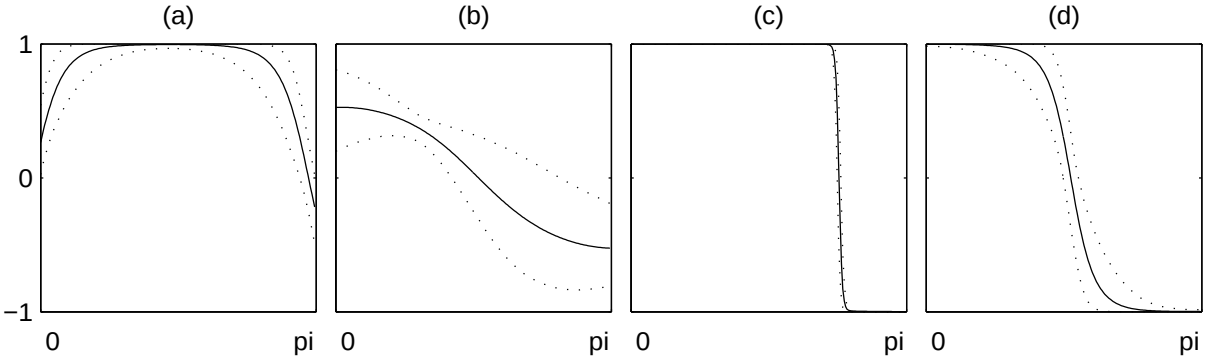

Figure 4: Biomedical data: Directional skewness of the distribution of BMI and PBF (a), BMI and LBM (b), PBF and SSF (c) and PBF and LBM (d), as a function of direction.

Finally, we present some statistics for the number of knots used in the spline modelling of the logarithm of the dispersion functions. Table 2 presents the minimum, median, mean and maximum values for the number of knots recorded in the MCMC sample. Overall, the number of knots used never exceeded eleven (so there is no overfitting) and was never smaller than two. Modelling directional dispersion for PBF/SSF required the smallest number of knots. This is in line with the simpler shape of the corresponding directional dispersion function presented in Figure 3.

6.2 Firm Size Data

The study of firm size is a perennial problem in economics. Here we analyse the joint distribution of a set of measures of the size of 300 publicly traded U.S. manufacturing firms in 1990, using two covariates. The data are described in Hall (1993). The three measures of firm size are market value, tangible assets and sales, and the two covariates used are research and development (R&D) effort

Table 2: Biomedical data: Posterior statistics for the number of knots used in the spline modelling of the logarithm of the dispersion functions (prior mean and variance is 10).

Statistic	BMI/PBF	BMI/SSF	BMI/LBM	PBF/SSF	PBF/LBM	SSF/LBM
Min.	4	6	3	2	6	6
Median	6	7	5	3	6	6
Mean	6.44	6.61	4.86	2.62	6.61	6.56
Max.	10	11	9	5	10	11

and investment. The size measures were expressed as the logarithm of the original values and were normalised to have mean zero and variance one. The covariates R&D and investment are measured as the ratio between quantity spent and total assets and both were also normalised. In addition, a constant term is included.

We assume a linear form for the regression function setting $h(B, x_i) = B'x_i$, where $B \in \Re^{k\times m}$ is a matrix of regression coefficients and x_i is the k-dimensional vector of covariates corresponding to observation $i \in \{1, \ldots, n\}$. For the firm size data problem $k = m = 3$. A matrix-variate Normal prior is chosen on B, with mean zero and variance $100I_{km}$. For this application we consider two directionally dispersed models resulting from two density generators. In addition to the Normal density generator, we consider the Student-t density generator with γ (unknown) degrees of freedom. For this so-called D-Student model, the prior on γ is chosen to be Exponential with mean ten. Sampling the components of B and γ is done independently using random walk Metropolis-Hastings moves.

The Student-t density generator is favoured, with an estimated log Bayes factor of 22.65 with respect to the D-Normal model, implying that heavier tails find much more support in the data.

For this three-dimensional problem, plotting the density is not trivial. However, we can gain substantial insight by analysing the directional dispersion function D. Figure 5 presents the posterior mean and standard deviation of D as a function of direction for the D-Student model. There are mainly two areas where mean directional dispersion has large values. These two peaks are roughly antipodal on the S^2 sphere, which might suggest a symmetric distribution. However, the different relative heights of the peaks provide evidence in favour of skewness. Panel (b) of Figure 5 indicates that the posterior on D is concentrated enough to conduct meaningful inference. A very similar mean directional dispersion function is obtained when using the Normal density generator. However, for this density generator standard deviation values are larger. This is probably due to the inadequacy of the Normal tail behaviour.

Figure 5 also illustrates the smoothness of the function D as modelled with hyperspherical log-splines, even when the value of l is chosen as low as two. Another interesting aspect is the complexity of the directional dispersion function measured by the number of bases that is used. This is also relevant for computation as, from the formulation in (7), the computational burden will increase linearly with the number of knots. Figure 6(a) shows the posterior probability mass function of the number of bases for the D-Student model. The latter assigns non-zero mass to values in the range 6-17 and has a mode of 12. The prior used on D clearly prevents overfitting and the concentration of the probability mass on relatively small values facilitates inference. Similar results are obtained for the D-Normal case.

Figure 6(b) presents the posterior density of the degrees of freedom of the Student-t density

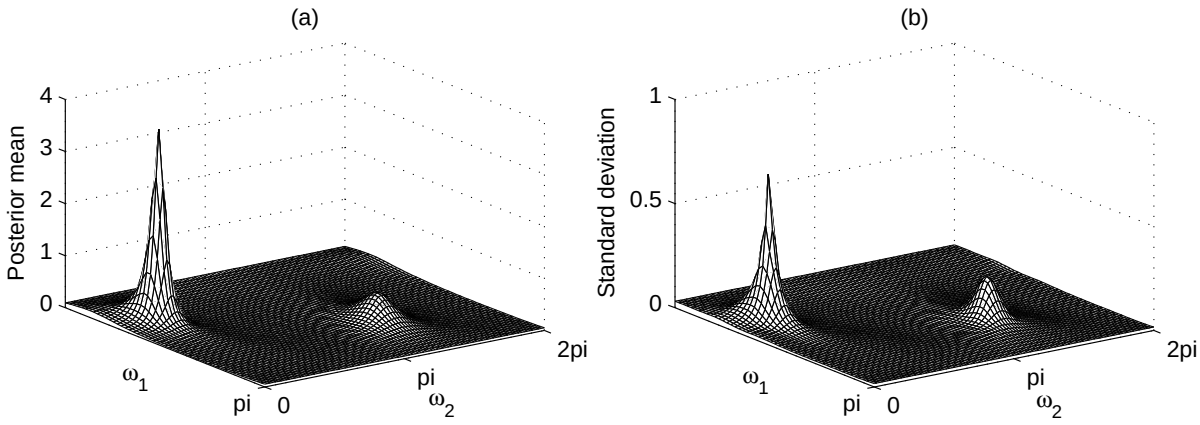

Figure 5: Firm size data: Posterior mean (panel (a)) and posterior standard deviation (panel (b)) of directional dispersion D as a function of direction $d_{y-\mu_y}$ parameterised as in Theorem 1 for the D-Student model.

generator. Mass is concentrated on fairly small values of γ with a 95% credible interval $(4.01, 8.74)$, thus confirming the evidence provided by the Bayes factor.

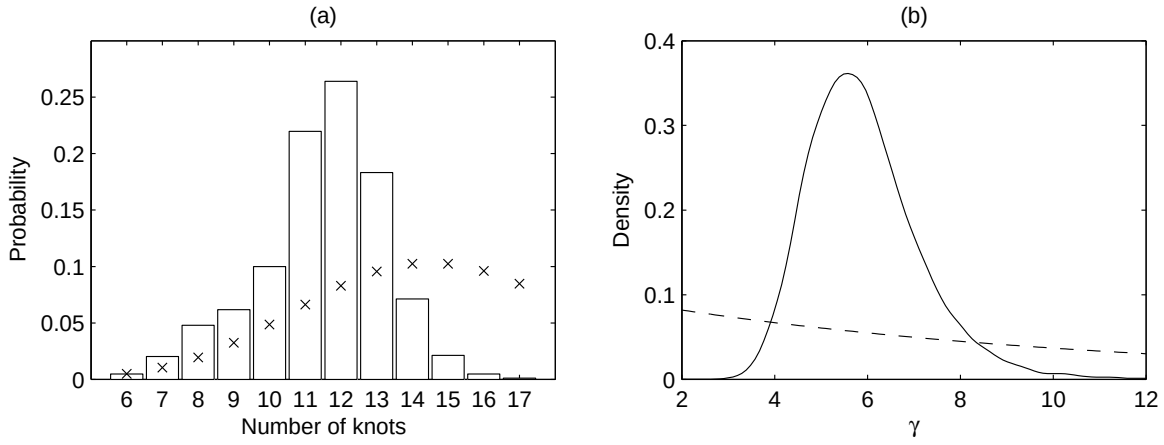

Figure 6: Firm size data: Prior (crosses) and posterior (bars) probability mass functions of the number of knots (panel (a)); Prior (dashed) and posterior (solid) densities of γ (panel (b)).

We conclude by analysing the location of the knots used in modelling D. Figure 7 exhibits a scatter plot of the locations of the knots, visited by the MCMC sampler, after the burn-in period, for the D-Student model. A similar pattern emerges for the Normal density generator. A clear connection can be made between the location of the knots and the form of D illustrated in Figure 5. Large areas of the ω space were only sparingly visited, and there are some regions where the locations of the knots concentrate. The latter are mainly situated around the peaks shown in Figure 5.

Finally, we stress that the inference on D is very similar for both choices of $g^{(m)}$, illustrating the complementarity of the dispersion function and the tails in modelling.

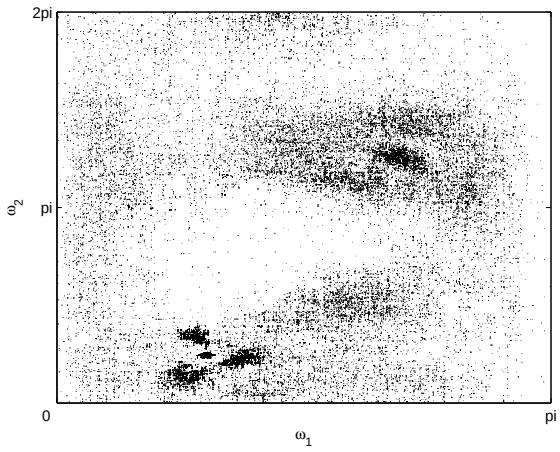

Figure 7: Firm size data: Sampled locations of the knots of the directional dispersion function.

7 Conclusion

This article introduces a novel class of multivariate distributions through explicit modelling of directional dispersion. This so-called D-class generalises the class of continuous elliptical distributions in an intuitive manner, and provides a useful framework for modelling. These distributions can be generated by the density generator of a spherical distribution while allowing dispersion to vary with the direction in the space. We show that, by controlling directional dispersion, it is possible to generate a very wide class of distributions, that we believe will prove useful for the description of real phenomena. By imposing a simple and empirically reasonable condition on the directional dispersion function, we show that unimodality can be controlled and that moment existence characteristics of the spherical distribution generated by the density generator are carried over to our more general setup. The D-class is also closed under affine linear transformations.

We suggest modelling directional dispersion using hyperspherical log-splines, where we can approximate any (smooth) function on a hypersphere to any degree and can control smoothness straightforwardly. Hyperspherical splines are well-defined in real spaces of any dimension and are relatively easy to implement. This provides a flexible, yet practical, modelling framework which allows the dimension of the parameterisation to vary in order to accommodate sufficiently complex behaviour.

The directionally dispersed class of distributions is used in the context of a Bayesian regression model. In the prior structure we separate the roles of the regression function and of the error distribution. We define a prior distribution on directional dispersion, using invariance arguments, through priors on the parameters of the hyperspherical log-spline. Inference for this regression framework is conducted using a reversible jump Markov chain Monte Carlo sampler.

Inevitably, some modelling choices have to be made. The required smoothness of the directional dispersion function D has to be chosen, as well as the prior distribution. A sensible choice for the prior on D is crucial in preventing overfitting. The prior proposed here for D is relatively universal and can be used safely in a wide variety of applications to reflect a lack of specific prior information. In the context of our applications, results were found to be relatively robust to moderate deviations from the choices made here.

Modelling directional dispersion is complementary to the choice of a suitable density generator. The latter issue is really the same as in the spherical case. In particular, the use of scale mixtures of

distributions can easily be implemented in our framework.

Acknowledgements: Part of the work of the first author was supported by grant SFRH BD 1399 2000 from Fundação para a Ciência e Tecnologia, Ministério para a Ciência e Tecnologia, Portugal. We thank the Joint Editor, the Associate Editor and two Referees for very helpful comments.

Proofs

Without loss of generality, in the sequel we assume that $\mu_y = 0$.

Proof of Theorem 1. We use the transformation to polar coordinates $y = \rho\, t(\omega)$, for which

$$p(\rho, \omega|D) = \frac{1}{K_D}\rho^{m-1}\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) g^{(m)}\left\{\frac{\rho^2}{D[t(\omega)]}\right\}. \tag{10}$$

Now, (5) denotes a density if and only if (10) integrates to unity. Defining $\rho = D[t(\omega)]^{1/2}r$, the integral becomes

$$\frac{1}{K_D}\int_\Omega\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) D[t(\omega)]^{m/2}d\omega \int_{\Re^+} r^{m-1}g^{(m)}(r^2)dr =$$

$$\frac{1}{K_D}\frac{\Gamma(m/2)}{2\pi^{m/2}}\int_\Omega\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) D[t(\omega)]^{m/2}d\omega, \tag{11}$$

which concludes the proof. □

Proof of Corollary 1. If for any direction d_y, $D(d_y) < C < \infty$, then there is a finite M such that for any d_y, $D[t(\omega)]^{m/2} < M$, and therefore

$$\int_\Omega\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) D[t(\omega)]^{m/2}d\omega < M\int_\Omega\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) d\omega < \infty.$$

□

Proof of Corollary 2. Follows directly from (11). □

Proof of Theorem 2. As $f_m(y|0, g^{(m)}, D)$ defined by (5) is proportional to $g^{(m)}\left[\frac{y'y}{D(d_y)}\right]$ and $0 < D(d_y) < C < \infty$, $g^{(m)}\left[\frac{y'y}{D(d_y)}\right]$ is decreasing in y for any direction d_y and, thus, the result follows. □

Proof of Theorem 3. The existence of the joint moment of order q_j in y_j, for $j = 1, \ldots, m$, is equivalent to

$$\int_{\Re^m}\prod_{j=1}^{m}|y_j|^{q_j} f_m(y|0, g^{(m)}, D)dy < \infty.$$

Transforming to polar coordinates and defining $q = \sum_{j=1}^m q_j$, the integral above equals

$\frac{1}{K_D}\int_{\Re^+\times\Omega}\rho^q\prod_{j=1}^m|t_j(\omega)|^{q_j}\rho^{m-1}\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right) g^{(m)}\left\{\frac{\rho^2}{D[t(\omega)]}\right\}d\rho d\omega =$

$\frac{1}{K_D}\int_\Omega\left(\prod_{i=1}^{m-2}\sin^{m-1-i}\omega_i\right)\prod_{j=1}^m|t_j(\omega)|^{q_j}D[t(\omega)]^{\frac{q+m}{2}}d\omega\int_{\Re^+} r^{q+m-1}g^{(m)}(r^2)dr \propto$

$\int_{\Re^+} r^{q+m-1}g^{(m)}(r^2)dr$, since the integral in ω is a finite positive number.

As the last integral above depends on $g^{(m)}$ alone, the proof is complete. □

Proof of Theorem 4. Follows from a variable transformation using the density function in (5). We use a nonsingular transformation as we need z to have a density in line with the definition of the D-class. □

Proof of Corollary 3. The sufficiency part follows directly from Theorem 4. The necessity follows from the fact that only elliptical distributions can be transformed to sphericity by linear transformations. □

References

Abramowitz, M. and Stegun, I. (eds) (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, New York: Dover.

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode, *Am. Statist.* **49**: 34–38.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *Am. Statist.* **49**: 327–335.

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*, New York: Wiley.

DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines, *Biometrika* **88**: 1055–1071.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall.

Fernández, C., Osiewalski, J. and Steel, M. F. J. (1995). Modeling and inference with v-spherical distributions, *J. Amer. Statist. Assoc.* **90**: 1331–1340.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004a). Bayesian multivariate regression analysis with a new class of skewed distributions, Mimeo, University of Warwick.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004b). On describing multivariate skewness: A directional approach, Statistics Research Report 429, University of Warwick.

Genton, M. G. (ed.) (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton: CRC Chapman & Hall.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Hall, B. H. (1993). The stock market's valuation of R&D investment during the 1980's, *Amer. Econ. Rev.* **83**: 259–263.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *J. Amer. Statist. Assoc.* **90**: 773–795.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generelization, *Sankhya* **32**: 419–430.

Müller, C. (1966). *Spherical Harmonics*, Vol. 17 of *Lecture Notes in Math.*, Springer-Verlag, New York.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion), *J. Roy. Statist. Soc. B* **56**: 3–48.

Osiewalski, J. and Steel, M. F. J. (1993). Robust Bayesian inference in l_q-spherical models, *Biometrika* **80**: 456–460.

Taijeron, H. J., Gibson, A. G. and Chandler, C. (1994). Spline interpolation and smoothing on hyperspheres, *SIAM J. Sci. Comput.* **15**: 1111–1125.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *Ann. Statist.* **22**: 1701–1762.

Wahba, G. (1975). Smoothing noisy data by spline functions, *Numer. Math.* **24**: 383–393.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere, *SIAM J. Sci. Statist. Comput.* **2**: 5–16.

Wahba, G. (1990). *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Apllied Mathematics, Philadelphia.