The Canadian Journal of Statistics
Vol. 31, No. ?, 2006, Pages ???-???
La revue canadienne de statistique

On describing multivariate skewed distributions: A directional approach

José T.A.S. FERREIRA and Mark F.J. STEEL

Key words and phrases: Bayesian methods; Directional skewness; Multivariate regression; Prior elicitation; Prior matching.
MSC 2000: Primary 62H05; 62E15; secondary 62F15.

Abstract: Most multivariate measures of skewness in the literature measure the overall skewness of a distribution. These measures were designed for testing the hypothesis of distributional symmetry and their relevance for describing skewed distributions is less obvious. In this article, we consider the problem of characterising the skewness of multivariate distributions. We define directional skewness as the skewness along a direction and analyse two parametric classes of skewed distributions using measures based on directional skewness. The analysis brings further insight into the classes, allowing for a more informed selection of classes of distributions for particular applications. In the context of Bayesian linear regression under skewed error we use the concept of directional skewness twice. First in the elicitation of a prior on the parameters of the error distribution, and then in the analysis of the skewness of the posterior distribution of the regression residuals.

Title in French:

Résumé : Most multivariate measures of skewness in the literature measure the overall skewness of a distribution. These measures were designed for testing the hypothesis of distributional symmetry and their relevance for describing skewed distributions is less obvious. In this article, we consider the problem of characterising the skewness of multivariate distributions. We define directional skewness as the skewness along a direction and analyse parametric classes of skewed distributions using measures based on directional skewness. The analysis brings further insight into the classes, allowing for a more informed selection of particular classes for particular applications. In the context of Bayesian linear regression under skewed error we use the concept of directional skewness twice. First in the elicitation of a prior on the parameters of the error distribution, and then in the analysis of the skewness of the posterior distribution of the regression residuals.

1. INTRODUCTION

Modelling skewness in the distribution of real phenomena is becoming common statistical practice, with recent years seeing the development of a number of classes of multivariate distributions designed for such tasks. However, the increased depth of the distributional toolbox available to the researcher was not complemented by tools that allow a characterisation of skewness. This article tries to fill part of this gap. We propose measures of multivariate skewness that are more informative for describing distributions than the traditional measures of overall or total skewness.

Quantifying multivariate skewness has been a perennial problem. Traditionally, the main objective of the measures was to provide statistics that could be used for testing the hypothesis that the distribution of the quantity of interest was symmetric (in the sense that random variable $X-\mu$ has the same distribution as $\mu - X$, for some constant vector μ). Therefore, the measures of multivariate skewness were primarily, and often uniquely, developed for testing lack of symmetry (*e.g.* see Henze 2002, Section 3 for a review of normality tests based on measures of skewness).

The measures of multivariate skewness in the literature can be broadly divided into three groups. The first group is made up of measures based on joint moments of the random variable (see *i.a.* Mardia 1970 and Móri, Rohatgi & Székely 1993). A different approach was taken by Malkovich & Afifi (1973) who made use of projections of the random variable onto a line. It then selects the direction along which the projection maximises some value of univariate skewness, and sets the measure of multivariate skewness as the square of the skewness value along that direction. The third class of measures was suggested by Oja (1983) and uses volumes of simplexes. Even though these groups are intrinsically distinct, they all have a number of common characteristics: they take values on the non-negative real line, are zero for symmetric distributions and are invariant to affine linear transformations. However, they measure overall skewness, and are uninformative about how skewness varies with direction. This makes them of limited use for characterising skewed distributions. Nevertheless, the $\beta_{1,p}$ measure suggested by Mardia (1970) has been applied in the characterisation of multivariate skewed distributions (*i.a.* Sahu, Dey & Branco 2003 and Ferreira and Steel 2004). Yet, in these studies the measure was mainly used to compare ranges of skewness values between different classes of distributions.

Our proposal is based on the key concept of directional skewness, *i.e.* the amount of skewness along a particular direction. Along any direction, the skewness of the distribution can be quantified using an univariate measure. In addition, several measures of univariate skewness are available, some of which are fairly interpretable. By associating a direction with an interpretable value of univariate skewness, we can gain greater insight into the properties of the multivariate distribution.

We present two alternatives for employing directional skewness. The first provides full information about the skewness of the multivariate distribution by quantifying skewness along every direction in the multivariate space. This is a feasible procedure if the distributions along the directions take a simple form. For more complicated setups, we suggest the use of partial information, consisting of measuring skewness along each one of a set of orthogonal directions, spanning the complete space. In particular, we suggest the use of a specific set of orthogonal directions, called principal axes of skewness, which are defined in Section 3.

We employ directional skewness to characterise two rather distinct classes of skewed distributions suggested in the literature: the skew-Normal of Azzalini & Dalla Valle (1996) and the skew-Normal of Ferreira and Steel (2004), henceforth ADV-Normal and FS-Normal, respectively. A comprehensive comparison, in terms of skewness, between members of these two classes is then immediate.

We apply the concepts of directional skewness in the context of a Bayesian regression model, where the errors have a distribution of the form ADV-Normal or FS-Normal. First we use directional skewness to perform prior matching between the parameters of both classes. We then use directional skewness to characterise the predictive posterior distributions. We analyse a well-known set of biometrical measurements data. It is worth stressing that the main aim of the paper is not to point out the relative merits of these two classes, but rather to illustrate how the various concepts related to directional skewness can be applied to some classes of multivariate distributions that are used in the literature, and how directional skewness can help in characterising the differences between these distributions.

In Section 2 we provide a brief review of measures of univariate skewness. Section 3 introduces the concepts of directional skewness and of principal axes of skewness. In Section 4 we analyse two classes of distributions using full information on directional skewness. In Section 5, we study the application of directional skewness to a Bayesian regression model. The final section groups some further remarks. Proofs are deferred to the Appendix, without explicit mention in the body of the text.

2. MEASURES OF UNIVARIATE SKEWNESS

Several measures of univariate skewness have been proposed, and here we provide a brief summary. For a more complete review of the literature see *e.g.* Arnold & Groeneveld (1995) and references therein. Throughout the paper, upper case symbols will denote, interchangeably, distributions or distribution functions, with the corresponding lower case versions denoting densities. We always assume that the densities exist.

Let F and G denote two univariate distributions, and let $X \sim F$. Following Oja (1981), a measure of skewness $Sk(\cdot)$ should satisfy the following four properties:

1. For any symmetric F, $Sk(F) = 0$.
2. Let $k_1 \in \Re_+$, $k_2 \in \Re$ and $k_1X + k_2 \sim G$, then $Sk(G) = Sk(F)$.
3. For any F, if $-X \sim G$ then $Sk(G) = -Sk(F)$.
4. If $G^{-1}[F(x)]$ is convex, where $F(\cdot)$ and $G(\cdot)$ denote the distribution functions of F and G, then $Sk(F) \leq Sk(G)$.

A number of functionals that meet the properties above have been proposed. Let X denote the random variable with distribution F, while μ, μ^+, μ^* denote mean, median and mode, respectively. Further, let Q_1, Q_3 denote the first and third quartiles of F and let σ denote the standard deviation. We mention three distinct measures:

$CE = E[(X-\mu)^3]/(\sigma^3)$, proposed by Edgeworth (1904) and Charlier (1905).

$B = (Q_3 + Q_1 - 2\mu^+)/(Q_3 - Q_1)$, suggested by Bowley (1920).

$AG = 1 - 2F(\mu^*)$, introduced by Arnold & Groeneveld (1995).

These measures are quite different, both in terms of how they quantify skewness and their applicability.

The CE measure has, perhaps, been the most widely used. As skewness is quantified by dividing the third central moment by the cubed standard deviation, it takes values on $\Re$ and its applicability is restricted to distributions for which the third moment exists. The second measure above is well defined for any distribution. It depends solely on the quartiles of F and takes values in $(-1, 1)$. Despite the generality of the measure, it is somewhat hard to interpret its results. For unimodal distributions, the AG measure, in $[-1, 1]$, is well defined. It quantifies skewness using the mass to the left of the mode. Like B, it makes no assumptions about the existence of moments of the distribution. The simplicity and interpretability make AG attractive for unimodal distributions.

3. CHARACTERISING MULTIVARIATE SKEWNESS

In this article, we restrict our attention to the characterisation of multivariate skewness for distributions that are unimodal. In fact, it is somewhat awkward to apply the concept of asymmetry to multimodal multivariate distributions.

The definition of directional skewness that will be introduced in Subsection 3.1 uses the concept of centre of a multivariate distribution. For unimodal skewed distributions, the unique mode is the obvious location for this centre and here we elaborate on directional skewness using the mode as the centre. However, other choices for the centre are possible, including the mean or some form of multivariate median. These locations would be suitable for examining asymmetry for multimodal distributions.

The quantification of directional skewness requires the use of a measure of univariate skewness, denoted by Sk, that follows properties 1-4 described in Section 2.

3.1 Directional skewness

Definition 1. Let $X \in \Re^m$ be a random variable with unimodal multivariate distribution F, and mode μ^*. Further, let Sk be a measure of univariate skewness, $d \in \Re^m$ denote a direction, represented by a vector with unitary norm and O^d be an orthogonal matrix with first column equal to d. Finally, let G be the distribution of $Y = (y_1, \ldots, y_m)' = (O^d)'(X - \mu^*)$. Then, the **directional skewness** of F along direction d is defined as

$$Sk_m(F, d) = Sk\left(G_{y_1|y_{-1}=0}\right) = Sk(F_d),$$

where y_{-1} denotes the last $m-1$ components of Y and $G_{y_1|y_{-1}=0}$ stands for the distribution of y_1 conditional on $y_{-1} = 0$, the so-called **directional distribution**, henceforth denoted by F_d.

Thus, directional skewness is obtained by centring the distribution on μ^* and measuring the skewness of the distribution of a univariate variable along the direction d conditional on all other (orthogonal) components equal to zero. Due to the orthogonality of O^d, the choice of its last $m-1$ columns does not matter.

Characterising multivariate skewness using directional skewness makes skewness direction-specific. By analysing $Sk_m(F, d)$ for varying d, we can gain substantial knowledge about the asymmetry of F. Further, the dimension m is conceptually irrelevant, as skewness is always quantified through measures on univariate distributions.

In the context of applications it may be especially important to evaluate skewness along certain interesting directions. For such cases, measuring total skewness would be of limited relevance. In contrast, directional skewness provides a much more informative measure.

For the directional distribution we use the conditional distribution of y_1. An obvious alternative would be to use the marginal distribution of y_1, G_{y_1}. One advantage of this alternative definition would be that the concept of centre of the distribution would not be required, therefore naturally extending the scope of the measure to multimodal distributions. However, using marginal distributions would have two major disadvantages, one conceptual and one practical. The conceptual and most important one is lack of interpretability. While the skewness of F_d has an immediate translation into the skewness of F along direction d, the same is not true for the skewness of G_{y_1}. It is not clear at all how $Sk(G_{y_1})$ would relate to F, especially for high dimensional distributions. See Arnold & Beaver (2002) for a general discussion of conditional modelling and its advantages for interpretation. The practical disadvantage is computational. To calculate the density f_d we require a one-dimensional integral. In contrast, an $(m-1)$-dimensional integral is necessary for calculating g_{y_1}. Apart from a few particular cases, the latter is much harder than the former, even for moderate m. The difficulties inherent in dealing with marginal distributions are well documented in the projection pursuit literature, such as discussed in Jones & Sibson (1987).

We now study some properties of $Sk_m(F, d)$.

Theorem 1. *If F is symmetric around the mode μ^* (i.e. $f(x - \mu^*) = f(\mu^* - x)$), then for any direction d, $Sk_m(F, d) = 0$.*

Theorem 2. *If $k_1 X + k_2 \sim H$, where $k_1 \in \Re_+$ and $k_2 \in \Re^m$, then $Sk_m(H, d) = Sk_m(F, d)$.*

Directional skewness preserves invariance to location-scale transformations. However, it is not invariant to multivariate linear transformations. We think that this is a desirable property of a measure meant to characterise multivariate asymmetry. Let us illustrate this with an example. Figure 1 presents contour plots for two bivariate skewed densities, which are linked by a linear transformation. In particular, both distributions displayed in Figure 1 are members of the FS-Normal class explained later in Subsection 4.2 and described in (6) and (7). The choice of the skewness coefficient vector γ is $(0.5, 1.5)'$ and the left plot corresponds to $A = A_1 = (1/\sqrt{2})I_2$ while the right plot corresponds to $A = A_2 \times A_1$ with $A_2 = (-0.9458, 0.6211; 0.3247, -1.2705)$. Thus, the random variable plotted in the right-hand panel is a linear transformation (through A_2) of the one characterised by the left-hand panel. The contours, however, are quite different and we feel it is sensible for a measure of skewness to reflect that. In particular, the right plot corresponds to higher maximum values of directional skewness for all the univariate skewness measures described in Section 2. Note that this property is not shared by the existing multivariate skewness measures mentioned in the Introduction. Of course, the latter are measures of overall skewness, whereas here we focus on characterising skewness as a function of direction.

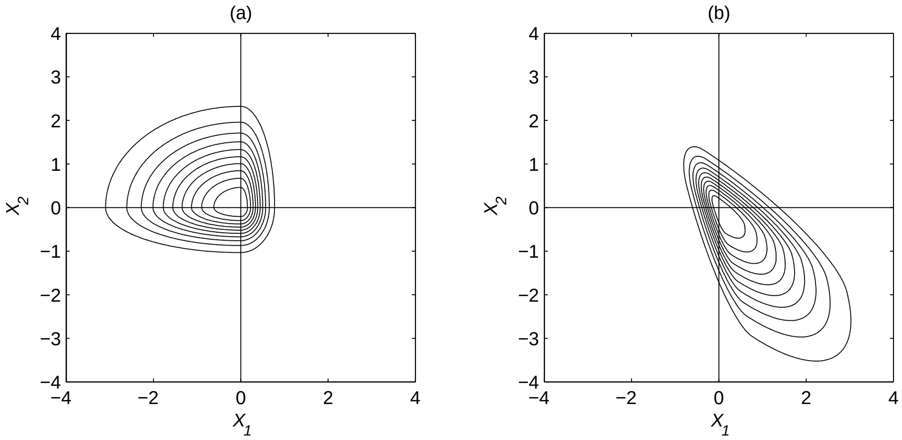

Figure 1: Contour plots of the densities of two bivariate skewed distributions. The distribution represented in (b) is the result of a linear transformation of the variable depicted in (a).

It is clear that $Sk_m(F, d) = -Sk_m(F, -d)$. This follows directly from the properties of the measure of univariate skewness Sk, as described in Section 2. As such, in order to completely describe the skewness of F, it is only necessary to calculate $Sk_m(F, d)$, for $d \in \mathcal{S}^{m-1}$, where $\mathcal{S}^{m-1}$ denotes half of the unit sphere in $\Re_m$.

3.2 Principal Axes of Skewness

A complete characterisation of skewness may be deemed infeasible, either because all that is needed is a simpler, but still informative, description of asymmetry, or because it would be too hard to compute. For these circumstances, we suggest the definition of principal axes of skewness.

Definition 2. Let F and Sk_m be as above, and let $D = \{d_1, \dots, d_m\},\ d_j \in \mathcal{S}^{m-1},\ j = 1, \dots, m$ be a set of orthogonal directions. Further, let $\mathcal{F}$ be a norm function in $\Re^m$. Then, D is a set of principal axes of skewness if

$$D = \arg\max_{\mathcal{S}^{m-1}} \mathcal{F}\left[Sk_m(F, d_1), \dots, Sk_m(F, d_m)\right].$$

A reasonable choice for the norm $\mathcal{F}$ is the l_∞ norm. Then, the axis along which directional skewness is maximal (in absolute value) is a principal axis of skewness. The remaining axes are chosen sequentially following a similar argument. As skewness can be measured in either direction along the axes (which merely changes the sign), we shall always take directional skewness to be non-negative.

It is clear that any F has at least one set of principal axes of skewness. However, there could be several such sets. For example, if F is symmetric, then any orthogonal set of m vectors in $\mathcal{S}^{m-1}$ is a set of principal axes of skewness. For most interesting skewed distributions, the set will be unique.

The direction of the principal axes of skewness and the skewness values along these axes allow the identification of sectors of large directional skewness and its quantification. For most parametric classes of distributions, $Sk_m(F, d)$ will be a well-behaved function of d and therefore, the measures at the principal axes of skewness will provide a good indication of the shape of the distribution.

3.3 Functionals of directional skewness

Skewness can be summarised even further, and characterised by a single quantity. For this, the measures of multivariate skewness mentioned in the Introduction are already available. Here, we analyse how directional skewness can be used to define other univariate measures of total skewness.

The most obvious single quantity of multivariate skewness that can be defined using directional skewness is the integrated directional skewness, IDS, defined as

$$IDS(F) = \left[\int_{\mathcal{S}^{m-1}} |Sk_m(F, r)|^q \, dr \right]^{\frac{1}{q}} \geq 0, \qquad (1)$$

where $q \in \Re_+$.

A measure closely related to the IDS is the mean directional skewness, MDS, defined as

$$MDS(F) = \frac{IDS(F)}{\left(\int_{\mathcal{S}^{m-1}} dr \right)^{\frac{1}{q}}} = \left[\frac{\Gamma(m/2)}{\pi^{m/2}} \right]^{\frac{1}{q}} IDS(F),$$

with $\Gamma(\cdot)$ denoting the gamma function. MDS does not depend on the dimension of F and it takes values on the same space as $|Sk|$.

The information available to construct a single measure of multivariate skewness can be the one contained in the principal axes of skewness, and the correspondent skewness values, leading to obvious discrete counterparts of the two measures above. The l_q norm of the vector $[Sk_m(F, d_1), \ldots, Sk_m(F, d_m)]$, is the discrete version of the IDS measure in (1), denoted by $DIDS$. Likewise, the definition of the discrete version of MDS, $DMDS = DIDS/m$.

It is immediate that all measures that we introduce here take non-negative values and are zero if and only if $Sk_m(F, d)$ is the constant null function of d. Also, as they are based on the concept of directional skewness, they inherit the properties in Theorem 2.

4. COMPLETE DESCRIPTION OF DIRECTIONAL SKEWNESS

In this section, we analyse in detail two classes of skewed distributions: the ADV-Normal class of Azzalini & Dalla Valle (1996) and the FS-Normal class introduced in Ferreira and Steel (2004). These are not distributions that are totally comparable, as they introduce skewness in different ways: the ADV model introduces skewness in a single direction while the FS model induces skewness in m directions. However, the object is not to pit these distributions against one another, but rather to illustrate how the concept of directional skewness can be used to characterise the difference between these two distributions.

For these distributions we present analytical forms for the directional distributions along any direction. These enable a complete description of directional skewness.

Throughout, we do not consider location parameters which, due to Theorem 2, brings no loss of generality.

4.1 The ADV-Normal class

Azzalini & Dalla Valle (1996) introduced a class of skewed normal distributions based on a conditioning argument. Let Σ be an $m \times m$ covariance matrix and $\alpha \in \Re^m$. Then, $X \in \Re^m$ has an ADV-Normal distribution with parameters Σ and α, denoted by $\mathrm{ADV}(\Sigma, \alpha)$, if its density is of the form

$$f_{\mathrm{ADV}(\Sigma,\alpha)}(x) = 2\phi_m(x|0,\Sigma)\Phi(\alpha' x), \tag{2}$$

where $\phi_m(\cdot|0,\Sigma)$ stands for the m-dimensional Normal density with mean zero and covariance Σ and $\Phi(\cdot)$ denotes the standard univariate Normal distribution function. Azzalini & Dalla Valle (1996) note that not all (Σ, α) pairs lead to a valid probability distribution.

We remind the reader that the directional distribution along d was defined in Definition 1 as the conditional distribution in the direction d given that all components corresponding to orthogonal directions are zero, and we can derive the following result:

THEOREM 3. *Let $X \sim ADV(\Sigma, \alpha)$, μ^* be the mode of $ADV(\Sigma, \alpha)$ and d be a vector in $\mathcal{S}^{m-1}$. Then, the density of the directional distribution of X along d is given by*

$$f_{ADV(\Sigma,\alpha),d}(y) = \frac{\phi\left(\frac{y-\mu_d}{\sigma_d}\right)\Phi\left[\delta_{0,d} + \delta_{1,d}\left(\frac{y-\mu_d}{\sigma_d}\right)\right]}{\sigma_d \Phi\left(\frac{\delta_{0,d}}{\sqrt{1+\delta_{1,d}^2}}\right)}, \tag{3}$$

where

$$\begin{array}{ll} \mu_d = -\frac{d'\Sigma^{-1}\mu^*}{d'\Sigma^{-1}d}, & \delta_{0,d} = \alpha'(\mu_d d + \mu^*) \\ \sigma_d = \left(d'\Sigma^{-1}d\right)^{-1/2}, & \delta_{1,d} = \sigma_d(\alpha' d). \end{array} \tag{4}$$

The mode of $\mathrm{ADV}(\Sigma, \alpha)$ is generally not available analytically. However, as (2) is well-behaved, it is easily found numerically, even in high dimensional spaces.

The density $f_{\mathrm{ADV}(\Sigma,\alpha),d}$ given in (3) coincides with (2.4) of Arnold & Beaver (2002), which was proposed as a generalisation of the skew-Normal distribution of Azzalini (1985). As the measures of univariate skewness that we employ are invariant to location and scale transformation, we have that $Sk_m\left[\mathrm{ADV}(\Sigma, \alpha), d\right]$ is equal to the skewness of the distribution with density

$$f(y) = \frac{\phi(y)\Phi\left[\delta_{0,d} + \delta_{1,d} y\right]}{\Phi\left(\frac{\delta_{0,d}}{\sqrt{1+\delta_{1,d}^2}}\right)}, \tag{5}$$

with $\delta_{i,d}$, $i = 0, 1$ as in (4). For the distribution generated by (5) the moment generating function is given by (2.5) of Arnold & Beaver (2002). The maximum achievable skewness as measured by CE is the same as for the univariate skew-Normal of Azzalini (1985), and, thus, $CE \in (-0.99527, 0.99527)$, where the bounds correspond to the values for the half-Normal.

Measures of skewness based solely on moment characteristics can then be calculated directly. For other measures, it is necessary to resort to numerical integration, which is quite feasible as (5) is simple to calculate. Figure 2 presents contour plots of the CE, B and AG measures of skewness as functions of $\delta_{0,d}$ and $\delta_{1,d}$, restricted to positive values of $\delta_{1,d}$, leading to positive skewness. For fixed

$\delta_{0,d}$, changing the sign of $\delta_{1,d}$ merely changes the sign of the measures of skewness. Darker contours correspond to larger values of skewness. In all plots, a non-trivial relationship between parameters and amount of skewness is revealed, with two quite different patterns of contours emerging, one corresponding to CE and B, the other to AG. For CE and B, if $\delta_{0,d}$ is fixed, the amount of skewness is a monotone increasing function of $\delta_{1,d}$ only when $\delta_{0,d} > 0$; for negative values of $\delta_{0,d}$, skewness is a unimodal function of $\delta_{1,d}$.

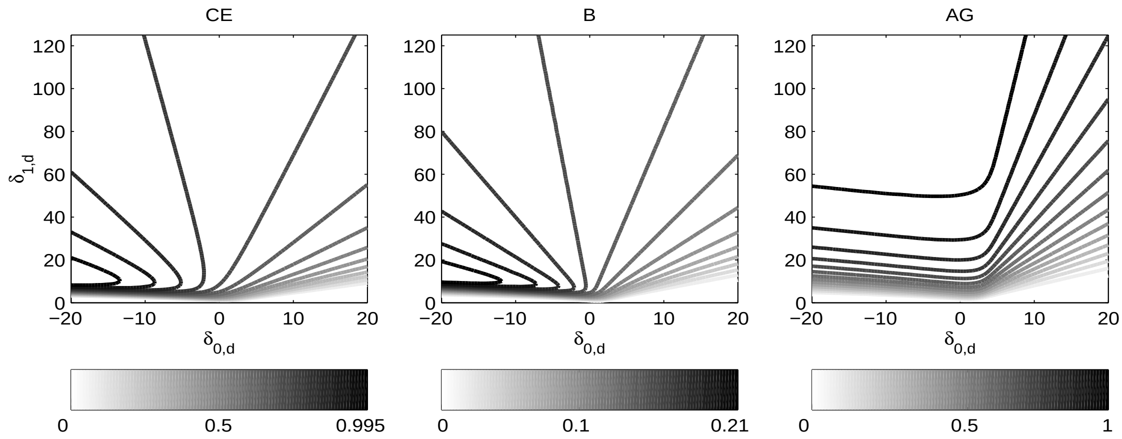

Figure 2: ADV-Normal distribution: Contour plots of the measures of univariate skewness for varying $\delta_{0,d}$ and $\delta_{1,d}$. Darker contours indicate larger values of skewness, as indicated by the greyscales.

4.1.1 Special case $\Sigma = \sigma^2 I$

A particular case of special relevance for the ADV-Normal distribution is when Σ is given equal to a constant σ^2 times the identity matrix. By Theorem 2b, Sk_m is invariant to scale transformations and, thus, here we restrict our attention to the case $\Sigma = I$.

By substituting $\Sigma = I$ in (2), we observe that for fixed $\|x\|$, $f_{\text{ADV}}(\Sigma, \alpha)(x)$ is maximised when $\alpha' x$ is maximal. The latter happens when x has the same direction as α. Therefore, it follows that the mode $\mu^* = k^* \alpha$, for some positive constant k^*.

By replacing $\mu^* = k^* \alpha$ in (4) we have that

$$\begin{aligned} \mu_d &= -k^* d' \alpha, \quad & \delta_{0,d} &= k^* \alpha' (I - dd') \alpha \\ \sigma_d &= 1, & \delta_{1,d} &= \alpha' d. \end{aligned}$$

As $I - dd'$ is non-negative definite, $\delta_{0,d} \geq 0$. Analytically, it is not possible to determine the directions that maximise directional skewness for any of the measures. This is due to the fact that k^* is unknown. Also, both B and AG measures of skewness do not have an analytical form. Nevertheless, we can still resort to numerical computations to draw interesting conclusions. As expected, the modulus of directional skewness, quantified by any of the measures reviewed in Section 2, is maximal if $d = \pm \frac{\alpha}{\|\alpha\|}$, corresponding to $\delta_{0,d} = 0$ and $\delta_{1,d} = \pm\|\alpha\|$. Zero skewness happens for directions perpendicular to α. Any set of m orthogonal vectors in $\mathcal{S}^{m-1}$ including $\pm \frac{\alpha}{\|\alpha\|}$ is a set of principal axes of skewness. Along these axes, directional skewness is non-zero for only one axis, namely the one collinear with α.

Figure 3 shows the directional skewness, for each of the three measures in Section 3, for the bivariate distribution of $\mathrm{ADV}(I, \alpha)$ where $\alpha = [k_\alpha, k_\alpha]'$, with k_α chosen so that maximum directional AG skewness equals $\frac{1}{2}$, and direction $d = [\cos(\theta), \sin(\theta)]'$. The shape of the curves in Figure 3 reveals the process used to generate the ADV class of distributions, namely that skewness is modified around one single direction, parameterised by $\frac{\alpha}{\|\alpha\|}$. Varying the direction of α, whilst keeping $\|\alpha\|$ constant does not change the shape of the curves in Figure 3, but only their location. Varying $\|\alpha\|$, whilst keeping the direction α constant, produces curves of a similar shape but with a different scale.

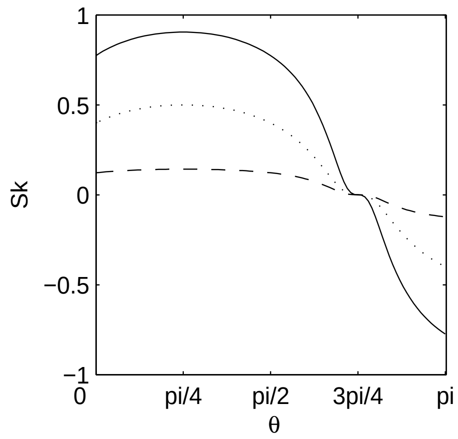

Figure 3: Directional skewness for a bivariate distribution of $\mathrm{ADV}(I, \alpha)$ as a function of θ, where $d = [\cos(\theta), \sin(\theta)]'$, and for the CE (solid), B (dashed) and AG(dotted) measures of univariate skewness.

4.2 The FS-Normal class

Ferreira and Steel (2004) introduced a class of skewed normal distributions based on linear transformations of univariate variables with independent, potentially skewed, distributions. The authors studied the case where the univariate skewed distributions are of the form discussed in Fernández and Steel (1998). Here we analyse their skewed version of the Normal distribution.

Let A be an $m \times m$ non-singular matrix and $\gamma = (\gamma_1, \dots, \gamma_m) \in \Re_+^m$. Then, $X \in \Re^m$ has an FS-Normal distribution, denoted by $\mathrm{FS}(A, \gamma)$, if its density is of the form

$$f_{FS}(x|A, \gamma) = \|A\|^{-1} \prod_{j=1}^{m} p(x' A_{\cdot j}^{-1} | \gamma_j), \tag{6}$$

where $A_{\cdot j}^{-1}$ denotes the j-th column of A^{-1}, $\|A\|$ denotes the absolute value of the determinant of A, and $p(\cdot)$ in (6) is given by

$$p(\epsilon_j|\gamma_j) = \frac{2}{\gamma_j + \frac{1}{\gamma_j}} \left\{ \phi(\gamma_j \epsilon_j) I_{(-\infty,0)}(\epsilon_j) + \phi\left(\frac{\epsilon_j}{\gamma_j}\right) I_{[0,\infty)}(\epsilon_j) \right\}, \tag{7}$$

with $I_S(\cdot)$ the indicator function on S.

For any A and γ, the distribution $\mathrm{FS}(A, \gamma)$ is unimodal and the mode is at zero.

THEOREM 4. *Let $X \sim FS(A,\gamma)$ and d be a vector in $\mathcal{S}^{m-1}$. Then, the density of the directional distribution of X along d is given by*

$$f_{FS(A,\gamma),d}(y) = \frac{2b_{1,d}b_{2,d}}{b_{1,d}+b_{2,d}} \left\{\phi(yb_{1,d})I_{(-\infty,0]}(y) + \phi(yb_{2,d})I_{(0,\infty)}(y)\right\} \tag{8}$$

where

$$b_{1,d} = \left[\sum_{j=1}^{m} \left(d'A_{\cdot j}^{-1}\right)^2 \gamma_j^{2sign(d'A_{\cdot j}^{-1})}\right]^{1/2} \qquad b_{2,d} = \left[\sum_{j=1}^{m} \left(d'A_{\cdot j}^{-1}\right)^2 \gamma_j^{-2sign(d'A_{\cdot j}^{-1})}\right]^{1/2}, \tag{9}$$

with $sign(\cdot)$ denoting the usual sign function.

A closer look reveals that (8) reverts to (7) when $b_{1,d} = \gamma_j$ and $b_{2,d} = \frac{1}{\gamma_j}$. Characterising univariate skewness of the distribution with density (8) using the measures introduced in Section 2 is straightforward. The moments of the distribution are given by

$$E\left[X^n|b_{1,d}, b_{2,d}\right] = \frac{2^{n/2}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}} \frac{b_{1,d}^{n+1} + (-1)^n b_{2,d}^{n+1}}{(b_{1,d}b_{2,d})^n(b_{1,d}+b_{2,d})}.$$

Calculating the CE measure is then immediate. For the B measure, only $\Phi(\cdot)$ is necessary. Finally, the AG measure is given by

$$AG\left[F_{\mathrm{FS}(A,\gamma),d}\right] = \frac{b_{1,d} - b_{2,d}}{b_{1,d} + b_{2,d}}.$$

Invariance of the measures of univariate skewness to scale transformations implies that the skewness of the distributions with density as in (8) is equivalent to that of (7) with $\gamma_j = \gamma_d = \sqrt{\frac{b_{1,d}}{b_{2,d}}}$. Thus, maximum values of directional CE skewness are the same as for the univariate skew-Normal of Fernández and Steel (1998) and coincide with the maximum directional CE skewness values of the ADV-Normal class. Figure 4 presents the three measures of univariate skewness as functions of γ_d. All the measures are strictly increasing functions of γ_d and are zero for $\gamma_d = 1$.

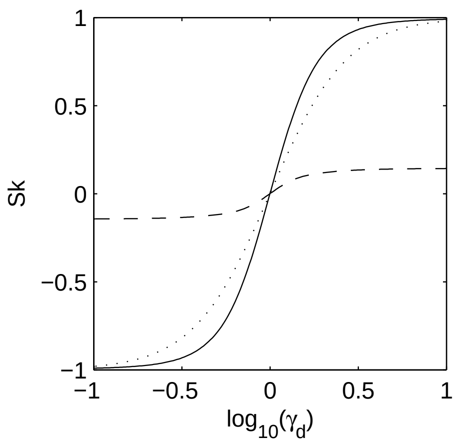

Figure 4: FS-Normal distribution: CE (solid), B (dashed) and AG (dotted) measures of directional skewness as functions of γ_d.

4.2.1 Special case $A = \sigma O$

Using Theorem 2, we can drop the constant σ and restrict our attention to the case when $A = O$, where O is an m-dimensional orthogonal matrix. Denoting the j-th row of matrix O by O_j we simply replace $d'A^{-1}_{\cdot j}$ by $O_j d$ in (9) to obtain $b_{1,d}$ and $b_{2,d}$. As both the rows of O and d have unitary norm, $|\log(\gamma_d)|$ takes maximum value when $d = \pm O'_{j^*}$, where $j^* \in \{1, \ldots, m\}$ is the index of the component of γ with largest absolute value of its logarithm. Following a similar argument, the axes of skewness are immediately identified as defined by the rows of O.

In Subsection 4.1.1 we analysed directional skewness for a bidimensional example of an ADV-Normal distribution with maximum directional skewness fixed and AG skewness along that axis equal to $\frac{1}{2}$. With fixed $\Sigma = I$, there were no more free parameters. We now perform a similar analysis for the FS-Normal class. Fixing the axes of skewness is equivalent to fixing the matrix O, and for simplicity we fix $O = I$. Selecting the first row of O as defining the axis along which skewness is maximal and AG skewness is equal to $\frac{1}{2}$, implies that $\gamma_1 = \sqrt{3}$. Choosing $|\log(\gamma_2)| < \log(\gamma_1)$ ensures that the direction along which skewness is maximal is left unchanged. Using $d = [\cos(\theta), \sin(\theta)]'$, directional skewness can then be examined as a function of both θ and γ_2.

Figure 5(a) shows a greyscale plot of the AG directional skewness. Varying γ_2 has a large effect on directional skewness. When $\gamma_2 = 1$, corresponding to a similar case as the one studied in Subsection 4.1.1, skewness is concentrated on directions close to the one defined by the first row of O, corresponding to $\theta = 0$. By increasing $|\log(\gamma_2)|$ the colour tones in the plot are made more extreme, indicating that there are bigger regions with high directional skewness. This is also shown in Figure 5(b), where MDS has minimum value when $\gamma_2 = 0$. In contrast with the ADV-Normal class, FS-Normal parameterises skewness using not one but m directions, given by the rows of O, and m scalars to model the amount of skewness, given by the elements of γ. This results in greater flexibility to describe phenomena in which skewness is not (mainly) manifested along one single direction.

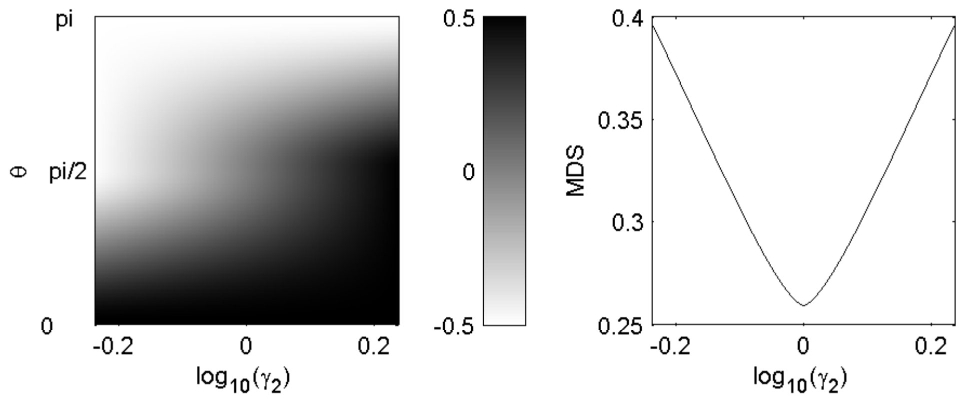

Figure 5: (a) Greyscale plot of the directional skewness, using the AG measure, for a bivariate distribution of FS(I, γ) as a function of θ, where $d = [\cos(\theta), \sin(\theta)]'$, and γ_2. (b) MDS as a function of γ_2.

5. ILLUSTRATION

We use 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). The data were collected for n =202 athletes at the Australian Institute of Sport and are described in Cook & Weisberg (1994). The dataset also contains information on three covariates: red cell count (RCC), white cell count (WCC) and plasma ferritin concentration (PFC).

These data have been used previously for the illustration of the use of skewed distributions. Azzalini & Capitanio (1999) used them without covariates, while Ferreira & Steel (2004) used the complete data in a linear regression model. We will use three datasets, differing in the number of variables included. The first dataset, denoted 2D, contains the variables BMI and PBF. The 3D dataset contains BMI, PBF and SSF. Finally, 4D is the complete dataset. In all cases we use the covariates, normalised to have mean zero and variance one. A constant term is also included.

5.1 Regression models

We consider n observations, each of which is given as a pair (y_i, x_i), $i = 1, \dots, n$. For each i, $y_i \in \Re^m$ represents the variable of interest and $x_i \in \Re^k$ is a vector of covariates. Thus, in the context of our application, y_i will be (a subset of) the four biomedical variables with m ranging from 2 to 4, and x_i will be the three-dimensional vector of covariates mentioned above. Throughout, we condition on x_i without explicit mention in the text.

We assume that the process generating the variable of interest can be described by independent sampling for $i = 1, \dots, n$ from the linear regression model

$$y_i = B'x_i + \eta_i,$$

where B is a $k \times m$ matrix of real regression coefficients, and $\eta_i \in \Re^m$ has a distribution of one of three possible forms: Normal with mean zero and variance Σ, ADV(Σ,α) as in Subsection 4.1 or FS(A,γ) as in Subsection 4.2.

5.2 Prior distributions

For the Normal model, we adopt the usual matrix-variate Normal-inverted Wishart prior on B and Σ, with parameters $B_0 \in \Re^{k \times m}$, M and Q covariance matrices with dimension k and m respectively, and v a positive constant, with density given by

$$p(B|\Sigma) \propto |M|^{-\frac{m}{2}} |\Sigma|^{-\frac{k}{2}} \exp\left[-\frac{1}{2} \text{ tr } \Sigma^{-1}(B - B_0)'M^{-1}(B - B_0)\right] \tag{10}$$

and

$$p(\Sigma) \propto |Q|^{\frac{v}{2}} |\Sigma|^{-\frac{m+v+1}{2}} \exp\left(-\frac{1}{2} \text{ tr } \Sigma^{-1}Q\right), \tag{11}$$

where tr denotes the trace operation.

The prior distributions on the parameters of the ADV- and FS-Normal models are defined taking two characteristics into consideration. The first is that they match the prior for the Normal case when α and γ have all components equal to zero and one, respectively, *i.e.* when the skewed models simplify to the symmetric distribution. The second assumption is that there is no prior information available on the direction of the distribution, *i.e.* the prior is invariant under orthogonal transformations.

In order to satisfy the first requirement we assume that $P_{B,\Sigma,\alpha} = P_\alpha P_{B,\Sigma}$ and $P_{B,A,\gamma} = P_\gamma P_{B,A}$. For the ADV-Normal case, the prior of B and Σ is simply given by (10)-(11). Strictly speaking, the prior for the ADV case should take into account the fact that not all pairs (Σ, α) are allowed, which would imply prior dependence between Σ and α. However, in our empirical application, these restrictions are not of practical importance and, thus, will be ignored here. Imposing them can simply be done by an extra rejection condition in the sampler used for inference (see next subsection).

The second characteristic imposes that Q in (11) must be of the form qI, with $q > 0$. To set the prior of B and A for the FS-Normal model, Ferreira & Steel (2004) considered the decomposition $A = OU$, where O is an m-dimensional orthogonal matrix and U is an upper triangular matrix with positive diagonal elements $u_{jj}, j = 1, \ldots, m$, and defined $\Sigma = A'A = U'U$. The prior on B and A is then given by the prior on B as in (10), given $\Sigma = U'U$, and a prior on O and U with density

$$p(O,U) \propto p(O)|Q|^{\frac{v}{2}} \prod_{j=1}^{m} u_{jj}^{m-j} |U|^{-(m+v)} \exp\left[-\frac{1}{2} \text{ tr } (U'U)^{-1}Q\right],$$

where $p(O)$ is the density on the set of m-dimensional orthogonal matrices invariant to linear orthogonal transformations (known as the Haar density).

The second characteristic imposed on the prior also implies that the prior on α and γ must be exchangeable. The simplest way to achieve this is to have $P_\alpha = \prod_{j=1}^{m} P_{\alpha_j}$, $P_\gamma = \prod_{j=1}^{m} P_{\gamma_j}$, with P_{α_j} and P_{γ_j} equal for all $j = 1, \ldots, m$.

We select P_{α_j} and P_{γ_j} based on directional skewness arguments, quantified using the AG measure defined in Section 2. As the prior structure is invariant under orthogonal transformations, the prior on directional skewness is the same for any direction. Let this prior be denoted by P_{AG}. We then choose P_{α_j} and P_{γ_j} so as to induce a prior on directional skewness that is closest, with respect to some distance function, to P_{AG}. We highlight the fact that both P_Σ and P_A have an effect on the prior of directional skewness. Therefore, we select P_{α_j} and P_{γ_j} conditional on P_Σ and P_A, respectively.

In this article, we assume that P_{AG} is a unimodal symmetric distribution with mode at zero, corresponding to a prior that puts identical mass on left and right skewness, concentrating most of the prior mass around symmetric directional distributions. We suggest a Beta prior on AG with both parameters equal to $a > 0$, rescaled to the interval (-1,1). As the value of a increases, the mass assigned by P_{AG} to heavily skewed distributions decreases.

Student-t priors with zero mean were chosen for α_j and $\log(\gamma_j)$, with the respective variances and degrees of freedom determined as to best approximate P_{AG}, using a Kullback-Leibler measure as suggested in Ferreira and Steel (to appear).

5.3 Inference

The hyperparameter B_0 is set to be the $k \times m$ zero matrix, $M = 100I_k$, $Q = I_m$ and $v = m + 2$. These settings correspond to a rather vague prior.

Inference is conducted using Markov chain Monte Carlo methods (MCMC). For brevity, we omit the details of the samplers. These can be obtained from the authors, as well as a Matlab implementation. MCMC chains of 120,000 iterations were used, retaining every 10th sample after a burn-in period of 20,000 draws.

We make use of Bayes factors to assess the relative adequacy of each model. Estimates of marginal likelihood are obtained using the p_4 measure in Newton & Raftery (1994), with their $\delta = 0.1$.

5.4 Results

5.4.1 Bayes factors

We start the analysis of the different problems by comparing the models using Bayes factors. Table 1 presents the logarithm of the Bayes factors for the different models with respect to the Normal alternative. Each row corresponds to a particular dimension. In all three problems, the skewed models were shown to be far superior to the symmetric one, with the difference between them increasing with the dimensionality of the space. When comparing the two skewed models, FS always outperforms ADV. In the remaining part of this section, we analyse how information about directional skewness can help to explain the different performance of the models. We restrict our attention to AG directional skewness but a similar analysis can easily be performed using any of the other skewness measures.

Table 1: Log of Bayes factors for the different skewed models with respect to the symmetric Normal.

Dimension	ADV	FS
2D	28.23	31.38
3D	31.05	39.38
4D	38.15	48.07

5.4.2 Skewness characterisations

For the two-dimensional problem, we can easily visualise the directional skewness for every direction. Figure 6 presents the mean posterior directional skewness as a function of θ, parameterising direction $d = [\cos(\theta), \sin(\theta)]'$. The first conclusion that can be drawn is that, as expected from the Bayes factors in Table 1, both skewed models lead to rather skewed distributions. The FS model puts a substantial amount of skewness, almost constant, in large intervals of θ, and makes a sharp transition between positive and negative skewness. The directional skewness for the ADV model increases more gradually and then decreases immediately. This shows that FS leads to an overall more skewed distribution than ADV. One interesting similarity between the models is that they both have maximum skewness, in absolute value, in similar directions. These findings are in close agreement with the characteristics of the two classes analysed in Section 4. The ADV model manages to capture adequately the most skewed part of the distribution, but in order to do so, employs all of its parameters, Σ and α (norm for amount and orientation for location of skewness). The FS model can induce skewness in a broader region. This greater flexibility is the result of employing two directions for the location of skewness besides two scalar parameters for the amount of skewness.

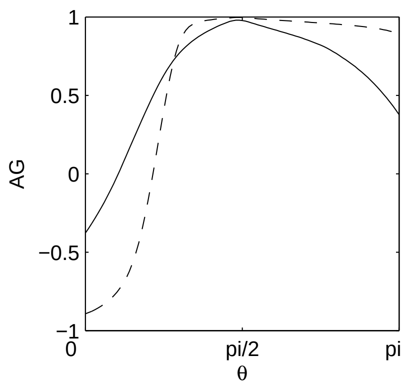

Figure 6: AG directional skewness for the 2D problem as a function of θ, where $d = [\cos(\theta), \sin(\theta)]'$, for the ADV- (solid) and FS-Normal (dashed) models.

For the higher dimensional problems, visualising directional skewness is not a simple task and we resort to summaries of directional skewness, namely to MDS and to DMDS. Figure 7 presents the posterior density of MDS for all models and for the three different dimensions. Note that MDS has values in the space of $|\mathrm{AG}|$, namely $[0, 1]$. The plot in 7(a) confirms the information provided by Figure 6, with the posterior mass of MDS more concentrated on large values for the FS model. The densities of MDS for the two other dimensions reveal quite distinct patterns. For the 3D problem, FS has most mass concentrated around MDS=0.2, whilst ADV concentrates mass around 0.75. The picture for the 4D problem is much closer to the one for the 2D problem, with the distribution of MDS being more concentrated on larger values for FS.

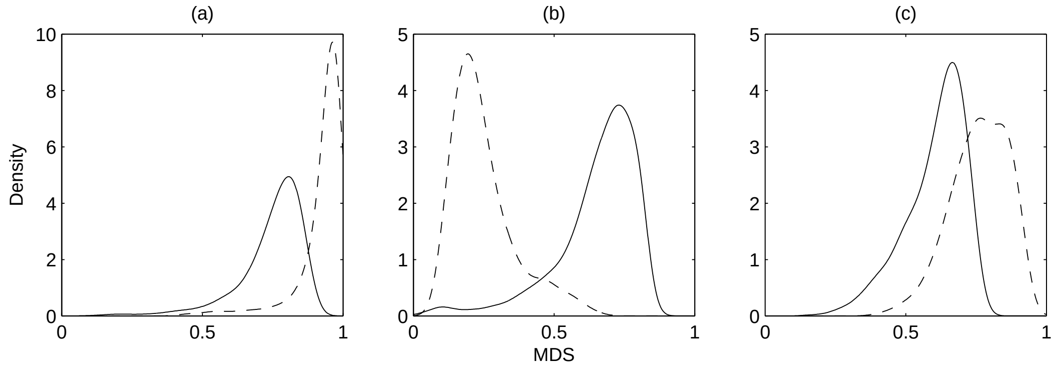

Figure 7: Posterior density of MDS for the 2D, 3D and 4D problems, respectively (a),(b) and (c). The solid line stands for the ADV models and the dashed line for the FS models.

Similar results are provided by the amounts of AG skewness along each one of the principal axes of skewness, choosing the l_∞ norm for $\mathcal{F}$ in Definition 2. Table 2 presents characteristics for the posterior distribution of these values for all problems. Heading AG_j stands for the amount of AG skewness along the j^{th} principal axis of skewness, ordered so that $\mathrm{AG}_i \geq \mathrm{AG}_j$, if $i < j$. For the 2D problem, AG_1 has similar values for both skewed models, with differences appearing for AG_2, where the statistics for FS have a larger value than for ADV. Inference for the 3D dataset exhibits differences for all three quantities. ADV leads to larger values than FS, with the difference being particularly evident for AG_2 and AG_3. These differences are replicated in DMDS. Lastly, for the 4D problem, FS leads to larger values than ADV. In this case, we call attention to the fact that AG_1, AG_2 and AG_3 have most mass close to one.

With the results on directional skewness that we have presented so far, it is possible to obtain a fairly comprehensive description of the skewed models. We now try to assess the reason for the differences between them. A useful tool is provided by plotting the residuals of the regression. Figure 8 presents the pairwise scatter plots for the residuals of the FS model corresponding to the modal values of the posterior. Plots obtained for the ADV models and/or for the more restricted datasets are similar.

The scatter plot between PBF and SSF provides the explanation for the low MDS and AG_2 and AG_3 values for the FS model in the 3D problem. As these two variables are very strongly correlated, they can both be captured by the same axis of skewness. As a consequence, the average skewness decreases when we go from the 2D to the 3D case. The ADV model does not seem to be able to account for this correlation in a fully adequate manner, as it basically induces skewness in a single direction. Thus, the ADV model focuses mainly on the most skewed direction of the distribution.

In the 4D dataset, the introduction of LBM brings additional skewness into the distribution,

Table 2: Characteristics of the posterior distribution of the amount of skewness along the principal axes of skewness, and of the posterior distribution of DMDS.

		ADV				FS			
Dimension	Measure	10%	Mean	Median	90%	10%	Mean	Median	90%
2D	AG_1	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
	AG_2	0.07	0.57	0.66	0.85	0.79	0.91	0.97	0.99
	DMDS	0.52	0.77	0.82	0.92	0.89	0.95	0.98	0.99
3D	AG_1	0.96	0.97	0.98	0.98	0.71	0.87	0.88	0.99
	AG_2	0.67	0.81	0.86	0.92	0.19	0.49	0.42	0.87
	AG_3	0.07	0.55	0.64	0.83	0.03	0.31	0.19	0.77
	DMDS	0.59	0.78	0.82	0.91	0.34	0.56	0.48	0.87
4D	AG_1	0.88	0.93	0.94	0.96	0.99	0.99	1.00	1.00
	AG_2	0.78	0.87	0.89	0.93	0.86	0.93	0.94	0.98
	AG_3	0.70	0.83	0.87	0.91	0.85	0.92	0.93	0.97
	AG_4	0.01	0.32	0.32	0.64	0.20	0.58	0.63	0.87
	DMDS	0.61	0.74	0.75	0.85	0.74	0.85	0.87	0.95

as can be seen by the pairwise scatter plots against the other variables. There are three different patterns of skewness (BMI vs. PBF/SSF, BMI vs. LBM and PBF/SSF vs. LBM). To model the joint distribution of the variables, both models employ distributions that have large values of MDS and AG_j, especially for $j = 1, 2, 3$. In addition, for this problem, both AG_4 and DMDS are higher for the FS model. This could be explained by the necessity of the skewed distribution to model also the interactions between BMI, PBF/SSF and LBM in the 4D space, not visible in Figure 8.

5.4.3 Predictive results

An important quality of a model is that it can provide useful forecasts. Thus, besides Bayes factors, predictive performance provides an additional benchmark for evaluating the model's adequacy. We consider predicting the observable y_f given the corresponding values of the regressors, grouped in a k-dimensional vector x_f.

Prediction naturally fits in the Bayesian paradigm as all parameters can be integrated out, formally taking parameter uncertainty into account.

We partition the sample into n observations on which we base our posterior inference and q observations which we retain in order to check the predictive accuracy of the model. As a formal criterion, we use the log predictive score (LPS), introduced by Good (1952). This is a strictly proper scoring rule, in the sense that it induces the forecaster to be honest in divulging her predictive distribution. For $f = n+1, \ldots, n+q$ (*i.e.*, for each observation in the prediction sample) we base our measure on the predictive density evaluated in these retained observations $y_{n+1}, \ldots, y_{n+q}$, namely:

$$LPS = -\frac{1}{q} \sum_{f=n+1}^{n+q} \ln p(y_f|Y),$$

where $Y = (y_1, \ldots, y_n)$ denotes the inference sample. Smaller values of LPS correspond to better performance in forecasting the prediction sample. The interpretation of LPS can be facilitated by considering that in the case of i.i.d. sampling, LPS approximates the sum of the Kullback-Leibler divergence between the actual sampling density and the predictive density and the entropy of the sampling distribution (Fernández, Ley and Steel 2001). So LPS captures uncertainty due to a lack of fit plus the inherent sampling uncertainty. Here we are faced with a different x_f for every

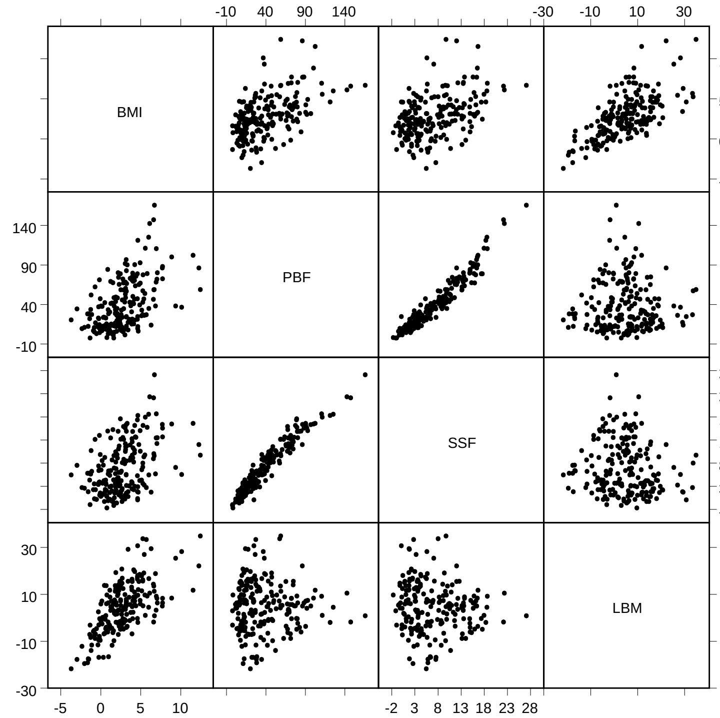

Figure 8: Pairwise scatter plots for the residuals of the FS model corresponding to the posterior mode.

forecasted observation, so we are not strictly in the i.i.d. framework, but the above interpretation may still shed some light on the calibration and comparison of LPS values.

Table 3 shows LPS values for our cross-validation analysis. It is based on five different partitions of the data leading to prediction sets of approximately the same size (three sets with $q = 40$ observations to predict and two with $q = 41$). The two skewed models under consideration were applied to each of these data sets. There are only marginal differences. Analysing the average LPS across the five cases, the ADV-Normal proved slightly more adequate for the 4D problem and the FS-Normal did slightly better for the 2D and 3D problems.

6. CONCLUSION

In this paper we introduce the concept of directional skewness, defined as the skewness along a particular direction, and study how this concept can facilitate the description and comparison of classes of multivariate skewed distributions. Focusing on a given direction d (through a conditional distribution) allows us to use univariate skewness measures to quantify directional skewness for any d. In contrast with existing measures of overall skewness, directional skewness will generally be affected by linear transformations. Whereas the latter single measures were primarily developed to test for symmetry, our directional skewness measure is intended to characterise the skewness properties of multivariate distributions. The full analysis of directional skewness completely describes the skewness of a distribution as a function of direction. We also suggest an alternative based on studying skewness along specific directions, given by the principal axes of skewness.

We analyse in detail two skewed classes of distributions based on the Normal distribution that

Table 3: LPS for the two skewed models applied to five predictive subsets.

	2D		3D		4D	
Folder	ADV	FS	ADV	FS	ADV	FS
1	6.79	6.95	8.59	8.93	12.02	12.40
2	7.08	7.19	8.93	8.71	12.34	12.09
3	6.98	6.80	8.63	8.51	11.73	11.84
4	7.03	7.03	8.77	8.79	12.02	12.06
5	6.94	6.75	8.84	8.58	11.88	11.75
Average	6.96	6.94	8.75	8.70	12.00	12.03

have appeared in the literature: the skew-Normal of Azzalini and Dalla Valle (1996) and the skew-Normal of Ferreira and Steel (2004). For these classes, it is possible to find simple forms for the directional distributions, which are shown to be slight generalisations of the univariate versions of these distributions. This allows for the complete analysis of directional skewness. A similar treatment is immediately applicable to classes of distributions that are generated as scale mixtures of either of these skew-Normal distributions.

We conduct Bayesian inference on regression problems of different dimension using the two classes of skewed distributions, along with a symmetric Normal model. Based on directional skewness arguments, we define prior distributions which are invariant under orthogonal transformations, representing prior ignorance about the direction of the skewness. In our application to biomedical variables we model distributions of dimensions 2,3 and 4, and find strong evidence for asymmetry, in line with earlier findings for these data. A complete and informative description of the skewness in terms of directional skewness is used to illustrate how the models differ in modelling skewness and is related to the specific properties of the data. We feel this leads to a better understanding of the reasons for the differences we find between models. Finally, we conduct a predictive analysis of the skewed models and find they are very similar in predictive performance.

The analysis of directional skewness also suggests a new approach to the definition of skewed distributions. One alternative that arises naturally is to model directional skewness explicitly through suitable functions of the direction. This is further developed in Ferreira and Steel (2005) and is the focus of current research.

APPENDIX

Proof of Theorem 1. Immediate. □

Proof of Theorem 2. If F has mode at μ^*, then the distribution of $X+a$ has mode at μ^*+a and $Y=(O^d)'(X-\mu^*)=(O^d)'[X+a-(\mu^*+a)]$. Thus, we have invariance to location transformations.

The distribution of $k_1X \sim H$ has mode at $k_1\mu^*$. Then $Sk_m(H,d) = Sk(G_{k_1y_1|y_1=0}) = Sk(G_{y_1|y_1=0})$, with the last equality following from the fact that the univariate measure of skewness is invariant to scale transformations. This establishes the result. □

Proof of Theorem 3. If $X \sim$ADV(Σ,α) has mode at μ^*, then $Z=X-\mu^*$ has mode at zero. Now, the density of Z is

$$f(z)=2\phi_m(z+\mu^*|0,\Sigma)\Phi[\alpha'(z+\mu^*)].$$

Now if $Y = (O^d)'Z$,

$$\begin{aligned} f(y_1|y_{-1}=0) &\propto \phi_m(dy_1+\mu^*|0,\Sigma)\Phi[\alpha'(dy_1+\mu^*)] \\ &\propto e^{-\frac{1}{2}(dy_1+\mu^*)'\Sigma^{-1}(dy_1+\mu^*)}\Phi[y_1\alpha'd+\alpha'\mu^*)] \\ &\propto \phi\left(\frac{y_1-\mu_d}{\sigma_d}\right)\Phi\left(\delta_{0,d}+\delta_{1,d}\frac{y_1-\mu_d}{\sigma_d}\right). \end{aligned}$$

Now, from (2.4) in Arnold & Beaver (2002), we obtain the integrating constant. □

Proof of Theorem 4. If $X \sim$FS(A,γ) and $Y = (O^d)'X$ then the density of Y is

$$f(y) \propto \prod_{j=1}^{m} p\left[y'(O^d)'A_{\cdot j}^{-1}|\gamma_j\right]$$

and, as such,

$$f(y_1|y_{-1}=0) \propto \prod_{j=1}^{m} p[y_1 d' A_{\cdot j}^{-1}|\gamma_j].$$

Now, simple manipulation reveals that

$$f(y_1|y_{-1}=0) \propto e^{-\frac{y_1^2}{2}\left[b_{1,d}^2 I_{(-\infty,0]}(y_1)+b_{2,d}^2 I_{(0,\infty)}(y_1)\right]}.$$

The proof follows by calculating the integral of $f(y_1|y_{-1}=0)$ for $y_1 \in \Re$. □

ACKNOWLEDGEMENTS

We thank two anonymous referees and an Associate Editor for insightful and helpful comments, which led to improvements in the paper. Part of the work of the first author was supported by a grant from Fundação para a Ciência e Tecnologia, Ministério para a Ciência e Tecnologia, Portugal. The work of the first author was conducted while he was affiliated to the University of Warwick.

REFERENCES

B. C. Arnold & R. J. Beaver (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11, 7–54.

B. C. Arnold & R. A. Groeneveld (1995). Measuring skewness with respect to the mode. *The American Statistician*, 49, 34–38.

A. Azzalini (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.

A. Azzalini & A. Capitanio (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B*, 61, 579–602.

A. Azzalini & A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.

A. L. Bowley (1920). *Elements of Statistics.* 4th edn, Scribner, New York.

C. V. L. Charlier (1905). Über das Fehlergesetz. *Archiv für Mathematik, Astronomi och Fysik*, 2, 8.

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

F. Y. Edgeworth (1904). The law of error. *Transactions of the Cambridge Philosophical Society*, 20, 26–65 and 113–114.

C. Fernández, E. Ley & M. F. J. Steel (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100, 381–427.

C. Fernández & M. F. J. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93, 359–371.

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

J. T. A. S. Ferreira & M. F. J. Steel (2005). Modelling directional dispersion through hyperspherical log-splines. *Journal of the Royal Statistical Society Series B*, 67, 599–616.

J. T. A. S. Ferreira & M. F. J. Steel (to appear). Model comparison of coordinate-free multivariate skewed distributions. *Journal of Econometrics*.

I. J. Good (1952). Rational decisions. *Journal of the Royal Statistical Society Series B*, 14, 107–114.

N. Henze (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers*, 43, 467–506.

M. C. Jones & R. Sibson (1987). What is projection pursuit?. *Journal of the Royal Statistical Society Series A*, 150, 1–36.

J. F. Malkovich & A. A. Afifi (1973). On tests for multivariate normality. *Journal of the American Statistical Association*, 68, 176–179.

K. V. Mardia (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530.

T. F. Móri, V. K. Rohatgi & G. J. Székely (1993). On multivariate skewness and kurtosis, *Theory of Probability and Its Applications*, 38, 547–551.

M. A. Newton & A. E. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society Series B*, 56, 3–48.

H. Oja (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, 8, 154–168.

H. Oja (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1, 327–332.

S. Sahu, D. K. Dey & D. Branco (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, 31, 129–150.

Received 1 July 2005
Accepted 8 March 2006

José T. A. S. FERREIRA: `tome.ferreira@endeavour-capital.com`
Endeavour Capital Management

London
United Kingdom

Mark F. J. STEEL: m.f.steel@stats.warwick.ac.uk
Department of Statistics, University of Warwick
Coventry
United Kingdom CV4 7AL