

# A unified view on skewed distributions arising from selections

Reinaldo B. ARELLANO-VALLE, Márcia D. BRANCO and Marc G. GENTON

*Key words and phrases:* Kurtosis; multimodal distribution; multivariate distribution; nonnormal distribution; selection mechanism; skew-symmetric distribution; skewness; transformation.

*MSC 2000:* Primary 60E05; secondary 62E10.

*Abstract:* Parametric families of multivariate nonnormal distributions have received considerable attention in the past few decades. The authors propose a new definition of a selection distribution that encompasses many existing families of multivariate skewed distributions. Their work is motivated by examples that involve various forms of selection mechanisms and lead to skewed distributions. They give the main properties of selection distributions and show how various families of multivariate skewed distributions, such as the skew-normal and skew-elliptical distributions, arise as special cases. The authors further introduce several methods of constructing selection distributions based on linear and nonlinear selection mechanisms.

## Une perspective intégrée des lois asymétriques issues de processus de sélection

*Résumé :* Les familles paramétriques de lois multivariées non gaussiennes ont suscité beaucoup d'intérêt depuis quelques décennies. Les auteurs proposent une nouvelle définition du concept de loi de sélection qui englobe plusieurs familles connues de lois asymétriques multivariées. Leurs travaux sont motivés par diverses situations faisant intervenir des mécanismes de sélection et conduisant à des lois asymétriques. Ils mentionnent les principales propriétés des lois de sélection et montrent comment diverses familles de lois asymétriques multivariées telles que les lois asymétriques normales ou elliptiques émergent comme cas particuliers. Les auteurs présentent en outre plusieurs méthodes de construction de lois de sélection fondées sur des mécanismes linéaires ou non linéaires.

## 1. INTRODUCTION

Since the collection and modelling of data plays a fundamental role in scientific research, multivariate distributions are central to statistical analyses. Indeed, multivariate distributions serve to model dependent outcomes of random experiments. Despite the key role of the classical multivariate normal distribution in statistics, there has been a sustained interest among statisticians in constructing nonnormal distributions. In recent years, significant progress has been made toward the construction of so-called multivariate skew-symmetric and skew-elliptical distributions and their successful application to problems in areas such as engineering, environmetrics, economics, and the biomedical sciences; see the book edited by Genton (2004a) for a first and unique synthesis of results and applications in this topic.

These multivariate skew-elliptical distributions and other related versions are essential for statistical modelling for two main reasons. Firstly, they provide flexible parametric classes of multivariate distributions for the statistician's toolkit that allow for modelling key characteristics such as skewness, heavy tails, and in some cases, surprisingly, multimodality. They also contain the multivariate normal distribution as a special case, enjoy pleasant theoretical properties, and do not involve the often difficult task of multivariate data transformation. Secondly, they appear in the natural and important context of selection models, that is, when latent variables influence the data collected. We start by discussing these aspects in the next two sections.

### 1.1. Parametric multivariate nonnormal distributions.

In the last few decades, applications from environmental, financial, and the biomedical sciences, among other fields, have shown that data sets following a normal law are more often the exception rather than the rule. Therefore, there has been a growing interest in the construction of

parametric classes of nonnormal distributions, particularly in models that can account for skewness and kurtosis. A popular approach to model departure from normality consists of modifying a symmetric (normal or heavy tailed) probability density function of a random variable/vector in a multiplicative fashion, thereby introducing skewness. This idea has been in the literature for a long time (e.g., Birnbaum 1950; Nelson 1964; Weinstein 1964; Roberts 1966; O'Hagan & Leonard 1976). But it was Azzalini (1985, 1986) who thoroughly implemented this idea for the univariate normal distribution, yielding the so-called skew-normal distribution. An extension to the multivariate case was then introduced by Azzalini & Dalla Valle (1996). Statistical applications of the multivariate skew-normal distribution were presented by Azzalini & Capitanio (1999), who also briefly discussed an extension to elliptical densities. Arnold, Castillo & Sarabia (1999, 2001) discussed conditionally specified distributions related to the above construction for skewness.

Since then, many authors have tried to generalize these developments to skewing arbitrary symmetric probability density functions with very general forms of multiplicative functions. Some of those proposals even allow for the construction of multimodal distributions. In the present article we will provide an overview of existing distributions and novel extensions from a unified point of view.

Besides the appeal of extending the classical multivariate normal theory, the current impetus in this field of research is its potential for new applications. For example, a skewed version of the *GARCH* model (generalized autoregressive conditional heteroscedasticity) in economics, the representation of the shape of a human brain with skewed distributions in image analysis, or a skewed Kalman filter for the analysis of climatic time series, are just but a few examples of these many applications; see Genton (2004a) for those and more.

### 1.2. Selection mechanisms.

In addition to the attractive features of the above distributions for parametric modelling of multivariate nonnormal distributions (a flexible parametric alternative to multivariate nonparametric density estimation), there is also a natural and important context where these distributions appear, namely selection models (see Heckman 1979, and more recently Copas & Li 1997). Indeed, consider the model where a random vector  $\mathbf{X} \in \mathbb{R}^p$  is distributed with symmetric probability density function  $f(\mathbf{x}; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a vector of unknown parameters. The usual statistical analysis assumes that a random sample  $\mathbf{X}_1, \dots, \mathbf{X}_n$  from  $f(\mathbf{x}; \boldsymbol{\theta})$  can be observed in order to make inference about  $\boldsymbol{\theta}$ . Nevertheless, there are many situations where such a random sample might not be available, for instance if it is too difficult or too costly to obtain. If the probability density function is distorted by some multiplicative nonnegative weight function  $w(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\eta})$ , where  $\boldsymbol{\eta}$  denotes some vector of additional unknown parameters, then the observed data is a random sample from a weighted distribution (see, e.g., Rao 1985) with probability density function

$$f(\mathbf{x}; \boldsymbol{\theta}) \frac{w(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\eta})}{E\{w(\mathbf{X}; \boldsymbol{\theta}, \boldsymbol{\eta})\}}. \quad (1)$$

In particular, if the observed data are obtained only from a selected portion of the population of interest, then (1) is called a selection model. For example, this can happen if the vector  $\mathbf{X}$  of characteristics of a certain population is measured only for individuals who manifest a certain disease due to cost or ethical reasons; see the survey article by Bayarri & DeGroot (1992) and references therein. If the weight function satisfies  $E\{w(\mathbf{X}; \boldsymbol{\theta}, \boldsymbol{\eta})\} = 1/2$ , the resulting probability density function (1) has a normalizing constant that does not depend on the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ . A weight function with such a property can naturally occur when the selection criterion is such that a certain variable of interest is larger than its expected value given other variables.

A simple example of selection is given by dependent maxima. Consider two random variables with bivariate normal distribution, mean zero, variance one, and correlation  $\rho$ . The distribution of the maximum of those two random variables is exactly skew-normal (Roberts 1966;

Loperfido 2002) with skewness depending on  $\rho$ . Note that when  $\rho = 0$ , the distribution of the maximum is still skewed. For instance, imagine students who report only their highest grade between calculus and geometry, say. The resulting distribution will be skewed due to this selective reporting, although the joint bivariate distribution of the two grades might be normal or elliptically contoured. The distribution of the minimum of the two random variables above is also skew-normal, but with a negative sign for the skewness parameter.

Another example of selection arises in situations where observations obey a rather well-behaved law (such as multivariate normal) but have been truncated with respect to some hidden covariables; see, e.g., Arnold & Beaver (2000a, 2002, 2004), Capitanio, Azzalini & Stanghellini (2003). A simple univariate illustration is the distribution of waist sizes for uniforms of airplane crews who are selected only if they meet a specific minimal height requirement. A bivariate normal distribution might well be acceptable to model the joint distribution of height and waist measurements. However, imposition of the height restriction will result in a positively skewed distribution for the waist sizes of the selected individuals. Note that this type of truncation might occur without our knowing of its occurrence. In that case, it is called a hidden truncation. Of course, selection mechanisms are not restricted to lower truncations, but can also involve upper truncations or various more complex selection conditions. For instance, airplane crews cannot be too tall because of the restrictive dimensions of the cabin. So the height requirements for airplane crews effectively involve both lower and upper truncations.

Astronomy is another setting where hidden truncations occur. The impact of observational selection effects, known generically in the literature as Malmquist bias (Malmquist 1920), has been at the centre of much controversy in the last decades. This effect occurs when an astronomer observes objects at a distance. Indeed, instruments set a limit on the faintest objects (such as stars for example) that we can see. In other words, in any observation there is a minimum flux density (measure of how bright objects seem to us from Earth) below which we will not detect an object. This flux density is proportional to the luminosity (measure of how bright objects are intrinsically) divided by the square of the distance, so it is possible to see more luminous objects out to larger distances than intrinsically faint objects. This means that the relative numbers of intrinsically bright and faint objects that we see may be nothing like the relative numbers per unit volume of space. Instead, bright objects are over-represented and the average luminosity of the objects we see inevitably increases with distance. There is not really anything that can be done about the Malmquist bias at this point in our stage of technology and observation, except merely to correct for it in the statistical analysis. Neglecting this selection bias will result in skewed conclusions.

### *1.3. Objectives and summary.*

The main objective of this article is to develop a unified framework for distributions arising from selections and to establish links with existing definitions of multivariate skewed distributions. Because many definitions have emerged in recent years, we believe that it is necessary to provide a crisp overview of this topic from a unified perspective. We hope that this account will serve as a resource of information and future research topics to experts in the field, new researchers, and practitioners.

Statistical inference for those families of multivariate skewed distributions is still mostly unresolved and problems have been reported already in the simplest case of the skew-normal distribution. It is currently the subject of vigorous research efforts and recent advances on this topic include work by Azzalini (2005), Azzalini & Capitanio (1999, 2003), DiCiccio & Monti (2004), Ma, Genton & Tsiatis (2005), Ma & Hart (2006), Pewsey (2000, 2006), and Sartori (2006). It is hence beyond the scope of this article to review inferential aspects. Future research on inferential issues for skewed distributions arising from selections, both from a frequentist and a Bayesian perspective, will however lead to the ultimate goal of facilitating the development of new applications.

The structure of the present article is the following. In Section 2, a definition of a selection

distribution that unifies many existing definitions of multivariate skewed distribution available in the literature is proposed. The main properties of selection distributions are also derived and therefore unify the properties of these existing definitions. In Section 3, we describe in detail the link between selection distributions and multivariate skewed distributions developed in the literature, such as skew-symmetric and skew-elliptical distributions, among many others. In Section 4, several methods for constructing selection distributions based on linear and nonlinear selection mechanisms are introduced. A discussion is provided in Section 5.

## 2. SELECTION DISTRIBUTIONS

We will begin by proposing a definition of a multivariate selection distribution and discuss some of its main properties.

**DEFINITION 1.** Let  $\mathbf{U} \in \mathbb{R}^q$  and  $\mathbf{V} \in \mathbb{R}^p$  be two random vectors, and denote by  $C$  a measurable subset of  $\mathbb{R}^q$ . We define a selection distribution as the conditional distribution of  $\mathbf{V}$  given  $\mathbf{U} \in C$ , that is, as the distribution of  $(\mathbf{V} | \mathbf{U} \in C)$ . We say that a random vector  $\mathbf{X} \in \mathbb{R}^p$  has a selection distribution if  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C)$ . We use the notation  $\mathbf{X} \sim SLCT_{p,q}$  with parameters depending on the characteristics of  $\mathbf{U}$ ,  $\mathbf{V}$ , and  $C$ .

This definition is simply a conditional distribution, but it serves to unify existing definitions of skewed distribution arising from selection mechanisms, as well as to provide new classes of multivariate symmetric and skewed distributions. Of course, if  $C = \mathbb{R}^q$ , then there is no selection and so we assume that  $0 < P(\mathbf{U} \in C) < 1$  in the sequel. If  $\mathbf{U} = \mathbf{V}$ , then the selection distribution is simply a truncated distribution described by  $\mathbf{V}$  given  $\mathbf{V} \in C$ . Moreover, if  $\mathbf{V} = (\mathbf{U}, \mathbf{T})$ , then a selection distribution corresponds to the joint distribution of  $\mathbf{U}$  and  $\mathbf{T}$  (marginally) truncated on  $\mathbf{U} \in C$ , from which we can obtain by marginalization both the truncated distribution of  $(\mathbf{U} | \mathbf{U} \in C)$  and the selection distribution of  $(\mathbf{T} | \mathbf{U} \in C)$ .

### 2.1. Main properties.

The main properties of a selection random vector  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C)$  can be studied easily by examining the properties of the random vectors  $\mathbf{V}$  and  $\mathbf{U}$ . Indeed, if  $\mathbf{V}$  in Definition 1 has a probability density function (pdf)  $f_{\mathbf{V}}$  say, then  $\mathbf{X}$  has a pdf  $f_{\mathbf{X}}$  given by:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x}) \frac{P(\mathbf{U} \in C | \mathbf{V} = \mathbf{x})}{P(\mathbf{U} \in C)}. \tag{2}$$

See, for example, Arellano-Valle and del Pino (2004) for this, and see Arellano-Valle, del Pino & San Martín (2002) for the particular choice of  $C = \{\mathbf{u} \in \mathbb{R}^q | \mathbf{u} > \mathbf{0}\}$ , where the inequality between vectors is meant componentwise. As motivated by the selection mechanisms in Section 1.2, our interest lies in situations where the pdf  $f_{\mathbf{V}}$  is symmetric. In that case, the resulting pdf (2) is generally skewed, unless the conditional probabilities satisfy  $P(\mathbf{U} \in C | \mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C | \mathbf{V} = -\mathbf{x})$  for all  $\mathbf{x} \in \mathbb{R}^p$ . For instance, this is the case if  $\mathbf{U}$  and  $\mathbf{V}$  are independent, but this is not a necessary condition.

Further properties of the selection random vector  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C)$  can be studied directly from its definition. For instance, we have:

(P1) When  $(\mathbf{U}, \mathbf{V})$  has a joint density,  $f_{\mathbf{U},\mathbf{V}}$  say, an alternative expression for the pdf of  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C)$  is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\int_C f_{\mathbf{U},\mathbf{V}}(\mathbf{u}, \mathbf{x}) d\mathbf{u}}{\int_C f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}}, \tag{3}$$

where  $f_{\mathbf{U}}$  denotes the marginal pdf of  $\mathbf{U}$ . Expression (3) is useful to compute the cumulative distribution function (cdf) as well as the moment generating function (mgf) of  $\mathbf{X}$ .

(P2) For any Borel function  $g$ , we have  $g(\mathbf{X}) \stackrel{d}{=} (g(\mathbf{V}) | \mathbf{U} \in C)$ . Moreover, since  $(g(\mathbf{V}) | \mathbf{U} \in C)$  is determined by the transformation  $(\mathbf{U}, \mathbf{V}) \rightarrow (\mathbf{U}, g(\mathbf{V}))$ , then the pdf of  $g(\mathbf{X})$  can be computed by replacing  $\mathbf{X}$  with  $g(\mathbf{X})$  and  $\mathbf{V}$  with  $g(\mathbf{V})$  in (2) (or (3)), respectively. If a family of distributions of  $\mathbf{V}$  is closed under a set of transformations  $g$  (for instance, all linear transformations) then the corresponding family of selection distributions is also closed under this set of transformations. If a family of distributions of  $\mathbf{V}$  is closed under marginalization (for instance, if  $\mathbf{V}$  is multivariate normal, then subvectors of  $\mathbf{V}$  are multivariate normal), then the corresponding family of selection distributions is also closed under marginalization.

(P3) The conditional selection distribution of  $(\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1)$ , where the joint distribution of  $(\mathbf{X}_1, \mathbf{X}_2) \stackrel{d}{=} [(\mathbf{V}_1, \mathbf{V}_2) | \mathbf{U} \in C]$ , can be obtained by noting that it is equivalent to consider the conditional selection distribution of  $(\mathbf{V}_2 | \mathbf{V}_1 = \mathbf{x}_1, \mathbf{U} \in C)$ . By (2), it has a pdf given by

$$f_{\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1}(\mathbf{x}_2) = f_{\mathbf{V}_2 | \mathbf{V}_1 = \mathbf{x}_1}(\mathbf{x}_2) \frac{P(\mathbf{U} \in C | \mathbf{V}_1 = \mathbf{x}_1, \mathbf{V}_2 = \mathbf{x}_2)}{P(\mathbf{U} \in C | \mathbf{V}_1 = \mathbf{x}_1)}. \tag{4}$$

Consider also the partition  $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$  such that  $(\mathbf{U}_1, \mathbf{V}_1)$  and  $(\mathbf{U}_2, \mathbf{V}_2)$  are independent, and assume that  $C = C_1 \times C_2$  is such that  $\mathbf{U} \in C$  is equivalent to  $\mathbf{U}_1 \in C_1$  and  $\mathbf{U}_2 \in C_2$ . Then under these conditions it follows directly from (4) that  $\mathbf{X}_1 \stackrel{d}{=} (\mathbf{V}_1 | \mathbf{U} \in C) \stackrel{d}{=} (\mathbf{V}_1 | \mathbf{U}_1 \in C_1)$  and  $\mathbf{X}_2 \stackrel{d}{=} (\mathbf{V}_2 | \mathbf{U} \in C) \stackrel{d}{=} (\mathbf{V}_2 | \mathbf{U}_2 \in C_2)$ , and that they are independent. For more details on these properties but considering the particular subset  $C = \{\mathbf{u} \in \mathbb{R}^q | \mathbf{u} > \mathbf{0}\}$ , see Arellano-Valle & Genton (2005) and Arellano-Valle & Azzalini (2006).

Selection distributions depend also on the subset  $C$  of  $\mathbb{R}^q$ . One of the most important selection subset is defined by

$$C(\boldsymbol{\beta}) = \{\mathbf{u} \in \mathbb{R}^q | \mathbf{u} > \boldsymbol{\beta}\}, \tag{5}$$

where  $\boldsymbol{\beta}$  is a vector of truncation levels. In particular, the subset  $C(\mathbf{0})$  leads to simple selection distributions. Note however that the difference between  $C(\boldsymbol{\beta})$  and  $C(\mathbf{0})$  is essentially a change of location. Another interesting subset consists of upper and lower truncations and is defined by

$$C(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \{\mathbf{u} \in \mathbb{R}^q | \boldsymbol{\alpha} > \mathbf{u} > \boldsymbol{\beta}\}; \tag{6}$$

see the example in Section 1.2 about height requirements for airplane crews with both lower and upper truncations. Arnold, Beaver, Groeneveld & Meeker (1993) considered the case of  $C(\boldsymbol{\alpha}, \boldsymbol{\beta})$  with  $p = q = 1$ . Many other subsets can be constructed, for instance quadratic-based subsets of the form  $\{\mathbf{u} \in \mathbb{R}^q | \mathbf{u}^\top Q \mathbf{u} > \beta\}$ , where  $Q$  is a matrix and  $\beta \in \mathbb{R}$  a truncation level; see Genton (2005) for more examples.

### 2.2. Selection elliptical distributions.

Further specific properties of selection distributions can be examined when we know the joint distribution of  $\mathbf{U}$  and  $\mathbf{V}$ . A quite popular family of selection distributions arises when  $\mathbf{U}$  and  $\mathbf{V}$  have a joint multivariate elliptically contoured (*EC*) distribution, that is

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim EC_{q+p} \left( \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_U \\ \boldsymbol{\xi}_V \end{pmatrix}, \Omega = \begin{pmatrix} \Omega_U & \Delta^\top \\ \Delta & \Omega_V \end{pmatrix}, h^{(q+p)} \right), \tag{7}$$

where  $\boldsymbol{\xi}_U \in \mathbb{R}^q$  and  $\boldsymbol{\xi}_V \in \mathbb{R}^p$  are location vectors,  $\Omega_U \in \mathbb{R}^{q \times q}$ ,  $\Omega_V \in \mathbb{R}^{p \times p}$ , and  $\Delta \in \mathbb{R}^{p \times q}$  are dispersion matrices, and, in addition to these parameters,  $h^{(q+p)}$  is a density generator function. We denote the selection distribution resulting from (7) by  $SLCT-EC'_{p,q}(\boldsymbol{\xi}, \Omega, h^{(q+p)}, C)$ . They typically result in skew-elliptical distributions, except for the cases: (a)  $\Delta = O$ , a matrix

of zeros, and  $h^{(q+p)}$  is the normal density generator function; or (b)  $\Delta = O$  and  $C = C(\xi_U)$  defined according to (5), since in both cases we have in (2) that  $P(U \in C | V = x) = P(U \in C)$  for all  $x$ . In fact, conditions in (a) imply that  $U$  and  $V$  are independent, while conditions in (b) imply that  $P(U \in C | V = x) = P(U \in C) = \Phi_q(0; 0, \Omega_U)$  for all  $x$  and any density generator function  $h^{(q+p)}$ , where  $\Phi_q(\cdot; 0, \Omega_U)$  denotes the cdf of the  $N_q(0, \Omega_U)$  distribution. In other words, the condition that  $U$  and  $V$  be correlated random vectors, i.e.,  $\Delta \neq O$ , is important in order to avoid in (2) that the skewing factor  $P(U \in C | V = x)/P(U \in C)$  be equal to one for all  $x$ : see also Section 4.

In general, any selection distribution can be constructed hierarchically by specifying a marginal distribution for  $V$  and then a conditional distribution for  $(U | V = x)$ . For example, the distribution  $SLCT-EC_{q,p}(\xi, \Omega, h^{(q+p)}, C)$  that follows from (7) according to Definition 1 can be obtained by specifying

$$\begin{aligned} V &\sim EC_p(\xi_V, \Omega_V, h^{(p)}) \quad \text{and} \\ (U | V = x) &\sim EC_q(\xi_U + \Delta^\top \Omega_V^{-1}(x - \xi_V), \Omega_U - \Delta^\top \Omega_V^{-1} \Delta, h_{v(x)}^{(q)}), \end{aligned} \tag{8}$$

where  $h_{v(x)}^{(q)}(u) = h^{(q+p)}(u + v(x))/h^{(p)}(v(x))$  is the induced conditional generator, and  $v(x) = (x - \xi_V)^\top \Omega_V^{-1}(x - \xi_V)$ . It is then straightforward to obtain the pdf of  $X \stackrel{d}{=} (V | U \in C)$  by means of (2). In fact, let

$$f_r(y; \xi_Y, \Omega_Y, h^{(r)}) = |\Omega_Y|^{-1/2} h^{(r)}((y - \xi_Y)^\top \Omega_Y^{-1}(y - \xi_Y))$$

be the pdf of a  $r$ -dimensional elliptically contoured random vector  $Y \sim EC_r(\xi_Y, \Omega_Y, h^{(r)})$ , and denote  $P(Y \in C)$  by  $\bar{F}_r(C; \xi_Y, \Omega_Y, h^{(r)})$ , i.e.,

$$\bar{F}_r(C; \xi_Y, \Omega_Y, h^{(r)}) = \int_C f_r(y; \xi_Y, \Omega_Y, h^{(r)}) dy.$$

Then, applying (8) in (2) we have the following generic expression for the pdf of  $X \stackrel{d}{=} (V | U \in C)$ :

$$f_X(x) = f_p(x; \xi_V, \Omega_V, h^{(p)}) \frac{\bar{F}_q(C; \Delta^\top \Omega_V^{-1}(x - \xi_V) + \xi_U, \Omega_U - \Delta^\top \Omega_V^{-1} \Delta, h_{v(x)}^{(q)})}{\bar{F}_q(C; \xi_U, \Omega_U, h^{(q)})}, \tag{9}$$

which can also be easily rewritten according to (3) as

$$f_X(x) = \frac{\int_C f_{q+p}(u, x; \xi, \Omega, h^{(q+p)}) du}{\int_C f_q(u; \xi_U, \Omega_U, h^{(q)}) du}. \tag{10}$$

Expression (10) can be helpful to compute the cdf and also the mgf of any selection elliptical random vector  $X \sim SLCT-EC_{q,p}(\xi, \Omega, h^{(q+p)}, C)$ .

Note finally that depending on the choice of  $h^{(q+p)}$ , (9) (or (10)) becomes the pdf of different selection elliptical distributions. For example, when  $h^{(q+p)}(u) = c(q+p, \nu)\{1+u\}^{-(q+p+\nu)/2}$ , where  $c(r, \nu) = \Gamma[(r+\nu)/2]\nu^{\nu/2}/(\Gamma[\nu/2]\pi^{r/2})$ , (9) becomes the selection- $t$  pdf with  $\nu$  degrees of freedom given by

$$f_X(x) = t_p(x; \xi_V, \Omega_V, \nu) \frac{\bar{T}_q(C; \Delta^\top \Omega_V^{-1}(x - \xi_V) + \xi_U, \frac{\nu + v(x)}{\nu + p}[\Omega_U - \Delta^\top \Omega_V^{-1} \Delta], \nu + p)}{\bar{T}_q(C; \xi_U, \Omega_U, \nu)}, \tag{11}$$

where  $t_r(\cdot; \xi, \Omega, \nu)$  denotes the pdf of a Student- $t$  random vector  $Y \sim \text{Student}_r(\xi, \Omega, \nu)$  of dimension  $r$ , and  $\bar{T}_r(C; \xi, \Omega, \nu) = P(Y \in C)$ . This produces a heavy tailed selection distribution that is useful for robustness purposes. Due to its importance, the normal case for which  $h^{(r)}(u) = \phi_r(\sqrt{u}) = (2\pi)^{-r/2} \exp\{-u/2\}$ ,  $u \geq 0$ , is considered next in a separate section.

### 2.3. The selection normal distribution.

Within the elliptically contoured class (7), one of the most important cases is undoubtedly when  $\mathbf{U}$  and  $\mathbf{V}$  are jointly multivariate normal, so that instead of (7), we have

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim N_{q+p} \left( \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{\mathbf{U}} \\ \boldsymbol{\xi}_{\mathbf{V}} \end{pmatrix}, \Omega = \begin{pmatrix} \Omega_{\mathbf{U}} & \Delta^{\top} \\ \Delta & \Omega_{\mathbf{V}} \end{pmatrix} \right). \quad (12)$$

We denote the selection distribution resulting from (12) by  $SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \Omega, C)$ . We establish two important theoretical characteristics of the  $SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \Omega, C)$  distributions, namely their pdf and their mgf, for any selection subset  $C \in \mathbb{R}^q$ . This result unifies the moment properties of existing definitions of skewed distributions based on the multivariate normal distribution. Specifically, let  $\mathbf{X}$  have a multivariate selection distribution according to Definition 1 with  $\mathbf{U}$  and  $\mathbf{V}$  jointly normal as in (12), that is,  $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \Omega, C)$ . Let  $\phi_r(\cdot; \boldsymbol{\xi}_{\mathbf{Y}}, \Omega_{\mathbf{Y}})$  be the pdf of an  $r$ -dimensional normal random vector  $\mathbf{Y} \sim N_r(\boldsymbol{\xi}_{\mathbf{Y}}, \Omega_{\mathbf{Y}})$ , and denote by  $\bar{\Phi}_r(C; \boldsymbol{\xi}_{\mathbf{Y}}, \Omega_{\mathbf{Y}}) = P(\mathbf{Y} \in C)$ . It is then straightforward to derive the conditional distribution in Definition 1, yielding the following pdf based on (2):

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \boldsymbol{\xi}_{\mathbf{V}}, \Omega_{\mathbf{V}}) \frac{\bar{\Phi}_q(C; \Delta^{\top} \Omega_{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}}) + \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}} - \Delta^{\top} \Omega_{\mathbf{V}}^{-1} \Delta)}{\bar{\Phi}_q(C; \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}})}, \quad (13)$$

which obviously follows also from (9) when  $h^{(q+p)}$  is chosen as the normal density generator. In addition, the moment generating function of  $\mathbf{X}$  is

$$M_{\mathbf{X}}(\mathbf{s}) = \exp \left\{ \mathbf{s}^{\top} \boldsymbol{\xi}_{\mathbf{V}} + \frac{1}{2} \mathbf{s}^{\top} \Omega_{\mathbf{V}} \mathbf{s} \right\} \frac{\bar{\Phi}_q(C; \Delta^{\top} \mathbf{s} + \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}})}{\bar{\Phi}_q(C; \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}})}. \quad (14)$$

The proof of this result is provided in the Appendix. From (13) and (14), we can also derive the following stochastic representation (or transformation) for the normal selection random vector  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C)$ , i.e., with  $\mathbf{U}$  and  $\mathbf{V}$  jointly normal as in (12). Let  $\mathbf{X}_0 \sim N_q(\mathbf{0}, \Omega_{\mathbf{U}})$  and  $\mathbf{X}_1 \sim N_p(\mathbf{0}, \Omega_{\mathbf{V}} - \Delta \Omega_{\mathbf{U}}^{-1} \Delta^{\top})$  be independent normal random vectors. Note by (12) that  $(\mathbf{U}, \mathbf{V}) \stackrel{d}{=} (\boldsymbol{\xi}_{\mathbf{U}} + \mathbf{X}_0, \boldsymbol{\xi}_{\mathbf{V}} + \Delta \Omega_{\mathbf{U}}^{-1} \mathbf{X}_0 + \mathbf{X}_1)$ . Thus, considering also a random vector  $\mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}] \stackrel{d}{=} (\mathbf{X}_0 | \mathbf{X}_0 \in C - \boldsymbol{\xi}_{\mathbf{U}})$ , which is independent of  $\mathbf{X}_1$ , we have

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\xi}_{\mathbf{V}} + \Delta \Omega_{\mathbf{U}}^{-1} \mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}] + \mathbf{X}_1. \quad (15)$$

All these results reduce to similar ones obtained in Arellano-Valle & Azzalini (2006) when  $C = C(\mathbf{0})$ . The stochastic representation (15) is useful for simulations from selection distributions and for computations of their moments. For instance, the mean vector and covariance matrix of  $\mathbf{X}$  are given by

$$\begin{aligned} E(\mathbf{X}) &= \boldsymbol{\xi}_{\mathbf{V}} + \Delta \Omega_{\mathbf{U}}^{-1} E(\mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}]), \\ \text{var}(\mathbf{X}) &= \Omega_{\mathbf{V}} + \Delta \Omega_{\mathbf{U}}^{-1} \{ \Omega_{\mathbf{U}} - \text{var}(\mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}]) \} \Omega_{\mathbf{U}}^{-1} \Delta^{\top}, \end{aligned}$$

respectively. Note that, in general, it is not easy to obtain closed-form expressions for the ‘‘truncated’’ mean vector  $E(\mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}])$  and covariance matrix  $\text{var}(\mathbf{X}_0[C - \boldsymbol{\xi}_{\mathbf{U}}])$  for any selection subset  $C$ . However, for the special case where  $\Omega_{\mathbf{U}}$  is diagonal and  $C = C(\boldsymbol{\alpha}, \boldsymbol{\beta})$ , the components of  $(\mathbf{X}_0 | \mathbf{X}_0 \in C)$  are independent. Therefore the computation of these moments is equivalent to computing marginal truncated means and variances of the form  $E(X | a < X < b)$  and  $\text{var}(X | a < X < b)$ , where  $X \sim N(\mu, \sigma^2)$ , for which the results of Johnson, Kotz & Balakrishnan (1994, § 10.1) can be used; see also Arellano-Valle & Genton (2005) and Arellano-Valle & Azzalini (2006) for more details on similar results.

2.4. Scale mixture of the selection normal distribution.

Another important particular case that follows from the elliptically contoured class (7) is obtained when the joint distribution of  $\mathbf{U}$  and  $\mathbf{V}$  is such that

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \mid W = w \sim N_{q+p} \left( \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{\mathbf{U}} \\ \boldsymbol{\xi}_{\mathbf{V}} \end{pmatrix}, w\Omega = \begin{pmatrix} w\Omega_{\mathbf{U}} & w\Delta^{\top} \\ w\Delta & w\Omega_{\mathbf{V}} \end{pmatrix} \right), \tag{16}$$

for some nonnegative random variable  $W$  with cdf  $G$ . This is equivalent to considering in (7) a density generator function such that  $h^{(q+p)}(u) = \int_0^\infty (2\pi w)^{-(q+p)/2} \exp\{-u/2w\} dG(w)$ . For example, if  $G$  is the cdf of the inverted gamma distribution with parameters  $\nu/2$  and  $\nu/2$ ,  $IG(\nu/2, \nu/2)$  say, then  $h^{(q+p)}(u) = c(q+p, \nu)\{1+u\}^{-(q+p+\nu)/2}$ , where  $c(r, \nu)$  is defined in Section 2.2, yielding the selection- $t$  distribution with  $\nu$  degrees of freedom and pdf (11).

The properties of any selection random vector  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} \mid \mathbf{U} \in C)$  that follows from (16) can be studied by noting that  $(\mathbf{X} \mid W = w) \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, w\Omega, C)$ , where  $W \sim G$ . In particular, for the pdf of  $\mathbf{X}$ , we have  $f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty f_{\mathbf{X} \mid W=w}(\mathbf{x}) dG(w)$ , where  $f_{\mathbf{X} \mid W=w}(\mathbf{x})$  is the selection normal pdf (13) with  $\Omega$  replaced by  $w\Omega$ , i.e.

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \phi_p(\mathbf{x}; \boldsymbol{\xi}_{\mathbf{V}}, w\Omega_{\mathbf{V}}) \frac{\bar{\Phi}_q(C; \Delta^{\top} \Omega_{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}}) + \boldsymbol{\xi}_{\mathbf{U}}, w\{\Omega_{\mathbf{U}} - \Delta^{\top} \Omega_{\mathbf{V}}^{-1} \Delta\})}{\bar{\Phi}_q(C; \boldsymbol{\xi}_{\mathbf{U}}, w\Omega_{\mathbf{U}})} dG(w).$$

Alternatively, from (3) or (10) it follows that the pdf of  $\mathbf{X}$  can also be computed as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\int_C \int_0^\infty \phi_{q+p}(\mathbf{u}, \mathbf{x}; \boldsymbol{\xi}, w\Omega) dG(w) d\mathbf{u}}{\int_C \int_0^\infty \phi_q(\mathbf{u}; \boldsymbol{\xi}_{\mathbf{U}}, w\Omega_{\mathbf{U}}) dG(w) d\mathbf{u}}.$$

Similar results for the mgf can be obtained by considering (14). Note, however, that the mean and covariance matrix of  $\mathbf{X}$  can be computed (provided they exist) by using the well-known facts that  $E(\mathbf{X}) = E[E(\mathbf{X} \mid W)]$  and  $\text{var}(\mathbf{X}) = \text{var}[E(\mathbf{X} \mid W)] + E[\text{var}(\mathbf{X} \mid W)]$ , where  $E(\mathbf{X} \mid W = w)$  and  $\text{var}(\mathbf{X} \mid W = w)$  are the mean and covariance matrix corresponding to the (conditional) selection normal random vector  $(\mathbf{X} \mid W = w) \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, w\Omega, C)$ ; see Section 2.3. Additional properties of this subclass of selection elliptical models can be explored for particular choices of  $C$ . For instance, when  $C = C(\boldsymbol{\xi}_{\mathbf{U}})$ , it is easy to show from (15) that

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\xi}_{\mathbf{V}} + \sqrt{W} \mathbf{X}_N, \tag{17}$$

where  $W \sim G$  and  $\mathbf{X}_N \sim SLCT\text{-}N_{p,q}(\mathbf{0}, \Omega, C(\boldsymbol{\xi}_{\mathbf{U}}))$ , and they are independent. In particular, if  $W$  has finite first moment, then (17) yields  $E(\mathbf{X}) = \boldsymbol{\xi}_{\mathbf{V}} + E(\sqrt{W})E(\mathbf{X}_N)$  and  $\text{var}(\mathbf{X}) = E(W) \text{var}(\mathbf{X}_N)$ . More generally, since  $\bar{\Phi}_q(C(\boldsymbol{\beta}); \boldsymbol{\xi}_{\mathbf{Y}}, \Omega_{\mathbf{Y}}) = \bar{\Phi}_q(\boldsymbol{\xi}_{\mathbf{Y}} - \boldsymbol{\beta}; \mathbf{0}, \Omega_{\mathbf{Y}})$ , the cdf of a  $N_q(\mathbf{0}, \Omega_{\mathbf{Y}})$  distribution evaluated at  $\boldsymbol{\xi}_{\mathbf{Y}} - \boldsymbol{\beta}$ , then from (17) and (14), we have for the mgf of  $\mathbf{X}$  that

$$M_{\mathbf{X}}(\mathbf{s}) = E[M_{\mathbf{V} \mid W}(\mathbf{s})] \frac{\bar{\Phi}_q(\Delta^{\top} \mathbf{s}; \mathbf{0}, \Omega_{\mathbf{U}})}{\bar{\Phi}_q(\mathbf{0}; \mathbf{0}, \Omega_{\mathbf{U}})}, \tag{18}$$

where  $M_{\mathbf{V} \mid W}(\mathbf{s}) = \exp\{\mathbf{s}^{\top} \boldsymbol{\xi}_{\mathbf{V}} + \frac{1}{2} \mathbf{s}^{\top} \Omega_{\mathbf{V}} \mathbf{s}\}$ , since  $(\mathbf{V} \mid W = w) \sim N_p(\boldsymbol{\xi}_{\mathbf{V}}, w\Omega_{\mathbf{V}})$  and  $W \sim G$ . The proof of this result is provided in the Appendix. From (18) it is clear that the moment computations reduce substantially when the matrix  $\Omega_{\mathbf{U}}$  is diagonal. The special case of the multivariate skew- $t$  distribution that follows when  $G$  is the  $IG(\nu/2, \nu/2)$  distribution is studied in Arellano-Valle & Azzalini (2006).

3. LINK WITH SKEWED MULTIVARIATE DISTRIBUTIONS

We show how many existing definitions of skew-normal, skew-elliptical, and other related distributions from the literature can be viewed as selection distributions in terms of assumptions



on the joint distribution of  $(\mathbf{U}, \mathbf{V})$  and on  $C$ . We start from the most general class of fundamental skewed distributions (Arellano-Valle & Genton 2005) and by successive specialization steps work down to the very specific skew-normal distribution of Azzalini (1985). We describe the links between those various distributions and provide a schematic illustrative summary of selection distributions and other skewed distributions.

3.1. *Fundamental skewed distributions.*

Fundamental skewed distributions have been introduced by Arellano-Valle & Genton (2005). They are selection distributions with the particular selection subset  $C(\mathbf{0})$  defined in (5), and therefore the pdf (2) becomes

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x}) \frac{P(\mathbf{U} > \mathbf{0} \mid \mathbf{V} = \mathbf{x})}{P(\mathbf{U} > \mathbf{0})}. \tag{19}$$

When the pdf  $f_{\mathbf{V}}$  is symmetric or elliptically contoured, the distributions resulting from (19) are called *fundamental skew-symmetric (FUSS)* and *fundamental skew-elliptical (FUSE)*, respectively. If  $f_{\mathbf{V}}$  is the multivariate normal pdf, then the distribution resulting from (19) is called *fundamental skew-normal (FUSN)*. In this case, note that the joint distribution of  $\mathbf{U}$  and  $\mathbf{V}$  is not directly specified, and in particular does not need to be normal. The only requirement is to have a model for  $P(\mathbf{U} > \mathbf{0} \mid \mathbf{V} = \mathbf{x})$ . Such an example is provided in Section 3.5. Finally, if  $\mathbf{U}$  and  $\mathbf{V}$  are jointly multivariate normal as in (12), then (19) takes the form of (13) with  $C = C(\mathbf{0})$  and  $\bar{\Phi}_r(\cdot; \cdot, \cdot) = \Phi_r(\cdot; \boldsymbol{\xi}, \Omega)$ , the cdf of a  $N_r(\boldsymbol{\xi}, \Omega)$  distribution. In particular, if  $\boldsymbol{\xi}_{\mathbf{U}} = \mathbf{0}$ ,  $\boldsymbol{\xi}_{\mathbf{V}} = \mathbf{0}$ ,  $\Omega_{\mathbf{U}} = I_q$ , and  $\Omega_{\mathbf{V}} = I_p$ , then the resulting distribution is called *canonical fundamental skew-normal (CFUSN)* with the pdf reducing to the form:

$$f_{\mathbf{X}}(\mathbf{x}) = 2^q \phi_p(\mathbf{x}) \Phi_q(\Delta^\top \mathbf{x}; \mathbf{0}, I_q - \Delta^\top \Delta). \tag{20}$$

Location and scale can be further reincorporated in (20) by considering the pdf of  $\boldsymbol{\mu} + \Sigma^{1/2} \mathbf{X}$ , where  $\boldsymbol{\mu} \in \mathbb{R}^p$  is a location vector,  $\Sigma \in \mathbb{R}^{p \times p}$  is a positive definite scale matrix. This family of distributions has many interesting properties such as a stochastic representation, a simple formula for its cdf, characteristics of linear transformations and quadratic forms, among many others, see Arellano-Valle & Genton (2005) for a detailed account. Fundamental skewed distributions appear also naturally in the context of linear combinations of order statistics; see Crocetta & Loperfido (2005) and Arellano-Valle & Genton (2006a,b).

3.2. *Unified skewed distributions.*

Unified skewed distributions have been introduced by Arellano-Valle & Azzalini (2006). They are *FUSN* or *FUSE* distributions with the additional assumption that the joint distribution of  $\mathbf{U}$  and  $\mathbf{V}$  is multivariate normal or elliptically contoured as in (12) or (7), respectively. These unified skewed distribution are called *unified skew-normal (SUN)* and *unified skew-elliptical (SUE)*, respectively, and they unify (contain or are equivalent under suitable reparameterizations to) all the skew-normal or skew-elliptical distributions derived by assuming (12) or (7), respectively. In fact, with some modification of the parameterization used by Arellano-Valle & Azzalini (2006), the *SUE* pdf is just the selection pdf that follows from (7) with  $C = C(\mathbf{0})$  and has the form

$$f_{\mathbf{X}}(\mathbf{x}) = f_p(\mathbf{x}; \boldsymbol{\xi}_{\mathbf{V}}, \Omega_{\mathbf{V}}, h^{(p)}) \frac{F_q(\boldsymbol{\xi}_{\mathbf{U}} + \Delta^\top \Omega_{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}}); \mathbf{0}, \Omega_{\mathbf{U}} - \Delta^\top \Omega_{\mathbf{V}}^{-1} \Delta, h_{v(\mathbf{x})}^{(q)})}{F_q(\boldsymbol{\xi}_{\mathbf{U}}; \mathbf{0}, \Omega_{\mathbf{U}}, h^{(q)})}, \tag{21}$$

where (as before)  $f_p(\mathbf{x}; \boldsymbol{\xi}_{\mathbf{V}}, \Omega_{\mathbf{V}}, h^{(p)}) = |\Omega_{\mathbf{V}}|^{-1/2} h^{(p)}(v(\mathbf{x}))$ ,  $v(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}})^\top \Omega_{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}})$  and  $h_{v(\mathbf{x})}^{(q)}(u) = h^{(p+q)}(u + v(\mathbf{x}))/h^{(p)}(v(\mathbf{x}))$ , and  $F_r(\mathbf{z}; \mathbf{0}, \Theta, g^{(r)})$  denote the cdf of the  $EC_r(\mathbf{0}, \Theta, g^{(r)})$  distribution, namely  $F_r(\mathbf{z}; \mathbf{0}, \Theta, g^{(r)}) = \int_{\mathbf{v} \leq \mathbf{z}} |\Theta|^{-1/2} g^{(r)}(\mathbf{v}^\top \Theta^{-1} \mathbf{v}) d\mathbf{v}$ . The *SUN* pdf follows by taking the normal generator  $h^{(p+q)}(u) = (2\pi)^{-(p+q)/2} e^{-u/2}$ ,  $u \geq 0$ ,

for which  $h_v^{(r)}(u) = h^{(r)}(u) = (2\pi)^{-r/2}e^{-u/2}$ ,  $u, v \geq 0$ , and so (21) takes the form of (13) with  $C = C(\mathbf{0})$  and  $\bar{\Phi}_r(\cdot; \cdot, \cdot) = \Phi_r(\cdot; \boldsymbol{\xi}, \Omega)$ . The multivariate skew- $t$  ( $ST$ ) distribution considered in Branco & Dey (2001) and Azzalini & Capitanio (2003) is extended in Arellano-Valle & Azzalini (2006) by considering in (21) the  $t$ -generator with  $\nu$  degrees of freedom defined by  $h^{(p+q)}(u) = c(p+q, \nu)\{\nu+u\}^{-(p+q+\nu)/2}$ ,  $u \geq 0$ , for which  $h^{(p)} = c(p, \nu)\{\nu+u\}^{-(p+\nu)/2}$  and  $h_v^{(q)}(u) = c(q, \nu+p)\{\nu+v+u\}^{-(p+q+\nu)/2}$ ,  $u, v \geq 0$ , where  $c(r, \omega) = \Gamma[(r+\omega)/2]\omega^{\omega/2}/\Gamma[\omega/2]\pi^{r/2}$ ; see also Section 2.4.

3.3. Closed skewed distributions.

The closed skew-normal ( $CSN$ ) distributions introduced by González-Farías, Domínguez-Molina & Gupta (2004a) is also  $FUSN$  by reparameterization of the covariance matrix of the normal joint distribution in (12) as  $\Delta = \Omega_V D^\top$  and  $\Omega_U = \Theta + D\Omega_V D^\top$ . Another closed skew-normal distribution, which was introduced in Arellano-Valle & Azzalini (2006), follows also as a  $FUSN$  when  $\Delta = A\Omega_U$  and  $\Omega_V = \Psi + A\Omega_U A^\top$  in (12). An important property of this last parameterization is that it simplifies the stochastic representation in (15) to a simpler structure of the form  $\mathbf{X} \stackrel{d}{=} \boldsymbol{\xi}_V + A\mathbf{X}_0[-\boldsymbol{\xi}_U] + \mathbf{X}_1$ , where  $\mathbf{X}_0[-\boldsymbol{\xi}_U] \stackrel{d}{=} (\mathbf{X}_0 | \mathbf{X}_0 > -\boldsymbol{\xi}_U)$ , with  $\mathbf{X}_0 \sim N_q(\mathbf{0}, \Omega_U)$ , and is independent of  $\mathbf{X}_1 \sim N_p(\mathbf{0}, \Psi)$ , thus generalizing the skew-normal distribution considered by Sahu, Dey & Branco (2003). As was shown in Arellano-Valle & Azzalini (2006), all these distributions are equivalent to the  $SUN$  distribution under some suitable reparameterizations. The term *closed* was motivated by the fact that  $CSN$  distributions are closed under marginalizations and also under conditioning when  $P(\mathbf{U} \in C) \neq 2^{-q}$ . Moreover, the  $CSN$  distributions have some particular additive properties, see González-Farías, Domínguez-Molina & Gupta (2004b).

3.4. Hierarchical skewed distributions.

Liseo & Loperfido (2006) generalized a pioneer result obtained by O’Hagan & Leonard (1976) from a Bayesian point of view, who used the selection idea to incorporate constraints on parameters in the prior specifications. This is just an interesting example of how Bayesian modelling can be viewed through selection distributions. In fact, suppose that  $\mathbf{V} | \boldsymbol{\mu} \sim N_p(\mathbf{a} + A\boldsymbol{\mu}, \Theta)$ ,  $\boldsymbol{\mu} \sim N_q(\mathbf{0}, \Gamma)$ , where we know also that  $\boldsymbol{\mu} + \mathbf{c} > \mathbf{0}$ . This is equivalent to considering the multivariate normal model (see also Arellano-Valle & Azzalini 2006)

$$\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{V} \end{pmatrix} \sim N_{q+p} \left( \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Gamma & \Gamma A^\top \\ A\Gamma & \Theta + A\Gamma A^\top \end{pmatrix} \right), \quad \text{with } \boldsymbol{\mu} + \mathbf{c} > \mathbf{0}.$$

Thus, the selection model we are considering is  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \boldsymbol{\mu} + \mathbf{c} > \mathbf{0})$  which by (19) has pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \mathbf{a}, \Sigma) \frac{\Phi_q(\mathbf{c} + \Delta^\top \Sigma^{-1}(\mathbf{x} - \mathbf{a}); \mathbf{0}, \Gamma - \Delta^\top \Sigma^{-1} \Delta)}{\Phi_q(\mathbf{c}; \mathbf{0}, \Gamma)}, \tag{22}$$

where  $\Delta = A\Gamma$  and  $\Sigma = \Theta + A\Gamma A^\top$ . A random vector with pdf (22) is said to have a *hierarchical skew-normal (HSN)* distribution and is also in the  $SUN$  family (Arellano-Valle & Azzalini 2006).

3.5. Shape mixture of skewed distributions.

The idea of *shape mixture of skew-symmetric (SMSS)* distributions is introduced by Arellano-Valle, Genton, Gómez & Iglesias (2006). The starting point is a skewed multivariate distribution with pdf of the form  $2^p f_p(\mathbf{x}) \prod_{i=1}^p G(\lambda_i x_i)$  or  $2f_p(\mathbf{x})G(\boldsymbol{\lambda}^\top \mathbf{x})$ , with a (prior) distribution for the shape parameter  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^\top$ . The distributions resulting from integration on  $\boldsymbol{\lambda}$  are  $SMSS$ , and they are in the  $FUSS$  family. When  $f_p = \phi_p$ , the resulting distribution is called a *shape mixture of skew-normal (SMSN)* distribution. Thus, as was shown in Arellano-Valle, Genton, Gómez & Iglesias (2006), a particular  $SMSN$  random vector  $\mathbf{X}$  is obtained by Bayesian modelling by considering that  $X_i | \boldsymbol{\lambda} \sim SN(\lambda_i)$ ,  $i = 1, \dots, p$ , are independent, where  $SN(\lambda)$  denotes the standard univariate skew-normal distribution introduced by

Azzalini (1985), see also Section 3.8, and adopting the prior  $\boldsymbol{\lambda} \sim N_p(\boldsymbol{\mu}, \Sigma)$ . This is equivalent to considering the selection vector  $\mathbf{X} \stackrel{d}{=} (\mathbf{V} | \mathbf{U} \in C(\mathbf{0}))$ , with  $\mathbf{V} \sim N_p(\mathbf{0}, I_p)$  and  $\mathbf{U} | \mathbf{V} = \mathbf{v} \sim N_p(D(\boldsymbol{\mu})\mathbf{v}, I_p + D(\mathbf{v})\Sigma D(\mathbf{v}))$ , where  $D(\boldsymbol{\mu})$  is the  $p \times p$  diagonal matrix with the components of the vector  $\boldsymbol{\mu}$  as the diagonal elements. This example illustrates a *FUSN* case where the marginal distribution of  $\mathbf{V}$  and the conditional distribution of  $\mathbf{U}$  given  $\mathbf{V} = \mathbf{v}$  are normal but the corresponding joint distribution is not normal. The case with  $p = 1$  is discussed in Arellano-Valle, Gómez & Quintana (2004b), which is called skew-generalized normal (*SGN*) distribution and has a pdf given by

$$f_X(x) = 2\phi(x)\Phi\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right),$$

where  $\lambda_1 \in \mathbb{R}$  and  $\lambda_2 \geq 0$ . Note that this family of pdf's contains both the  $N(0, 1)$  pdf when  $\lambda_1 = 0$  and the  $SN(\lambda_1)$  when  $\lambda_2 = 0$ . From a shape mixture point of view it can be obtained by assuming that  $X | \lambda \sim SN(\lambda)$ , with  $\lambda \sim N(\lambda_1, \lambda_2)$ , and from a selection approach it corresponds to the pdf of  $X \stackrel{d}{=} (V | U > 0)$ , where  $V \sim N(0, 1)$  and  $U | V = x \sim N(\lambda_1 x, 1 + \lambda_2 x^2)$ .

### 3.6. Skew-symmetric and generalized skew-elliptical distributions.

For all the skewed distributions considered in this and the next two sections, we have the case with dimension  $q = 1$ , i.e., selection models considering constraints on one random variable  $U$  only. Skew-symmetric (*SS*) distributions have been introduced by Wang, Boyer & Genton (2004a). They are selection distributions with the particular selection subset  $C(\mathbf{0})$  defined in (5), the dimension  $q = 1$ , a symmetric random variable  $U$ , and a symmetric random vector  $\mathbf{V}$ . Therefore, we have  $P(U > 0) = 1/2$ , and the function  $\pi(\mathbf{x}) = P(U > 0 | \mathbf{V} = \mathbf{x})$ , called a *skewing function*, must satisfy  $0 \leq \pi(\mathbf{x}) \leq 1$  and  $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$ . The pdf of a skew-symmetric distribution based on (2) then takes the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2f_{\mathbf{V}}(\mathbf{x})\pi(\mathbf{x}), \quad (23)$$

which is similar to the formulation in Azzalini & Capitanio (2003) except for a different but equivalent representation of  $\pi(\mathbf{x})$ .

Generalized skew-elliptical (*GSE*) distributions have been introduced by Genton & Loperfido (2005). They are skew-symmetric distributions with pdf given by (23) and with the additional assumption that the distribution of  $\mathbf{V}$  is elliptically contoured. When the pdf  $f_{\mathbf{V}}$  is multivariate normal or multivariate  $t$ , the resulting distributions are called *generalized skew-normal* (*GSN*) and *generalized skew-t* (*GST*), respectively.

Both skew-symmetric and generalized skew-elliptical distributions have a simple stochastic representation and possess straightforward properties for their linear transformations and quadratic forms. Genton (2004b) provides a detailed overview of skew-symmetric and generalized skew-elliptical distributions with an illustrative data fitting example.

### 3.7. Flexible skewed distributions.

Flexible skewed distributions have been introduced by Ma & Genton (2004). They are skew-symmetric distributions with a particular skewing function. Indeed, any continuous skewing function can be written as  $\pi(\mathbf{x}) = H(w(\mathbf{x}))$  where  $H: \mathbb{R} \rightarrow [0, 1]$  is the cdf of a continuous random variable symmetric around zero, and  $w: \mathbb{R}^p \rightarrow \mathbb{R}$  is a continuous function satisfying  $w(-\mathbf{x}) = -w(\mathbf{x})$ . By using an odd polynomial  $P_K$  of order  $K$  to approximate the function  $w$ , we obtain a flexible skew-symmetric (*FSS*) distribution with pdf of the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2f_{\mathbf{V}}(\mathbf{x})H(P_K(\mathbf{x})). \quad (24)$$

Under regularity conditions on the pdf  $f_{\mathbf{V}}$ , Ma & Genton (2004) have proved that the class of *FSS* distributions is dense in the class of *SS* distributions under the  $L^\infty$  norm. This means that

*FSS* distributions can approximate *SS* distributions arbitrarily well. Moreover, the number of modes in the pdf (24) increases with the dimension  $K$  of the odd polynomial  $P_K$ , whereas only one mode is possible when  $K = 1$ . When the pdf  $f_{\mathbf{V}}$  is multivariate normal or  $t$ , the resulting distributions are called *flexible skew-normal (FSN)* and *flexible skew- $t$  (FST)*, respectively. Thus, flexible skewed distributions can capture skewness, heavy tails, and multimodality.

### 3.8. Skew-elliptical distributions.

Skew-elliptical (*SE*) distributions have been introduced by Branco & Dey (2001) as an extension of the skew-normal distribution defined by Azzalini & Dalla Valle (1996) to the elliptical class. They are selection distributions considering a random variable  $U$  (i.e.  $q = 1$ ) and a random vector  $\mathbf{V}$ , which have a multivariate elliptically contoured joint distribution, as given in (7), with  $\xi_{\mathbf{U}} = 0$ ,  $\Omega_{\mathbf{U}} = 1$ ; i.e., they are in the *SUE* subclass with dimension  $q = 1$ . Thus, by (21), the *SE* pdf has the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2f_p(\mathbf{x}; \xi_{\mathbf{V}}, \Omega_{\mathbf{V}}, h^{(p)})F_1(\alpha^\top(\mathbf{x} - \xi_{\mathbf{V}}), h_{v(\mathbf{x})}^{(1)}),$$

with  $\alpha = (1 - \Delta^\top \Omega_{\mathbf{V}}^{-1} \Delta)^{-1/2} \Omega_{\mathbf{V}}^{-1} \Delta$ , and  $v(\mathbf{x})$  and  $h_{v(\mathbf{x})}^{(1)}$  defined in Section 3.2.

The idea of an *SE* distribution was also considered by Azzalini & Capitanio (1999), who introduced a different family of *SE* distributions than the above class and suggested exploring their properties as a topic for future research. The relationship between both families is discussed in Azzalini & Capitanio (2003), who have shown that they are equivalent in many important cases, like the skew- $t$  (*ST*) distribution studied by Branco & Dey (2001) and Azzalini & Capitanio (2003), and extended for  $q > 1$  in Arellano-Valle & Azzalini (2006) as a member of *SUE* distributions. The *ST* pdf takes the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2t_p(\mathbf{x}; \xi_{\mathbf{V}}, \Omega_{\mathbf{V}}, \nu)T_1\left(\left(\frac{\nu + 1}{\nu + v(\mathbf{x})}\right)^{1/2} \alpha^\top(\mathbf{x} - \xi_{\mathbf{V}}); \nu + 1\right),$$

where  $t_p(\mathbf{x}; \theta, \Theta, \vartheta)$  is the pdf of the  $p$ -dimensional Student- $t$  distribution with location  $\theta$ , scale  $\Theta$ , and  $\vartheta$  degrees of freedom, and  $T_1(x; \vartheta)$  is the cdf of the standard univariate Student- $t$  distribution with  $\vartheta$  degrees of freedom. Other versions of the skew- $t$  distribution were proposed by Branco & Dey (2002) and Sahu, Dey & Branco (2003). Another important case is of course the multivariate  $SN_p(\xi_{\mathbf{V}}, \Omega_{\mathbf{V}}, \alpha)$  distribution introduced by Azzalini & Dalla Valle (1996), whose pdf has the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2\phi_p(\mathbf{x}; \xi_{\mathbf{V}}, \Omega_{\mathbf{V}})\Phi(\alpha^\top(\mathbf{x} - \xi_{\mathbf{V}})), \quad (25)$$

and is just a member of the *SUN* distributions with dimension  $q = 1$ . If in addition  $p = 1$ , then (25) reduces to the pdf of the seminal univariate skew-normal distribution of Azzalini (1985, 1986); see also Gupta & Chen (2004) for a slightly different definition of a multivariate skew-normal distribution. In all of the above models,  $\alpha$  is a  $p$ -dimensional vector controlling the skewness, and  $\alpha = \mathbf{0}$  recovers the symmetric parent distribution.

Other particular skew-elliptical distributions include skew-Cauchy (Arnold & Beaver 2000b), skew-logistic (Wahed & Ali 2001), skew-exponential power (Azzalini 1986), and skew-slash (Wang & Genton 2006).

A fundamental property of all the selection distributions satisfying  $P(U \in C) = 1/2$  has been mentioned extensively in the literature, see, e.g., Azzalini (2005), Azzalini & Capitanio (2003), Genton (2005), Genton & Loperfido (2005), Wang, Boyer & Genton (2004b) for recent accounts. In brief, the distribution of even functions of a random vector from such a distribution does not depend on  $P(\mathbf{U} \in C | \mathbf{V})$ , i.e., the skewness. This distributional invariance is of particular interest with respect to quadratic forms because of their extensive use in Statistics.

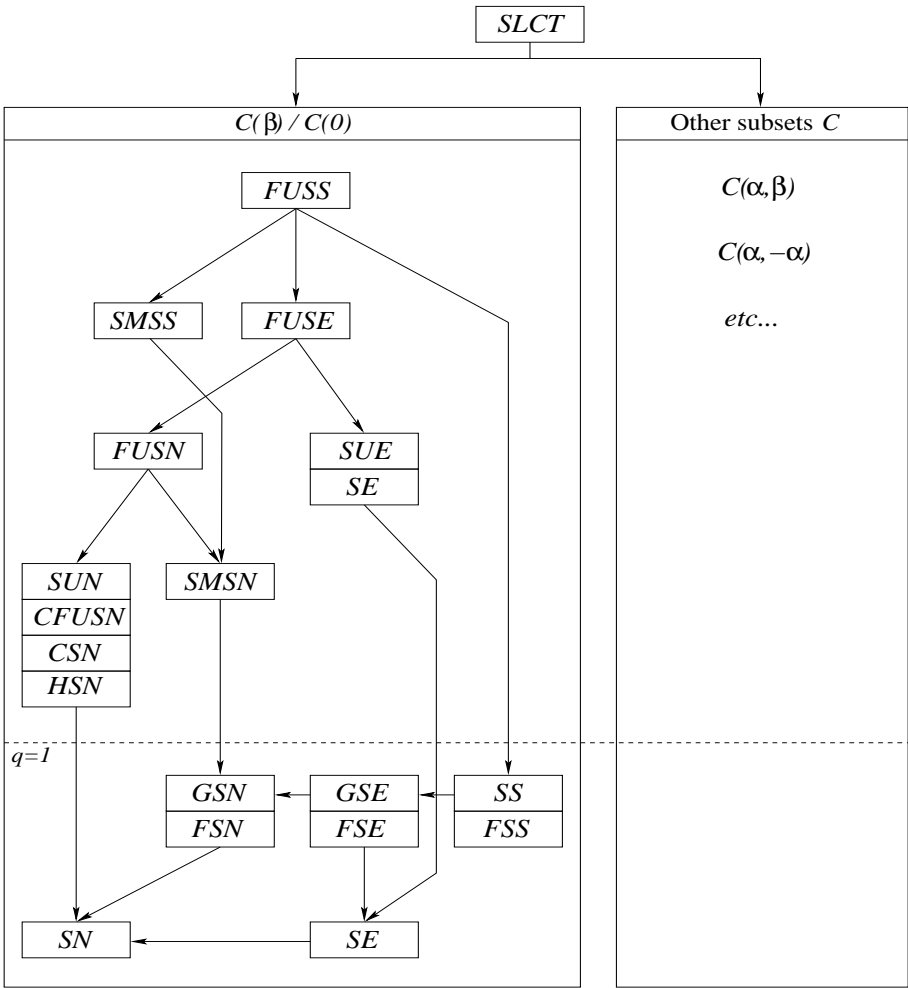


FIGURE 1: Schematic illustrative summary of selection distributions.

3.9. A brief summary of skewed distributions.

The field of skew-normal and related distributions has been growing fast in recent years and many new definitions of univariate and multivariate skewed distributions have been proposed. We believe that new researchers in this field (as well as experts) can be confused by the number of different definitions and their links. In addition to the unified overview described in the previous sections, we propose in Figure 1 a schematic illustrative summary of the main multivariate skewed distributions defined in the literature based on our selection distribution point of view.

The left panel of Figure 1 describes selection distributions based on the selection subset  $C(\beta)$  defined in (5), and essentially contains the definitions of univariate and multivariate skewed distributions described previously. Note that up to a location shift, this is equivalent to using  $C(0)$ . The top part corresponds to  $q > 1$  conditioning variables, whereas the bottom part is for  $q = 1$ . The right panel of Figure 1 describes selection distributions based on arbitrary selection subsets  $C$ , for example such as  $C(\alpha, \beta)$  defined in (6), or  $C(\alpha, -\alpha)$ . Those selection distributions are mainly unexplored except for some general properties derived in Section 2. In Table 1, we provide also an alphabetical summary of the abbreviations of skewed distributions in order to complement Figure 1, as well as some of their main references. Note that for the sake of clarity in Figure 1 we represent only skewed distributions based on normal, elliptically contoured, and symmetric probability density functions. Skewed distributions such as skew- $t$  and skew-Cauchy

are implicitly included in the skew-elliptical families. Similarly, Table 1 is not meant to be exhaustive since so many definitions have appeared in the literature, but it should be useful to trace back some of the main references.

TABLE 1: Alphabetical summary of selection distributions with main references.

ABBREVIATION	NAME OF DISTRIBUTION	MAIN REFERENCES
<i>CFUSN</i>	Canonical fundamental skew-normal	Arellano-Valle & Genton (2005)
<i>CSN</i>	Closed skew-normal	González-Farías et al. (2004a) Arellano-Valle & Azzalini (2006)
<i>FSN, FSS, FST</i>	Flexible skew: -normal, -symmetric, - <i>t</i>	Ma & Genton (2004)
<i>FUSE, FUSN, FUSS</i>	Fundamental skew: -elliptical, -normal, -symmetric	Arellano-Valle & Genton (2005)
<i>GSE, GSN, GST</i>	Generalized skew: -elliptical, normal, - <i>t</i>	Genton & Loperfido (2005)
<i>HSN</i>	Hierarchical skew-normal	Liseo & Loperfido (2006)
<i>SMSN, SMSS</i>	Shape mixture skew: -normal, -symmetric	Arellano-Valle et al. (2006)
<i>SE</i>	Skew-elliptical	Azzalini & Capitanio (1999, 2003) Arnold & Beaver (2000a, 2002) Branco & Dey (2001, 2002) Sahu, Dey & Branco (2003)
<i>SL</i>	Skew-logistic	Wahed & Ali (2001)
<i>SN</i>	Skew-normal	Azzalini (1985, 1986) Azzalini & Dalla Valle (1996) Arnold & Beaver (2000a, 2002) Capitanio et al. (2003) Gupta & Chen (2004)
<i>SS</i>	Skew-symmetric	Wang, Boyer & Genton (2004a)
<i>SSL</i>	Skew-slash	Wang & Genton (2006)
<i>ST</i>	Skew- <i>t</i>	Azzalini & Capitanio (2003) Branco & Dey (2001, 2002)
<i>SUE, SUN</i>	Unified skew: -elliptical, -normal	Arellano-Valle & Azzalini (2006)

#### 4. CONSTRUCTION OF SELECTION DISTRIBUTIONS

As was discussed in Section 2.2, when  $P(\mathbf{U} \in C \mid \mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C)$  for all  $\mathbf{x}$ , we have in (2) that  $f_{\mathbf{X}} = f_{\mathbf{V}}$ , and so the original model is unaffected by the selection mechanism that we are considering. When  $C = C(\mathbf{0})$ , this fact can occur in many situations where  $\mathbf{U}$  and  $\mathbf{V}$  are uncorrelated random vectors, even though they are not independent. In order to assure an explicit association between  $\mathbf{U}$  and  $\mathbf{V}$ , we propose in this section several constructions of explicit selection mechanisms, from which new classes of multivariate symmetric and skewed distributions can be obtained. We start with a random vector  $\mathbf{Z} \in \mathbb{R}^k$  and consider two functions  $u: \mathbb{R}^k \rightarrow \mathbb{R}^q$  and  $v: \mathbb{R}^k \rightarrow \mathbb{R}^p$ . Denote these by  $\mathbf{U} = u(\mathbf{Z})$  and by  $\mathbf{V} = v(\mathbf{Z})$ . We then investigate various selection distributions according to Definition 1 based on  $\mathbf{U}$  and  $\mathbf{V}$ , and a measurable subset  $C$  of  $\mathbb{R}^q$ .

4.1. Linear selection distributions.

The most important case of selection distributions is undoubtedly the one based on linear functions  $u$  and  $v$ , i.e.,  $u(\mathbf{Z}) = A\mathbf{Z} + \mathbf{a}$  and  $v(\mathbf{Z}) = B\mathbf{Z} + \mathbf{b}$ , where  $A \in \mathbb{R}^{q \times k}$  and  $B \in \mathbb{R}^{p \times k}$  are matrices,  $\mathbf{a} \in \mathbb{R}^q$  and  $\mathbf{b} \in \mathbb{R}^p$  are vectors, and we assume  $p \leq k$  and  $q \leq k$ . Under this approach, it is natural to assume that the distribution of  $\mathbf{Z}$  is closed under linear transformations,  $\mathbf{Z} \sim EC_k(\mathbf{0}, I_k, h^{(k)})$  say, implying that  $\mathbf{U} = A\mathbf{Z} + \mathbf{a}$  and  $\mathbf{V} = B\mathbf{Z} + \mathbf{b}$  will have joint distribution as in (7), with  $\xi_{\mathbf{U}} = \mathbf{a}$ ,  $\xi_{\mathbf{V}} = \mathbf{b}$ ,  $\Omega_{\mathbf{U}} = AA^T$ ,  $\Omega_{\mathbf{V}} = BB^T$ , and  $\Delta = AB^T$ . It is then straightforward to establish the connection between (9) and the selection pdf resulting by using the present approach. We now examine the particular case  $f_{\mathbf{Z}}(\cdot) = \phi_k(\cdot)$ , the multivariate standard normal pdf, for the random vector  $\mathbf{Z} \in \mathbb{R}^k$ . We further focus on subsets  $C(\beta)$  of  $\mathbb{R}^q$  defined in (5), with the associated property that, for  $\mathbf{Y} \sim N_r(\xi_{\mathbf{Y}}, \Omega_{\mathbf{Y}})$  with cdf  $\Phi_r(\mathbf{y}; \xi_{\mathbf{Y}}, \Omega_{\mathbf{Y}})$ :

$$\bar{\Phi}_r(C(\beta); \xi_{\mathbf{Y}}, \Omega_{\mathbf{Y}}) = P(\mathbf{Y} > \beta) = P(-\mathbf{Y} < -\beta) = \Phi_r(-\beta; -\xi_{\mathbf{Y}}, \Omega_{\mathbf{Y}}) = \Phi_r(\xi_{\mathbf{Y}}; \beta, \Omega_{\mathbf{Y}}). \tag{26}$$

Using (5) and (26), the pdf (13) simplifies to:

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \mathbf{b}, BB^T) \frac{\Phi_q(\mathbf{a} + AB^T(BB^T)^{-1}(\mathbf{x} - \mathbf{b}); \beta, AA^T - AB^T(BB^T)^{-1}BA^T)}{\Phi_q(\mathbf{a}; \beta, AA^T)}. \tag{27}$$

Apart from some differences in parameterization, the skew-normal pdf in (27) is analogous to the unified skew-normal (SUN) pdf introduced recently by Arellano-Valle & Azzalini (2006). As is shown by these authors, from the SUN pdf most other existing SN probability density functions can be obtained by appropriate modifications in the parameterization.

Another interesting application arises when we consider linear functions of an exchangeable random vector  $\mathbf{Z}$ , which can be used to obtain the density of linear combinations of the order statistics obtained from  $\mathbf{Z}$ ; see Arellano-Valle & Genton (2006a,b).

4.2. Nonlinear selection distributions.

Here we discuss two types of special constructions of a selection distribution by letting: (i)  $\mathbf{V} = g(\mathbf{U}, \mathbf{T})$ ; or (ii)  $\mathbf{U} = g(\mathbf{V}, \mathbf{T})$  where  $g$  is an appropriate Borel function, possibly nonlinear, and  $\mathbf{T} \in \mathbb{R}^r$  is a random vector independent of (or uncorrelated with)  $\mathbf{U}$  or  $\mathbf{V}$ , depending on the situation (i) or (ii), respectively. We obtain stochastic representations as a by-product.

Let  $\mathbf{V} = g(\mathbf{U}, \mathbf{T})$ , where  $\mathbf{U}, \mathbf{T}$  are independent (or uncorrelated) random vectors. For example, the most important case arises when  $g$  is a linear function, i.e.  $\mathbf{V} = A\mathbf{U} + B\mathbf{T}$ . In such situation, the independence assumption implies that  $(\mathbf{V} | \mathbf{U} \in C) \stackrel{d}{=} A(\mathbf{U} | \mathbf{U} \in C) + B\mathbf{T}$  which is a stochastic representation of convolution type. This type of representation, jointly with the independence assumption, has been used by some authors to define skewed distributions (see, e.g., Arnold & Beaver 2002), but considering the particular subset of the form  $C(\beta)$  given in (5) only. However, this definition does not include many other families of skewed distributions, such as the class of unified skew-elliptical distributions, or SUE, introduced by Arellano-Valle & Azzalini (2006). The latter class has also a convolution type representation for the case  $C(\mathbf{0})$ , and does not require the independence assumption. In fact, as was shown by Arellano, del Pino & San Martín (2002) (see also Arellano-Valle & del Pino 2004), the independence assumption can be relaxed by considering that the random vector  $(\mathbf{U}, \mathbf{T})$  is in the  $\mathcal{C}$ -class (renamed  $\mathcal{SI}$ -class in Arellano-Valle & del Pino 2004). Specifically,  $\mathcal{C}$  is the class of all symmetric random vectors  $\mathbf{Z}$ , with  $P(\mathbf{Z} = \mathbf{0}) = 0$  and such that:  $|\mathbf{Z}| = (|Z_1|, \dots, |Z_m|)^T$  and  $\text{sign}(\mathbf{Z}) = (W_1, \dots, W_m)^T$  are independent, and  $\text{sign}(\mathbf{Z}) \sim U_m$ , where  $W_i = +1$ , if  $Z_i > 0$  and  $W_i = -1$ , if  $Z_i < 0$ ,  $i = 1, \dots, m$ , and  $U_m$  is the uniform distribution on  $\{-1, 1\}^m$ . In that case,  $(g(\mathbf{U}, \mathbf{T}) | \mathbf{U} > \mathbf{0}) \stackrel{d}{=} g(\mathbf{U}[\mathbf{0}], \mathbf{T})$ , where  $\mathbf{U}[\mathbf{c}] \stackrel{d}{=} (\mathbf{U} | \mathbf{U} > \mathbf{c})$ . As we show in the next section, other families of asymmetric distributions such as those considered by Fernández & Steel (1998), Mudholkar & Hutson (2000), and Arellano-Valle, Gómez & Quintana (2005c) can also be viewed as selection distributions by using this framework.

Now let  $\mathbf{U} = g(\mathbf{V}, \mathbf{T})$ , with similar conditions as above. Linear and nonlinear truncated conditions on  $\mathbf{V}$  can be introduced by this approach. An important example is when we consider the function  $\mathbf{U} = w(\mathbf{V}) - \mathbf{T}$ , with  $w: \mathbb{R}^p \rightarrow \mathbb{R}^q$ , and the independence assumption between  $\mathbf{V}$  and  $\mathbf{T}$ . In that case, (2) with  $C = C(\mathbf{0})$  reduces to

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x}) \frac{F_{\mathbf{T}}(w(\mathbf{x}))}{P(\mathbf{U} > \mathbf{0})}. \tag{28}$$

If  $\mathbf{U}$  has a symmetric distribution (about  $\mathbf{0}$ ), then  $P(\mathbf{U} > \mathbf{0}) = F_{\mathbf{U}}(\mathbf{0})$ , which holds, e.g., if  $w(-\mathbf{x}) = -w(\mathbf{x})$  and  $\mathbf{V}$  and  $\mathbf{T}$  are symmetric random vectors. In particular, if  $\mathbf{U}$  is a random vector in the  $\mathcal{C}$ -class, then  $F_{\mathbf{U}}(\mathbf{0}) = 2^{-q}$  and (28) yields

$$f_{\mathbf{X}}(\mathbf{x}) = 2^q f_{\mathbf{V}}(\mathbf{x}) F_{\mathbf{T}}(w(\mathbf{x})), \tag{29}$$

which is an extension of a similar result given by Azzalini & Capitanio (1999, 2003). Note that (28) is not altered if the independence condition between  $\mathbf{V}$  and  $\mathbf{T}$  is replaced by the condition that  $\mathbf{V}$  and  $\mathbf{T}$  are conditionally independent given  $w(\mathbf{V})$ . The result can be extended however for the case where  $\mathbf{V}$  and  $\mathbf{T}$  are not independent by replacing the marginal cdf  $F_{\mathbf{T}}$  by the conditional cdf  $F_{\mathbf{T}|\mathbf{V}=\mathbf{x}}$ . Thus, flexible families of multivariate skewed probability density functions as described in Section 3.7 can be obtained from (28) or (29). For example, if  $\mathbf{T} \sim N_q(\mathbf{0}, I_q)$  is independent of  $\mathbf{V}$  with a symmetric probability density function  $f_p$ , then  $\mathbf{U} | \mathbf{V} = \mathbf{x} \sim N_q(w(\mathbf{x}), I_q)$  and its marginal distribution is symmetric since  $w(\cdot)$  is an odd function. In that case, (28) reduces to  $K^{-1} f_p(\mathbf{x}) \Phi_q(w(\mathbf{x}))$ , where  $K = E[\Phi_q(w(\mathbf{V}))]$ . Finally, we note that this approach was simply the starting point for Azzalini (1985). He introduced the univariate family of probability density functions  $2f(x)G(\lambda x)$ ,  $x \in \mathbb{R}$ ,  $\lambda \in \mathbb{R}$ , by considering the random variable  $U = \lambda V - T$ , with the assumption that  $V$  and  $T$  are symmetric (about 0) and independent random variables with  $f_V = f$  and  $F_T = G$ . This implies that  $U$  is a symmetric random variable (about 0) and so  $P(U \geq 0) = P(T \leq \lambda V) = \int_{-\infty}^{\infty} G(\lambda v) f(v) dv = 1/2$  for any value of  $\lambda$ .

On the other hand, as we mentioned in the Introduction, symmetric selection models can also be obtained depending on the conditioning constraints. For example, if  $\mathbf{V} = R\mathbf{V}_0$  is a spherical random vector, i.e.,  $R \stackrel{d}{=} \|\mathbf{V}\|$  and  $\mathbf{V}_0 \stackrel{d}{=} \mathbf{V}/\|\mathbf{V}\|$  and they are independent, then the selection random vector  $(\mathbf{V} | \|\mathbf{V}\| \in C) \stackrel{d}{=} (R | R \in C)\mathbf{V}_0$ . Its distribution is also spherical and reduces to the original spherical distribution, i.e., the distribution of  $\mathbf{V}$  when  $C = C(\mathbf{0})$ . Thus, new families of symmetric distributions can be obtained from the general class of selection distributions, and of course they can be used to obtain many more families of new skewed distributions.

### 4.3. Other classes of skewed distributions.

As was shown by Arellano-Valle, Gómez & Quintana (2005c), another very general family of univariate skewed distributions is defined by the pdf of a random variable  $X$  given by

$$f_X(x) = \frac{2}{a(\alpha) + b(\alpha)} \left[ f\left(\frac{x}{a(\alpha)}\right) I\{x \geq 0\} + f\left(\frac{x}{b(\alpha)}\right) I\{x < 0\} \right], \quad x \in \mathbb{R}, \tag{30}$$

where  $I\{\cdot\}$  denotes the indicator function,  $f(-x) = f(x)$  for all  $x \in \mathbb{R}$ ,  $\alpha$  is an asymmetry parameter (possibly vectorial), and  $a(\alpha)$  and  $b(\alpha)$  are known and positive asymmetric functions. The skewed class of Fernández & Steel (1998) can be obtained from (30) by choosing  $a(\gamma) = \gamma$  and  $b(\gamma) = 1/\gamma$ , for  $\gamma > 0$ . In turn, the  $\varepsilon$ -skew-normal pdf considered in Mudholkar & Hutson (2000) follows when  $a(\varepsilon) = 1 - \varepsilon$  and  $b(\varepsilon) = 1 + \varepsilon$ , with  $|\varepsilon| < 1$ . A particular characteristic of these families is that they all have the same mode as their parent symmetric probability density functions. Moreover, Arellano-Valle, Gómez & Quintana (2005c) showed that (30) corresponds to the pdf of  $X = W_\alpha |Z|$ , where  $Z$  has pdf  $f$  and is independent of  $W_\alpha$ , which is a discrete random variable with  $P(W_\alpha = a(\alpha)) = a(\alpha)/(a(\alpha) + b(\alpha))$  and



$P(W_\alpha = -b(\alpha)) = b(\alpha)/(a(\alpha) + b(\alpha))$ . Thus, since  $|Z| \stackrel{d}{=} (Z | Z > 0)$ , from a selection point of view, (30) can be obtained also from (2) since  $X \stackrel{d}{=} (V | U \in C)$ , with  $V = W_\alpha Z, U = Z$  and  $C = C(0)$ . For an arbitrary selection set  $C$ , more general and flexible families of probability density functions can be obtained following this approach. Moreover, their corresponding location-scale extensions can be derived by considering the usual transformation  $Y = \xi + \tau X$ , with  $\xi \in \mathbb{R}$  and  $\tau > 0$ . Multivariate extensions when the above ingredients are replaced by vectorial quantities and the connection with the  $\mathcal{C}$ -class are discussed in Arellano-Valle, Gómez & Quintana (2005c). Another approach for obtaining multivariate extensions by considering nonsingular affine transformations of independent skewed random variables is discussed by Ferreira & Steel (2004). Finally, a thorough discussion of the connection between the selection framework and the fact that any random variable can be determined by its sign and its absolute value can be found in Arellano-Valle & del Pino (2004).

### 5. DISCUSSION

In this article, we have proposed a unified framework for selection distributions and have described some of their theoretical properties and applications. However, much more research needs to be carried out to investigate unexplored aspects of these distributions, in particular inference. Indeed, the currently available models and their properties are mainly based on the selection subset  $C(\beta)$  defined by (5) and summarized in the left panel of Figure 1. Azzalini (2005) presents an excellent overview of this particular case of selection mechanisms that complements our present article nicely. However, multivariate distributions resulting from other selection mechanisms remain mostly unexplored, see the right panel of Figure 1. The present article serves as a unification of the theory of many known multivariate skewed distributions. Genton (2005) discusses several possibilities for combining semiparametric and nonparametric techniques with selection distributions. Arellano-Valle & Azzalini (2006) gave a unified treatment of most existing multivariate skew-normal distributions and discussed extensions to the skew-elliptical class. They also considered the case of the singular multivariate skew-normal distribution where the key covariance matrix  $\Omega$  does not have full rank.

Selection distributions provide a useful mechanism for obtaining a joint statistical model when we have additional information about the occurrence of some related random events. We have illustrated this point with various examples. There is an unlimited potential for other applications necessitating similar models describing selection procedures. For instance, Chen (2004) describes applications of multivariate skewed distributions as a link function in discrete selection models, and Bázan, Branco & Bolfarine (2006) propose a new skewed family of models for item response theory. We expect to see an important growth of applications of selection distributions in the near future.

### APPENDIX

*Moment generating function of SLCT- $N_{p,q}$  distributions.* The moment generating function of  $\mathbf{X} \sim \text{SLCT-}N_{p,q}$  follows directly from the form of (13). We have:

$$\begin{aligned}
 M_{\mathbf{X}}(\mathbf{s}) &= E[\exp\{\mathbf{s}^\top \mathbf{X}\}] \\
 &= \int \exp(\mathbf{s}^\top \mathbf{x}) \phi_p(\mathbf{x}; \boldsymbol{\xi}_V, \Omega_V) \frac{\bar{\Phi}_q(C; \Delta^\top \Omega_V^{-1}(\mathbf{x} - \boldsymbol{\xi}_V) + \boldsymbol{\xi}_U, \Omega_U - \Delta^\top \Omega_V^{-1} \Delta)}{\bar{\Phi}_q(C; \boldsymbol{\xi}_U, \Omega_U)} d\mathbf{x} \\
 &= \exp\left\{\mathbf{s}^\top \boldsymbol{\xi}_V + \frac{1}{2} \mathbf{s}^\top \Omega_V \mathbf{s}\right\} \frac{1}{\bar{\Phi}_q(C; \boldsymbol{\xi}_U, \Omega_U)} \\
 &\quad \times \int \phi_p(\mathbf{x}; \boldsymbol{\xi}_V + \Omega_V \mathbf{s}, \Omega_V) \bar{\Phi}_q(C; \Delta^\top \Omega_V^{-1}(\mathbf{x} - \boldsymbol{\xi}_V) + \boldsymbol{\xi}_U, \Omega_U - \Delta^\top \Omega_V^{-1} \Delta) d\mathbf{x} \\
 &= \exp\left\{\mathbf{s}^\top \boldsymbol{\xi}_V + \frac{1}{2} \mathbf{s}^\top \Omega_V \mathbf{s}\right\} \frac{1}{\bar{\Phi}_q(C; \boldsymbol{\xi}_U, \Omega_U)} E[P(\tilde{U} \in C | \tilde{V} = \mathbf{x})],
 \end{aligned}$$

where  $\tilde{\mathbf{U}} | \tilde{\mathbf{V}} = \mathbf{x} \sim N_q(\Delta^\top \Omega_{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\xi}_{\mathbf{V}}) + \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}} - \Delta^\top \Omega_{\mathbf{V}}^{-1} \Delta)$  and  $\tilde{\mathbf{V}} \sim N_p(\boldsymbol{\xi}_{\mathbf{V}} + \Omega_{\mathbf{V}} \mathbf{s}, \Omega_{\mathbf{V}})$ , and thus  $\tilde{\mathbf{U}} \sim N_q(\Delta^\top \mathbf{s} + \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}})$ . The result is then obtained from the relation  $E[\mathbf{P}(\tilde{\mathbf{U}} \in C | \tilde{\mathbf{V}} = \mathbf{x})] = \mathbf{P}(\tilde{\mathbf{U}} \in C) = \bar{\Phi}_q(C; \Delta^\top \mathbf{s} + \boldsymbol{\xi}_{\mathbf{U}}, \Omega_{\mathbf{U}})$ .  $\square$

*Moment generating function of scale mixture of SLCT- $N_{p,q}$  distributions.* The moment generating function of  $\mathbf{X}$  such that  $(\mathbf{X} | W = w) \sim \text{SLCT-}N_{p,q}(\boldsymbol{\xi}, w\Omega, C(\boldsymbol{\xi}_{\mathbf{U}}))$ , where  $W \sim G$ , follows directly from the form of (17) and (14). We have:

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{s}) &= \exp\{\mathbf{s}^\top \boldsymbol{\xi}_{\mathbf{V}}\} E[M_{\mathbf{X}_N}(\sqrt{W} \mathbf{s})] \\ &= \exp\{\mathbf{s}^\top \boldsymbol{\xi}_{\mathbf{V}}\} \frac{\int_0^\infty \exp\{\frac{1}{2} w \mathbf{s}^\top \Omega_{\mathbf{V}} \mathbf{s}\} \Phi_q(\sqrt{w} \Delta^\top \mathbf{s}; \mathbf{0}, w\Omega_{\mathbf{U}}) dG(w)}{\Phi_q(\mathbf{0}; \mathbf{0}, \Omega_{\mathbf{U}})} \\ &= \exp\{\mathbf{s}^\top \boldsymbol{\xi}_{\mathbf{V}}\} \int_0^\infty \exp\left\{\frac{w}{2} \mathbf{s}^\top \Omega_{\mathbf{V}} \mathbf{s}\right\} dG(w) \frac{\Phi_q(\Delta^\top \mathbf{s}; \mathbf{0}, \Omega_{\mathbf{U}})}{\Phi_q(\mathbf{0}; \mathbf{0}, \Omega_{\mathbf{U}})} \\ &= E[M_{\mathbf{V}|W}(\mathbf{s})] \frac{\Phi_q(\Delta^\top \mathbf{s}; \mathbf{0}, \Omega_{\mathbf{U}})}{\Phi_q(\mathbf{0}; \mathbf{0}, \Omega_{\mathbf{U}})}. \end{aligned}$$

$\square$

## ACKNOWLEDGEMENTS

The authors are grateful to the Editor, the Associate Editor and two anonymous referees for insightful comments on the manuscript. This project was partially supported by a grant from the Fondo Nacional de Desarrollo Científico y Tecnológico in Chile. The work of Branco was partially supported by the Brazilian Conselho Nacional de Desenvolvimento Científico and Tecnológico (CNPq). The work of Genton was partially supported by a grant from the United States National Science Foundation.

## REFERENCES

- R. B. Arellano-Valle & A. Azzalini (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33, 561–574.
- R. B. Arellano-Valle & G. E. del Pino (2004). From symmetric to asymmetric distributions: a unified approach. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 113–130.
- R. B. Arellano-Valle, G. E. del Pino & E. San Martín (2002). Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*, 58, 111–121.
- R. B. Arellano-Valle & M. G. Genton (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, 96, 93–116.
- R. B. Arellano-Valle & M. G. Genton (2006a). On the exact distribution of the maximum of dependent random variables. *Statistics & Probability Letters*, revised version submitted for publication.
- R. B. Arellano-Valle & M. G. Genton (2006b). On the exact distribution of linear combinations of order statistics from dependent random variables. *Journal of Multivariate Analysis*, revised version submitted for publication.
- R. B. Arellano-Valle, M. G. Genton, H. W. Gómez & P. Iglesias (2006). Shape mixtures of skewed distributions, with application in regression analysis. Submitted for publication.
- R. B. Arellano-Valle, H. W. Gómez & F. A. Quintana (2004b). A new class of skew-normal distributions. *Communications in Statistics: Theory and Methods*, 33, 1465–1480.
- R. B. Arellano-Valle, H. W. Gómez & F. A. Quintana (2005c). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128, 427–443.
- B. C. Arnold & R. J. Beaver (2000a). Hidden truncation models. *Sankyā Series A*, 62, 22–35.
- B. C. Arnold & R. J. Beaver (2000b). The skew-Cauchy distribution. *Statistics & Probability Letters*, 49, 285–290.
- B. C. Arnold & R. J. Beaver (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11, 7–54.

- B. C. Arnold & R. J. Beaver (2004). Elliptical models subject to hidden truncation or selective sampling. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 101–112.
- B. C. Arnold, R. J. Beaver, R. A. Groeneveld & W. Q. Meeker (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58, 471–478.
- B. C. Arnold, E. Castillo & J. M. Sarabia (1999). *Conditional Specification of Statistical Models*. Springer-Verlag, New York.
- B. C. Arnold, E. Castillo & J. M. Sarabia (2001). Conditionally specified distributions: an introduction (with discussion). *Statistical Science*, 16, 249–274.
- A. Azzalini (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.
- A. Azzalini (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199–208.
- A. Azzalini (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32, 159–188.
- 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. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew  $t$  distribution. *Journal of the Royal Statistical Society Series B*, 65, 367–389.
- A. Azzalini & A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.
- M. J. Bayarri & M. DeGroot (1992). A BAD view of weighted distributions and selection models. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, April 15–20, 1991* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds.), Oxford University Press, pp. 17–29.
- J. L. Bázan, M. D. Branco & H. Bolfarine (2006). A skew item response model. *Bayesian Analysis*, in press.
- Z. W. Birnbaum (1950). Effect of linear truncation on a multinormal population. *Annals of Mathematical Statistics*, 21, 272–279.
- M. D. Branco & D. K. Dey (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79, 99–113.
- M. D. Branco & D. K. Dey (2002). Regression model under skew elliptical error distribution. *The Journal of Mathematical Sciences*, 1, 151–169.
- A. Capitanio, A. Azzalini & E. Stanghellini (2003). Graphical models for skew-normal variates. *Scandinavian Journal of Statistics*, 30, 129–144.
- M.-H. Chen (2004). Skewed link models for categorical response data. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 131–151.
- J. B. Copas & H. G. Li (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society Series B*, 59, 55–95.
- C. Crocetta & N. Loperfido (2005). The exact sampling distribution of  $L$ -statistics. *Metron*, 63, 1–11.
- T. J. DiCiccio & A. C. Monti (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99, 439–450.
- C. Fernández & M. F. Steel (1998). On Bayesian modelling 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 skewed regression modelling with an application to firm size. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 175–190.
- M. G. Genton, ed. (2004a). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman & Hall/CRC, Boca Raton, Florida.
- M. G. Genton (2004b). Skew-symmetric and generalized skew-elliptical distributions. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 81–100.

- M. G. Genton (2005). Discussion of: The skew-normal distribution and related multivariate families, by A. Azzalini. *Scandinavian Journal of Statistics*, 32, 189–198.
- M. G. Genton & N. Loperfido (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, 57, 389–401.
- G. González-Farías, J. A. Domínguez-Molina & A. K. Gupta (2004a). The closed skew-normal distribution. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.), Chapman & Hall/CRC, Boca Raton, Florida, pp. 25–42.
- G. González-Farías, J. A. Domínguez-Molina & A. K. Gupta (2004b). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, 126, 521–534.
- A. K. Gupta & J. T. Chen (2004). A class of multivariate skew-normal models. *Annals of the Institute of Statistical Mathematics*, 56, 305–315.
- J. Heckman (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153–161.
- N. L. Johnson, S. Kotz & N. Balakrishnan (1994). *Continuous Univariate Distributions 1*, Wiley, New York.
- B. Liseo & N. Loperfido (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference*, 136, 373–389.
- N. Loperfido (2002). Statistical implications of selectively reported inferential results. *Statistics & Probability Letters*, 56, 13–22.
- Y. Ma & M. G. Genton (2004). A flexible class of skew-symmetric distributions. *Scandinavian Journal of Statistics*, 31, 459–468.
- Y. Ma, M. G. Genton & A. A. Tsiatis (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *Journal of the American Statistical Association*, 100, 980–989.
- Y. Ma & J. Hart (2006). Constrained local likelihood estimators for semiparametric skew-normal distributions. *Biometrika*, in press.
- K. G. Malmquist (1920). A study of the stars of spectral type A. *Meddelande från Lunds Astronomiska Observatorium, Serie II*, 22, 1–39.
- G. S. Mudholkar & A. D. Hutson (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83, 291–309.
- L. S. Nelson (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6, 469–471.
- A. O'Hagan & T. Leonard (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201–202.
- A. Pewsey (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859–870.
- A. Pewsey (2006). Some observations on a simple means of generating skew distributions. In *Advances in Distribution Theory, Order Statistics and Inference*, (B. C. Arnold, N. Balakrishnan, E. Castillo & J.-M. Sarabia, eds.), Birkhäuser, Boston, pp. 75–84.
- C. R. Rao (1985). Weighted distributions arising out of methods of ascertainment: What populations does a sample represent? In *A Celebration of Statistics: The ISI Centenary Volume* (A. C. Atkinson, S. E. Fienberg, eds.), Springer-Verlag, New York, pp. 543–569.
- C. Roberts (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61, 1184–1190.
- S. K. Sahu, D. K. Dey & M. Branco (2003). A new class of multivariate skew distributions with application to Bayesian regression models. *The Canadian Journal of Statistics*, 31, 129–150.
- N. Sartori (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew  $t$  distributions. *Journal of Statistical Planning and Inference*, 136, 4259–4275.
- A. Wahed & M. M. Ali (2001). The skew-logistic distribution. *Journal of Statistical Research*, 35, 71–80.
- J. Wang, J. Boyer & M. G. Genton (2004a). A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, 14, 1259–1270.
- J. Wang, J. Boyer & M. G. Genton (2004b). A note on an equivalence between chi-square and generalized skew-normal distributions. *Statistics & Probability Letters*, 66, 395–398.

- J. Wang & M. G. Genton (2006). The multivariate skew-slash distribution. *[Journal of Statistical Planning and Inference](#)*, 136, 209–220.
- M. A. Weinstein (1964). The sum of values from a normal and a truncated normal distribution. *[Technometrics](#)*, 6, 104–105.
- 

Received 21 December 2005

Accepted 25 July 2006

Reinaldo B. ARELLANO-VALLE: reivalle@mat.puc.cl

*Departamento de Estadística  
Pontificia Universidad Católica de Chile  
Santiago 22, Chile*

Márcia D. BRANCO: mbranco@ime.usp.br

*Departamento de Estatística  
Universidade São Paulo  
São Paulo, SP 05315-970, Brasil*

Marc G. GENTON: genton@stat.tamu.edu

*Department of Statistics  
Texas A&M University  
College Station, TX 77843-3143, USA*