Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* 11, *No.* 1, pp. 7–54

Skewed multivariate models related to hidden truncation and/or selective reporting

Barry C. Arnold and Robert J. Beaver
Department of Statistics.
University of California. Riverside, USA

Abstract

The univariate skew-normal distribution was introduced by Azzalini in 1985 as a natural extension of the classical normal density to accommodate asymmetry. He extensively studied the properties of this distribution and in conjunction with coauthors, extended this class to include the multivariate analog of the skew-normal. Arnold et al. (1993) introduced a more general skew-normal distribution as the marginal distribution of a truncated bivariate normal distribution in which X was retained only if Y satisfied certain constraints. Using this approach more general univariate and multivariate skewed distributions have been developed. A survey of such models is provided together with discussion of related inference questions.

Key Words: Skew-normal, skew-Cauchy, skew-Laplace, skew multivariate distributions.

AMS subject classification: 62F03, 62A05.

1 Introduction

Skewed multivariate distributions may be encountered in situations in which observations obey a relatively well behaved law (perhaps multivariate normal) but have been truncated with respect to some hidden covariable. A simple univariate example of such a phenomenon involves the distribution of waist sizes for uniforms of elite troops who are selected only if they meet a specific minimal height requirement. A classical bivariate normal model might well be acceptable for 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.

*Correspondence to: Barry C. Arnold and Robert J. Beaver. Department of Statistics. University of California. Riverside, CA 92521-0128 USA. Email: barnold@ucrstat2.ucr.edu

Received: March 2002; Accepted: April 2002

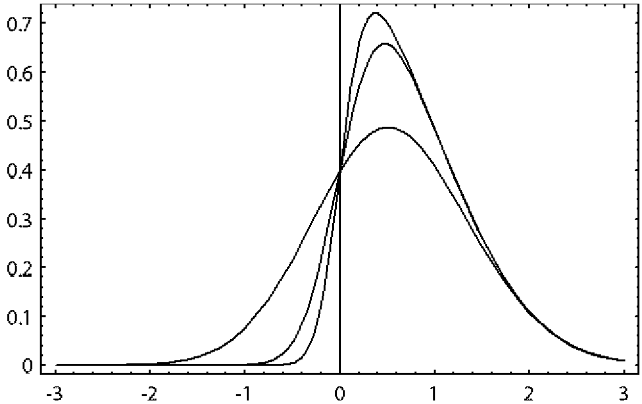

Figure 1: Probability density function in (2.1) for $\lambda = 1, 3$ and 5.

A normal model will not adequately fit the waist size data in this setting. In many situations, truncation such as that just described will have occurred without us having knowledge of its occurrence. We thus envision the possibility that quite often some "hidden" truncation will have occurred. The present paper will survey a spectrum of distributions available to help us adequately model (and make inferences about) such data configurations. We begin by considering models involving underlying normal distributions, so called skew-normal models. As we shall see, straightforward extensions to non-normal models are feasible.

Azzalini (1985, 1986) introduces the basic skew-normal model. His name, together with the names of his coworkers, will occur frequently in the subsequent discussion. An independent development, using different notation, can be traced in a series of papers beginning with Arnold et al. (1993). The words "selective reporting" appearing in the title can be related to our hidden truncation scenario in two ways. First it may be that the investigators whose data is being analyzed have indeed selectively reported their data without our knowing it (perhaps for example only measuring and reporting archeological pottery shards that were large enough to not fall through the sieve used on-site). They may indeed themselves be also unaware that they selectively reported the data (or equivalently subjecting it to a hidden truncation). In this sense selective reporting and hidden truncation are synonyms. But there is another form of selective reporting that is not as obviously related to hidden truncation. Consider the following univariate scenario. Suppose normally distributed observations are available. But now suppose that each reported observation is actually the maximum of two independent identically distributed normal observations (i.e. we pick the "best" one of each pair of observations). The resulting distribution in such a setting is again skewed normal. A suitable higher dimensional analog of this selective reporting scenario will yield a skewed multivariate normal model just the same as the model observed via hidden truncation. The bottom line of this discussion is that hidden truncation may well be quite prevalent in data analysis settings and it will behoove us to have suitable models available to account for such phenomena.

2 The basic univariate skew-normal model

A random variable Z will be said to have a skew-normal distribution with parameter λ if its density function is of the form (see Chiogna (1998)):

$$f(z;\lambda) = 2\varphi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \tag{2.1}$$

where $\varphi(z)$ and $\Phi(z)$ denote the standard normal ($N(0,1)$) density and distribution functions respectively. The expression (2.1) is a valid density for any $\lambda \in \mathbb{R}$. The density is symmetric if $\lambda = 0$ (in which case it coincides with the standard normal density). It is reasonable to call λ the skewness parameter. The family of densities (2.1) was apparently first discussed in detail in Azzalini (1985). The suggestion that (2.1) might provide a reasonable family of skewed alternatives to the classical normal model certainly is due to Azzalini. It is quite likely that the family (2.1) did appear in earlier research on normal order statistics but, not withstanding any such earlier references which may be encountered, it seems appropriate to call (2.1) the Azzalini skew-normal distribution. See the note at the end of this section for some historical precedents.

Examples of the pdf (2.1) for $\lambda = 1, 3$ and 5 are given in Figure 1.

Azzalini and Dalla Valle (1996) discuss three interesting settings in which skew-normal densities of the form (2.1) can arise. All three scenarios can be fruitfully employed in the generation of extensions to the basic Azzalini model and consequently merit our consideration.

First consider two independent identically distributed standard normal random variables Y and W. Now define Z to be equal to Y conditionally on the event $\{\lambda Y > W\}$. The resulting distribution of Z will be given by

$$\begin{aligned} P(Z \le z) &= P(Y \le z|\lambda Y > W) \\ &= P(Y \le z, \lambda Y > W)/P(\lambda Y > W) \\ &= \int_{-\infty}^{z}\int_{-\infty}^{\lambda y} \varphi(y)\varphi(w)dwdy/P(\lambda Y > W) \\ &= \int_{-\infty}^{z} \varphi(y)\Phi(\lambda y)dy/P(\lambda Y > W). \end{aligned} \quad (2.2)$$

Since $P(\lambda Y > W) = P(\lambda Y - W > 0) = 1/2$ (because $\lambda Y - W$ has a normal distribution with mean zero), it follows on differentiating (2.2) with respect to z that Z has the skew-normal density (2.1).

The second scenario involves hidden truncation. Suppose that (Z, Y) has a bivariate normal distribution with standardized ($N(0,1)$) marginals and correlation δ. In this context consider the conditional density of Z given $Y > 0$ (i.e. keeping only Z's whose corresponding Y value is above average). It follows that

$$\begin{aligned} P(Z \le z|Y > 0) &= P(Z \le z, Y > 0)/P(Y > 0) \\ &= 2P(Z \le z, Y > 0). \end{aligned}$$

Now, without loss of generality, we may assume that $Y = \delta Z - \sqrt{1-\delta^2}W$ where W and Z are i.i.d. $N(0,1)$ variables. Consequently the event $\{Y > 0\}$ is equivalent to $\{\delta Z - \sqrt{1-\delta^2}W > 0\}$ or $\{\frac{\delta}{\sqrt{1-\delta^2}}Z > W\}$ and it follows that the conditional distribution of Z given $\{Y > 0\}$ is skew-normal (λ) where $\lambda = \delta/\sqrt{1-\delta^2}$.

The third development begins with the i.i.d. standard normal random variables Y_1, Y_2 and a constant $\delta \in (-1, 1)$. If we define

$$Z = \delta|Y_1| + \sqrt{1-\delta^2}Y_2$$

then a simple convolution computation may be used to verify that Z has a skew-normal (λ) distribution (with $\lambda = \delta/\sqrt{1-\delta^2}$ again) (Henze (1986)).

On reflection, scenarios 1 and 2 are essentially equivalent while scenario 3 appears to be different. This will become more apparent when we consider extensions to multivariate and/or non-normal cases below.

A fourth derivation of the skew-normal distribution is the one alluded to in Section 1 involving dependent maxima. For it, consider (W_1, W_2) to be a bivariate normal random vector with standard normal marginals and correlation δ. Define now

$$Z = \max(W_1, W_2).$$

Loperfido (2001) observed that in this case Z has a skew-normal (λ) distribution with $\lambda = \sqrt{(1-\delta)/(1+\delta)}$.

Before considering multivariate and non-normal extensions of these alternative skew-normal genesis scenarios, we remark that the hidden truncation scenario admits a simple but very useful extension. Our original hidden truncation discussion involved retention of X values subject to their concomitant Y value being at least as large as a specific threshold. The corresponding family of densities can be viewed as being a location-scale extension of the following basic skewed density

$$f(z; \lambda_0, \lambda_1) = \frac{\varphi(z)\Phi(\lambda_0 + \lambda_1 z)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)}, \tag{2.3}$$

where $\lambda_0 \in \mathbb{R}$ and $\lambda_1 \in \mathbb{R}$. The Azzalini distribution (2.1) corresponds to the choice $\lambda_0 = 0$ in (2.3) and may be interpreted as having arisen by keeping observations only if the corresponding concomitant variable is above average. Use of a more general threshold for retention will result in the more general family of densities (2.3). The model (2.3) is discussed extensively in Arnold et al. (1993). The model (2.3) can alternatively be envisioned as having arisen as follows. Begin with W and U i.i.d. standard normal variables. Consider the conditional density of W given that $\{\lambda_0 + \lambda_1 W > U\}$. This density will be that one displayed in (2.3). The family (2.3) in addition to its obvious increased flexibility for modelling varieties of skewness will also turn out to be important in that it will permit us to construct multivariate skewed models that have marginal and conditional densities that are again of the same general multivariate skewed form.

One example of the pdf (2.3) for $\lambda_0 = \lambda_1 = 2$ is shown in Figure 2.

It is natural to add further flexibility to the model (2.3) (or the submodel (2.1)) by the introduction of location and scale parameters. In this way we arrive at a 4 parameter skewed normal density of the form

$$f(z; \lambda_0, \lambda_1, \mu, \sigma) = \frac{\varphi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\lambda_0 + \lambda_1\left(\frac{z-\mu}{\sigma}\right)\right)}{\sigma\Phi\left(\frac{\lambda_0}{\sqrt{1-\lambda_1^2}}\right)} \tag{2.4}$$

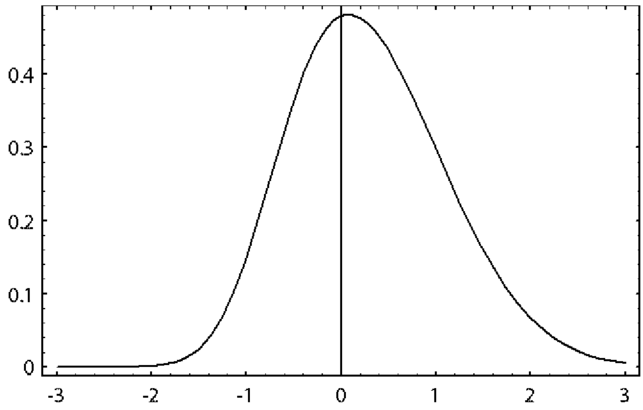

Figure 2: Probability density function in (2.3) for $\lambda_0 = \lambda_1 = 2$.

where $\mu, \lambda_0, \lambda_1 \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. A random variable will have density (2.4) if it admits a representation as $\mu + \sigma Z$ where Z has density (2.3).

An elementary computation (see Arnold et al. (1993)) yields the moment generating function of the density (2.3) in the form

$$M(t) = e^{t^2/2} \frac{\Phi\left(\frac{\lambda_0 + \lambda_1 t}{\sqrt{1+\lambda_1^2}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)}. \tag{2.5}$$

Moments can then be readily obtained by differentiation of (2.5).

Remark 2.1. Cain (1994) derived the moment generating function of the minimum of X_1, X_2 where (X_1, X_2) has a general bivariate normal density. His results show that the distribution of $\min(X_1, X_2)$ is a mixture of two skew-normal densities of the form (2.3). If $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ then $\min(X_1, X_2)$ indeed has a skew-normal density (2.3).

Remark 2.2 (Historical Note). The density (2.3) has made cameo appearances in several earlier papers. It may be found in Birnbaum (1950), introduced in a psychological testing context (see the first equation of his Section 2). He described the phenomenon as linear truncation of a multivariate normal random variable. It also appears as equation (1) in O'Hagan and Leonard (1976) where it is suggested as a possible prior for a normal mean. See also Roberts (1966) who discusses the distribution of the minimum of two correlated normal variates. Nelson (1964) and Weinstein (1964) discuss the distribution of the sum of a normal variable and an independent truncated normal variable. This turns out to have density (2.3). See the remark following equation (4.5) for an extension of this paradigm. Perhaps readers will be able to suggest even earlier sources for discussion of the model (2.3).

3 Non-normal univariate skewed distributions

It is evident that the role played by normal variates in the development of Azzalini's distribution (2.1) and in the development of (2.3) could be played by other distributions. For example in the hidden truncation development we could begin with two independent random variables W, U but now do not assume they are standard normal variables. Instead assume that W has density (distribution) function $\psi_1(\Psi_1)$ and that U has density (distribution) function $\psi_2(\Psi_2)$. In this setting the conditional density of W given that $\{\lambda_0 + \lambda_1 W > U\}$ is of the form:

$$f(w; \lambda_0, \lambda_1) = \frac{\psi_1(w)\Psi_2(\lambda_0 + \lambda_1 w)}{P(\lambda_0 + \lambda_1 W > U)} \ . \tag{3.1}$$

Simplification occurs in (3.1) when $\lambda_0 = 0$ and when ψ_1 and ψ_2 are symmetric densities. In such a situation, the denominator in (3.1) equals 1/2 and our density is given by

$$f(w; \lambda_1) = 2\psi_1(w)\Psi_2(\lambda_1 w) \ , \tag{3.2}$$

analogous to Azzalini's original skew-normal model (2.1).

In the general formulation (3.1), computation of the denominator, $P(\lambda_0 + \lambda_1 W > U)$, can be troublesome. If W and U are stable random variables of the same type then $\lambda_0 + \lambda_1 W - U$ will also be stable and in principle the denominator in (3.1) can be calculated. Arnold and Beaver (2000b) discuss the Cauchy case. Balakrishnan and Ambagaspitiya (1994) consider the Laplace case in which the denominator of (3.1) is equal to 1/2 when λ_0 is assumed to be equal to 0.

It will be recalled that hidden truncation was not the only scenario leading to Azzalini's model (2.1). It is instructive to consider non-normal variants of the other possible derivation of (2.1). One approach began with Y_1 and Y_2 i.i.d. standard normal variables and considered

$$Z = \delta|Y_1| + \sqrt{1-\delta^2}Y_2 \ . \tag{3.3}$$

Arnold and Beaver (2000c) consider the models derivable from (3.3) by permitting Y_1 and Y_2 to be i.i.d. with common non-normal density ψ. More generally we could allow Y_1 and Y_2 in (3.3) to have different densities ψ_1 and ψ_2. Variant skew-Cauchy and skew-Laplace distributions obtained in this fashion are discussed in Arnold and Beaver (2000c).

The other model leading to the Azzalini density (2.1) begins also with two i.i.d. standard normal random variables (Y_1, Y_2) but then considers a linear transformation $\underline{U} = A\underline{Y}$, where $\underline{Y} = (Y_1, Y_2)$ and $\underline{U} = (U_1, U_2)$, chosen to have standard normal marginals but non-zero correlation. One then considers $Z = \max(U_1, U_2)$, which turns out to have density of the form (2.1). Here too we can allow Y_1 and Y_2 to have densities ψ_1 and ψ_2 (instead of assuming that each has a standard normal density). This leads to new classes of skewed distributions.

4 Multivariate skew-normal models

In the hidden truncation setting, it is quite plausible that we might have k observed variables (instead of 1) that have been truncated with respect to an unobserved but related variable. A skewed observed multivariate distribution will then be encountered. More formally we may begin with a $(k+1)$ dimensional random vector $(X_1, \ldots, X_k, Y)$ with a general covariance structure. Then set $\underline{Z} = \underline{X}$ conditional on $Y > a$. Azzalini and Dalla Valle (1996) were the first to consider such a model though they restricted attention to be the case $a = \mu_Y$ (retaining observations only if their Y values were above average). This provides a natural extension of the basic skew-normal model (2.1). Allowing a to be arbitrary (as was done in Arnold and Beaver (2000a)) yields a natural extension of the more general skewed normal model (2.3). As we shall see, the general k-dimensional model has the attractive feature of being "closed under conditioning", i.e. such models have not only marginals but also conditional distributions of the same type.

To closely parallel the univariate development we begin with

$$W_1, W_2, \ldots, W_k, U$$

which are i.i.d. $N(0,1)$ random variables. Now consider the conditional distribution of $\underline{W}$ given that $\lambda_0 + \underline{\lambda}'\underline{W} > U$ where $\lambda_0 \in \mathbb{R}$ and $\underline{\lambda}_1 \in \mathbb{R}^k$. This formulation which involves a linear transformation of the hidden truncation model described in the previous paragraph, is particularly easy to analyze. If we define the event A by

$$A = \{\lambda_0 + \underline{\lambda'}_1 \underline{W} > U\},$$

then obviously the conditional density of $(\underline{W}, U)$ given A is

$$f_{\underline{W},U|A}(\underline{w}, u) \propto \left[\prod_{i=1}^{k} \varphi(w_i)\right] \varphi(u) I(\lambda_0 + \underline{\lambda}_1' \underline{w} > u) \ . \tag{4.1}$$

Integrating with respect to u, we get

$$f_{\underline{W}|A}(\underline{w}) = \frac{\left[\prod_{i=1}^{k} \varphi(w_i)\right] \Phi(\lambda_0 + \underline{\lambda}_1' \underline{w})}{P(A)} \ . \tag{4.2}$$

However,

$$P(A) = P(U - \underline{\lambda}_1' \underline{W} < \lambda_0) = \Phi\left(\frac{\lambda_0}{\sqrt{1 + \underline{\lambda}_1' \underline{\lambda}_1}}\right),$$

since $U - \underline{\lambda}_1' \underline{W} \sim N(0, 1 + \underline{\lambda}_1' \underline{\lambda}_1)$. Thus the k-dimensional hidden truncation density analogous to (2.3) is given by

$$f(\underline{w}; \lambda_0, \underline{\lambda}_1) = \frac{\left[\prod_{i=1}^{k} \varphi(w_i)\right] \Phi(\lambda_0 + \underline{\lambda}_1' \underline{w})}{\Phi\left(\frac{\lambda_0}{\sqrt{1 + \underline{\lambda}_1' \underline{\lambda}_1}}\right)}. \tag{4.3}$$

One example of the pdf (4.3) for $\lambda_0 = 1$; $\underline{\lambda_1} = (2, 3)$ and the corresponding contours are shown in Figure 3.

The Azzalini and Dalla Valle (1996) density can be identified with the choice $\lambda_0 = 0$ in (4.3). Note that in this case the denominator in (4.3) is equal to $1/2$ and the resulting density takes the form

$$f(\underline{w}; 0, \underline{\lambda}_1) = 2\left[\prod_{i=1}^{k} \varphi(w_i)\right] \Phi(\underline{\lambda}_1' \underline{w}) \ . \tag{4.4}$$

Azzalini and Dalla Valle observed that the joint density (4.4) has marginal densities of the same type. We will show below that this is true for the

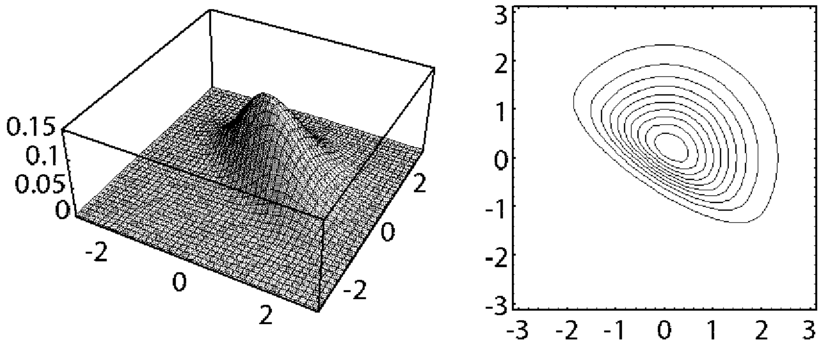

Figure 3: Probability density function in (4.3) for $\lambda_0 = 1; \underline{\lambda_1} = (2, 3)$ and corresponding contours.

extended model (4.3) and moreover we will verify that (4.3) also has conditional densities of the same type.

An alternative genesis of model (4.4) was described by Azzalini and Dalla Valle. For it, we can begin with $(k+1)$ i.i.d. $N(0,1)$ random variables, this time labelled $Y_0, Y_1, \ldots, Y_k$. Now we define $\underline{W}$ as follows

$$W_j = \delta_j |Y_0| + \sqrt{1 - \delta_j^2}\, Y_j, \quad j = 1, 2, \ldots, k\ . \tag{4.5}$$

The resulting joint density of $\underline{W}$ is k-variate skewed normal (in other words there exists a $k \times k$ matrix B such that $B\underline{W}$ has density (4.4)).

Remark 4.1. Suppose we begin with $X_0, X_1, \ldots, X_k$ i.i.d. $N(0,1)$ and, for $c \in \mathbb{R}$, define $X_0(c)$ to be X_0 truncated below at c. Now define (analogous to (4.5)) $Y_j = \delta_j X_0(c) + \sqrt{1 - \delta_j^2} X_i$, $i = 1, 2, \ldots, k$. This will lead to densities of the more general form (4.3) (i.e., there will exist a matrix B such that $B\underline{Y}$ has density (4.3)). See Arnold and Beaver (2002a) for details on this construction.

Perhaps the best way to identify the nature of the marginal distributions of the general hidden truncation model (4.3) is to consider its moment generating function. Arnold and Beaver (2000a) provide the following expression for this moment generating function:

$$M(\underline{t}) = \frac{e^{\underline{t}'\underline{t}/2}\Phi\left(\frac{\lambda_0+\underline{\lambda}_1'\underline{t}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)} \;. \tag{4.6}$$

To get the moment generating function of a k_1 dimensional marginal of the distribution (4.3), i.e. of $\underline{\dot{W}}$ where $\underline{W} = (\underline{\dot{W}}, \underline{\ddot{W}})$ has been partitioned into subvectors of dimension k_1 and $k - k_1$, we need only to similarly partition $\underline{t} = (\underline{\dot{t}}, \underline{\ddot{t}})$ and set $\underline{\ddot{t}} = \underline{0}$ in (4.6). It is then obvious that the marginal m.g.f. of $\underline{\dot{W}}$ is of the same form as (4.5) (though now of dimension k_1).

Specifically $\underline{\dot{W}}$ has density given by

$$f(\underline{\dot{w}}) = \frac{\left[\prod_{i=1}^{k_1}\varphi(w_i)\right]\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\ddot{\lambda}}_1'\underline{\ddot{\lambda}}_1}} + \frac{\underline{\dot{\lambda}}_1'\underline{\dot{w}}}{\sqrt{1+\underline{\ddot{\lambda}}_1'\underline{\ddot{\lambda}}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)} \tag{4.7}$$

and analogously $\underline{\ddot{W}}$ has density of the form:

$$f(\underline{\ddot{w}}) = \frac{\left[\prod_{i=k_1+1}^{k}\varphi(w_i)\right]\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}} + \frac{\underline{\ddot{\lambda}}_1'\underline{\ddot{w}}}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)} \tag{4.8}$$

It follows that the conditional density of $\underline{\dot{W}}$ given $\underline{\ddot{W}} = \underline{\ddot{w}}$ is (taking the ratio of (4.3) and (4.8))

$$f(\underline{\dot{w}}|\underline{\ddot{w}}) = \frac{\left[\prod_{i=1}^{k_1}\varphi(w_i)\right]\Phi(\lambda_0 + \underline{\ddot{\lambda}}_1'\underline{\ddot{w}} + \underline{\dot{\lambda}}_1'\underline{\dot{w}})}{\Phi\left(\frac{\lambda_0+\underline{\ddot{\lambda}}_1'\underline{\ddot{w}}}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}}\right)} \;. \tag{4.9}$$

Thus all conditionals as well as all marginals of the density (4.3) are of the same type.

If we introduce location and scale parameters in our model (4.3), the property of having marginals and conditionals of the same type continues to hold. If we define

$$\underline{X} = \underline{\mu} + \Sigma^{1/2}\underline{W} \tag{4.10}$$

where $\underline{\mu} \in \mathbb{R}^k$ and $\Sigma^{1/2}$ is positive definite and $\underline{W}$ has density (4.3), then the joint density of $\underline{X}$ (our full family of hidden truncation skewed multivariate normal models) is:

$$f_{\underline{X}}(\underline{x}) \propto \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right) \times \Phi\left(\frac{\delta_0 + \underline{\delta}_1'\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}}\right) \tag{4.11}$$

with $\delta_0 = \lambda_0/\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}$ and $\underline{\delta}_1 = \underline{\lambda}_1/\sqrt{1+\underline{\lambda}'\underline{\lambda}_1}$. Now partition $\underline{x}' = (\underline{\dot{x}}', \underline{\ddot{x}}')$ with a corresponding partition of Σ and $\underline{\lambda}_1$. Using the moment-generating function, the marginals are of the form

$$f_{\underline{\ddot{X}}}(\underline{\ddot{x}}) \propto \exp\left(-\frac{1}{2}(\underline{\ddot{x}}-\underline{\ddot{\mu}})'\Sigma_{22}^{-1}(\underline{\ddot{x}}-\underline{\ddot{\mu}})\right) \times \Phi\left(\frac{\delta_0 + \underline{\ddot{\delta}}_1'\Sigma_{22}^{-\frac{1}{2}}(\underline{\ddot{x}}-\underline{\ddot{\mu}})}{\sqrt{1-\underline{\ddot{\delta}}_1'\underline{\ddot{\delta}}_1}}\right). \tag{4.12}$$

The conditional density of $\underline{\dot{X}}|\underline{\ddot{X}}$ is

$$\begin{aligned} f_{\underline{\dot{X}}|\underline{\ddot{X}}}(\underline{\dot{x}}|\underline{\ddot{x}}) \quad \propto \quad & \exp\left(-\frac{1}{2}(\underline{\dot{x}}-\underline{\dot{\mu}}(\underline{\ddot{x}}))'(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(\underline{\dot{x}}-\underline{\dot{\mu}}(\underline{\ddot{x}}))\right) \\ & \times \Phi\left(\frac{\delta_0 + \underline{\delta}_1'\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}}\right), \end{aligned} \tag{4.13}$$

where

$$\dot{\mu}(\underline{\ddot{x}}) = \underline{\dot{\mu}} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{\ddot{x}} - \underline{\ddot{\mu}}).$$

Clearly both densities (4.12) and (4.13) are of the same type as (4.11).

Moments of the density (4.3) are readily obtained from the moment generating function (4.6). One finds

$$E(W_i) = \frac{\lambda_{1i}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}} \frac{\varphi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}, \tag{4.14}$$

and

$$cov(W_i, W_j) = \delta_{ij} - \frac{\lambda_{1i}\lambda_{1j}}{(1+\underline{\lambda}_1'\underline{\lambda}_1)}\left[\frac{\lambda_0}{\sqrt{1\underline{\lambda}_1'\underline{\lambda}_1}} g(\lambda_0, \underline{\lambda}_1) + g^2(\lambda_0, \underline{\lambda}_1)\right], \tag{4.15}$$

where δ_{ij} is the Kronecker delta symbol, and where

$$g(\lambda_0, \underline{\lambda}_1) = \frac{\varphi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}. \tag{4.16}$$

Considerable simplification occurs when $\lambda_0 = 0$, in which case the expression in (4.16) simplifies to $\sqrt{\frac{2}{\pi}}$.

5 Non-normal skewed multivariate models

The basic random variables $W_1, W_2, \ldots, W_k, U$ used in the development of the hidden truncation model in Section 4 were assumed to be i.i.d. standard normal random variables. If they are permitted to have different distributions we will be led to new types of skewed multivariate distributions.

Let us suppose that $W_1, W_2, \ldots, W_k$ and U are independent random variables with densities (distributions) given by $\psi_1(\Psi_1), \psi_2(\Psi_2), \ldots, \psi_k(\Psi_k)$, $\psi_0(\Psi_0)$ respectively. We then consider the conditional distribution of $\underline{W}$ given that $\lambda_0 + \underline{\lambda}_1'\underline{W} > U$, as before, where $\lambda_0 \in \mathbb{R}$ and $\underline{\lambda}_1 \in \mathbb{R}^k$. The conditional density of $(\underline{W}, U)$ given $A = \{\lambda_0 + \lambda_1 \underline{W} > U\}$ is of the form

$$f_{\underline{W},U|A}(\underline{w}, u) \propto \left[\prod_{i=1}^{k} \psi_i(w_i)\right] \psi_0(u) I(\lambda_0 + \underline{\lambda}_1'\underline{w} > u). \tag{5.1}$$

Integrating with respect to u yields

$$f_{\underline{W}|A}(\underline{w}) = \frac{\left[\prod_{i=1}^{k} \psi_i(w_i)\right] \Psi_0(\lambda_0 + \underline{\lambda}_1'\underline{w})}{P(A)} . \tag{5.2}$$

The quantity $P(A)$ which appears in the denominator of (5.2) may prove to be difficult to evaluate. In certain cases it will be easily accomplished. The simplest case is that in which $\lambda_0 = 0$ and all of the densities $\psi_1, \psi_2, \ldots, \psi_k, \psi_0$ are symmetric (about 0). In that case $P(A) = 1/2$. This yields the following analog of (4.4)

$$f_{\underline{W}|A}(\underline{w}) = 2\left[\prod_{i=1}^{k} \psi_i(w_i)\right] \Psi_0(\underline{\lambda}_1'\underline{w}). \tag{5.3}$$

The models (5.2) and (5.3) are readily extended by the usual linear transformation of the form

$$\underline{X} = \underline{\mu} + \Sigma^{1/2}\underline{W} \tag{5.4}$$

When $\lambda_0 \neq 0$ in (5.2), it is usually not possible to evaluate $P(A)$. If the W_i's and U are all stable variables of the same type, then $\underline{\lambda}_1'W - U$ will also be stable and, in principle, one can then evaluate $P(A)$. Of course this can be done if the variables are normally distributed. Arnold and Beaver (2000b) consider the other tractable case in which the W_i's and U are Cauchy variables. The resulting skew-Cauchy distribution has all marginals and all conditionals again of the skew-Cauchy form (closely paralleling the skew-normal paradigm).

An even more general skewed multivariate model can be obtained by not assuming that the W_i's are independent, but instead that their joint density is given by $\psi(\underline{w})$, say. In this case (5.2) is replaced by:

$$f_{\underline{W}|A}(\underline{w}) = \frac{\psi(\underline{w})\Psi_0(\lambda_0 + \underline{\lambda}_1'\underline{w})}{P(A)} \tag{5.5}$$

where, as before, $A = \{\lambda_0 + \underline{\lambda}_1'\underline{W} > U\}$. In order to justify calling these "skewed" models, it is usually reasonable to assume that the marginal densitites of $\underline{W}$ are symmetric, though the formulas remain valid without this assumption. Perhaps in asymmetric cases, we should revert to calling them hidden truncation models.

It will be recalled that there was another scenario not involving hidden truncation, for developing the Azzalini skew-normal model (2.1) and its k-variate analog (4.4). A non-normal version of this model may be described as follows.

Begin with independent random variables $Y_0, Y_1, Y_2, \ldots, Y_k$ with corresponding densities $\psi_0, \psi_1, \psi_2, \ldots, \psi_k$. Define $\underline{Z}$ by

$$Z_j = |Y_0| + \tau_j Y_j, \quad j = 1, 2, \ldots, k \tag{5.6}$$

(this is essentially the same as (4.5) used in the normal case by Azzalini and Dalla Valle). The joint density of $\underline{Z}$ is then given by

$$f_{\underline{Z}}(\underline{z}; \underline{\tau}) = \int_0^\infty \left[\prod_{i=1}^k \frac{1}{\tau_i}\psi_i\left(\frac{z_i - u}{\tau_i}\right)\right] [\psi_0(u) + \psi_0(-u)]\, du. \tag{5.7}$$

It may be difficult to perform the integration in (5.7). Tractable examples are ones in which all of the ψ_i's ($i = 0, 1, \ldots, k$) are standard normal (or Cauchy, Laplace or logistic). The normal case, as remarked earlier, leads to (4.4) again. The other cases lead to essentially new models. The model (5.6) is referred to by Arnold and Beaver (2000c) as a distribution skewed by an additive component (as distinguished from distributions skewed by conditioning or hidden truncation). Such additive component models have the desirable property of having k_1-dimensional marginals again of the same (lower dimensional) type. The representation (5.6) also permits ready computation of means, variances and covariances together with the joint moment generating function (or characteristic function).

Remark 5.1. If we replace $|Y_0|$ in (5.6) by $Y_0(c)$ (Y_0 truncated below at c), we can get new non-normal extensions of the model (4.3). See for example Arnold and Beaver (2002a).

It bears remarking at this point also that in these general settings it does make a difference whether we consider models based on a constraint of the form $\{\lambda_0 + \underline{\lambda}_1' \underline{W} > U\}$ or one of the form $\{\lambda_0 + \underline{\lambda}_1' \underline{W} < U\}$. In the following brief section we describe certain hidden truncation survival models where this distinction is important. It is not, of course, in the context of symmetric models such as the skew-normal density (4.4), where either use of "$> U$" or "$< U$" will lead to the same kind of models.

6 Hidden truncation survival models

Suppose that $(\underline{X}, Y)$ is a $k+1$ dimensional random vector with non-negative coordinates. Consider the conditional distribution of $\underline{X}$ given $\underline{\lambda}'\underline{X} < Y$. We visualize the X_i's as survival times of k components in a system. The variable Y is a related unobservable variable, perhaps related to operating conditions. Assume that the X_i's and Y have corresponding densities $\psi_1, \psi_2, \ldots, \psi_k, \psi_0$ and survival functions $\bar{\Psi}_1, \bar{\Psi}_2, \ldots, \bar{\Psi}_k, \bar{\Psi}_0$. Under such circumstances, the conditional density of $\underline{X}$ given $\{\underline{\lambda}'\underline{X} < Y\}$ takes the form:

$$f_{\underline{X}}(\underline{x}; \underline{\lambda}) \propto \left[\prod_{i=1}^{k} \psi_i(x_i)\right] \bar{\Psi}_0(\underline{\lambda}'\underline{x}). \tag{6.1}$$

The normalizing constant needed to guarantee that (6.1) integrates to 1 is, here too, likely to be difficult to evaluate. There is one case in which it is not difficult; the case in which Y has an exponential distribution. The hidden truncation model (6.1) has some remarkable features in that case due to the nature of the exponential density.

Suppose that Y in (6.1) has an exponential (δ_0) distribution. Then our hidden truncation density becomes

$$f_{\underline{X}}(\underline{x};\underline{\lambda}) \propto \left[\prod_{i=1}^{k} \psi_i(x_i)\right] e^{-\delta_0 \underline{\lambda}'\underline{x}} I(\{\underline{x} > \underline{0}\}), \tag{6.2}$$

which factors, indicating independent marginals. If we denote the Laplace transform of density ψ_i by M_i, i.e.

$$M_i(t) = \int_0^\infty e^{-tx}\psi_i(x)dx, \tag{6.3}$$

it follows that we can write

$$f_{\underline{X}}(\underline{x};\underline{\lambda}) = \prod_{i=1}^{k} \left[\frac{\psi_i(x_i)e^{-\delta_0\lambda_i}}{M_i(\delta_0\lambda_i)}\right] I(\{\underline{x} > \underline{0}\}). \tag{6.4}$$

Arnold and Beaver (2002b) verify that a hidden truncation model of the form (6.1) will have independent marginals iff $\Psi_0(y) = e^{-\delta y}$, for some $\delta > 0$, i.e. only when Y has an exponential distribution.

One could of course consider a more general version of (6.1) in which we condition on the event $\{\lambda_0 + \underline{\lambda}'\underline{X} < Y\}$ where $\lambda_0 > 0$ (instead of $\lambda_0 = 0$ in (6.1)). The resulting density takes the form

$$f(\underline{x};\underline{\lambda}) \propto \tilde{\psi}(\underline{x})\bar{\Psi}_0(\underline{\lambda}'\underline{x} + \lambda_0) \;\; I(\{\underline{x} > \underline{0}\}). \tag{6.5}$$

If Y has an exponential distribution then $\bar{\Psi}_0(\underline{\lambda}'\underline{x} + \lambda_0) \propto \bar{\Psi}_0(\underline{\lambda}'\underline{x})$ and the model with $\lambda_0 > 0$ will be identical to the model with $\lambda_0 = 0$. Thus we would be unable to determine from the data whether the threshhold λ_0 was positive or zero.

7 Distributional aspects of the k-variate normal hidden truncation model

Suppose that $\underline{X} = \underline{\mu} + \Sigma^{1/2}\underline{W}$, where $\underline{W}$ has the k-variate density (4.3). We have already noted that $\underline{X}$ will have all marginals and all conditional densities in the location scale extension of the family (4.3). In the absence of hidden truncation, linear and quadratic forms in the variables are well behaved. What happens in the skewed case in which $\underline{\lambda}_0$ and $\underline{\lambda}_1'$ are not all zero.

For linear forms there is no difficulty of course. If $\underline{Y} = A\underline{X} + \underline{b}$ where A is a $k \times k$ matrix, then

$$\underline{Y} = A\underline{\mu} + \underline{b} + A\Sigma^{1/2}\underline{W}$$

which clearly is again an affine transformation of $\underline{W}$ with density (4.3). If $\underline{Y}$ has dimension less than k (i.e., if A is a $k_1 \times k$ matrix with $k_1 < k$) then $\underline{Y}$ is readily seen to still have a distribution associated with an affine transformation of a k_1 dimensional variable with density of the form (4.3). So linear forms in the X_i's continue to behave well, i.e., to be again skewed normal variables.

Turning to quadratic forms, there is some evidence that quadratic forms in skewed normal variables will have scaled χ^2 distributions just as they do for classical multivariate normal variables. The first hint of this in the literature is to be found in the paper by Roberts (1966). He considered the square of the minimum of two correlated normal random variables with zero means and unit variances. He showed that this variable has a χ_1^2 distribution. Recalling the Loperfido (2002) discussion of the skew normal nature of maxima (or minima) of correlated normal variables, it is evident that Roberts was considering one case in which the square of a skew-normal variable (with density (2.1)) has a χ_1^2 distribution. Azzalini and Capitanio (1999) showed that a χ^2 distribution describes the behavior of a broad class of quadratic forms in variables with density (4.4) (when $\lambda_0 = 0$). The following computation confirms that this will not occur in the general model (4.3). We may readily evaluate the moment generating function of $\sum_{i=1}^k W_i^2$ when $\underline{W}$ has density (4.3) as follows:

$$E\left(e^{t\Sigma_{i=1}^k W_i^2}\right) = \int \cdots \int_{\mathbb{R}^k} e^{t\underline{w}'\underline{w}} \left[\prod_{i=1}^k \varphi(w_i)\right] \frac{\Phi\left(\frac{\lambda_0 + \underline{\lambda}_1'\underline{w}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)} d\underline{w}$$

$$= \quad (1-2t)^{-k/2}\int\cdots\int_{\mathbb{R}^k}\left[\prod_{i=1}^{k}\varphi(v_i)\right]\frac{\Phi\left(\frac{\lambda_0+(1-2t)^{-1/2}\underline{\lambda}'\underline{v}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)}d\underline{v}$$

$$= \quad (1-2t)^{-k/2}\frac{\Phi\left(\frac{\lambda_0}{\sqrt{1+\frac{\underline{\lambda}_1'\underline{\lambda}_1}{1-2t}}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)},$$

where, in the second equality, $\underline{v} = (1-2t)^{1/2}\underline{w}$ This reduces to the χ^2_k moment generating function ($(1-2t)^{-k/2}$) only when $\lambda_0 = 0$. In general, then, $\Sigma_{i=1}^k W_i^2$ will not have a χ^2_k distribution. If $\lambda_0 = 0$, then, as observed by Azzalini and Capitanio (1999), $\Sigma_{i=1}^k W_i^2$ will have a χ^2_k distribution for any $\underline{\lambda}_1$. Azzalini and Capitanio (1999) also point out cases in which quadratic forms in skew-normal variables can (with density (4.4)) have independent χ^2 distributions, paralleling Cochran's theorem in the non-skewed case. See Loperfido (2001) and Genton et al. (2001) for further discussion of quadratic forms in skew normal variables.

8 The conditional specification approach

The k-variate skewed normal density (4.3) was seen to have conditional densities which are all affine transformations of the general form (4.3). In particular the conditional density of W_i given the other variables $\{W_j : j \neq i\}$ is of the form (2.3), for every i. Arnold et al. (2002) discussed the nature of multivariate densities with such skew-normal conditional structure. To describe their results, we first introduce some convenient notational conventions. For a random vector $\underline{W}$ of dimension k, we define $(k-1)$-dimensional subvectors $\underline{W}_{(1)}, \underline{W}_{(2)}, \ldots, \underline{W}_{(k)}$ such that, for each i, $\underline{W}_{(i)}$ is the vector $\underline{W}$ with the ith coordinate W_i deleted. Analogously for any real vector $\underline{x}$ we define $\underline{x}_{(i)}$. We will say that a random vector $\underline{W}$ has skewed normal conditionals (of the type (2.3)) if for every $\underline{w}_{(i)} \in \mathbb{R}^{k-1}$ and for every i, the conditional random variables

$$W_i|\underline{W}_{(i)} = \underline{w}_{(i)} \sim LSN(\lambda_0(\underline{w}_{(i)}), \lambda_1(\underline{w}_{(i)})), \tag{8.1}$$

where by a $LSN(\lambda_0, \lambda_1)$ distribution we mean one with density (2.3) and $\sim$ denotes "is distributed as". Arnold et al. (2002) speak of (2.3) as being a

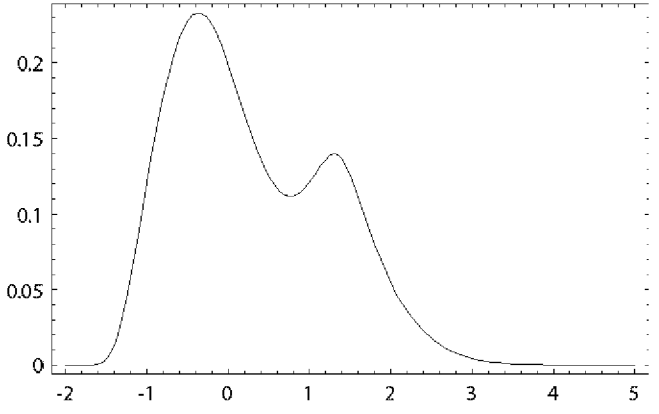

Figure 4: Example of a pdf of the form (8.4).

linearly skewed normal density as distinguished from (2.1) which is simply called a skew normal density and the following density called a quadratically skewed normal density:

$$f(z; \lambda_0, \lambda_1, \lambda_2) \propto \varphi(z)\Phi(\lambda_0 + \lambda_1 z + \lambda_2 z^2). \tag{8.2}$$

A k-dimensional density with linearly skewed normal conditionals (i.e. satisfying (8.1)) is given by

$$f(w_1, w_2, \ldots, w_k) \propto \left[\prod_{i=1}^{k} \varphi(w_i)\right] \Phi\left(\sum_{\underline{s}\in S_k} \lambda_{\underline{s}} \prod_{i=1}^{k} w_i^{s_i}\right) \tag{8.3}$$

where $S_k = \{0,1\}^k$ is the set of all vectors of 0's and 1's of dimension k. Location and scale parameters can be introduced in (8.3) to extend the model retaining linearly skewed normal conditionals.

The pdf

$$f(w_1) \propto \phi(w_1)\phi(w_2)\Phi(w_1(w_1-1)(w_1+1)), \tag{8.4}$$

is shown in Figure 4.

The pdfs

$$f(w_1, w_2) \propto \phi(w_1)\phi(w_2)\Phi(w_1 + 2w_2 + 6w_1 w_2) \tag{8.5}$$

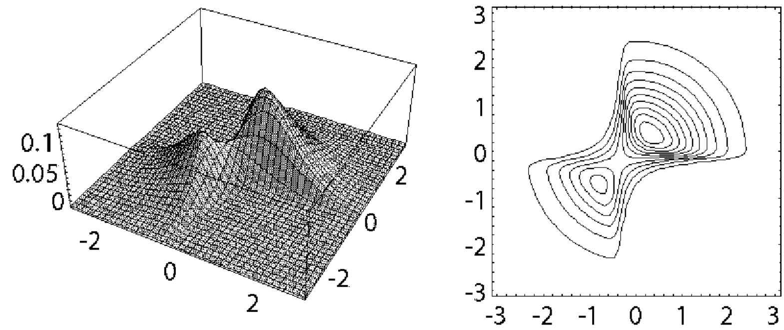

Figure 5: Example of a pdf of the form (8.5).

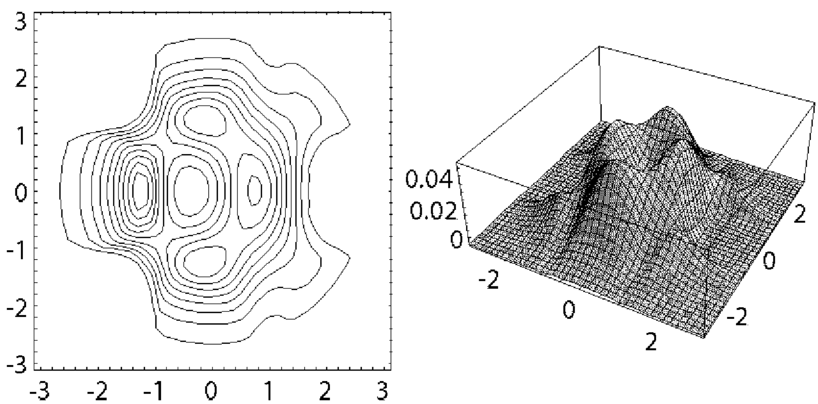

Figure 6: Example of a pdf and contours of the form (8.6).

and

$$f(w_1, w_2) \propto \phi(w_1)\phi(w_2)\Phi((w_1 - 1)^2(w_1 + 1)(w_2 - 1)(w_2 + 1)) \qquad (8.6)$$

and their contours are displayed in Figures 5 and 6.

The reader is referred to Arnold et al. (2002) for discussion of more general models with linearly and/or quadratically skewed normal conditionals. Additionally, that paper includes discussion of non-normal variants of these distributions.

9 Multiple constraint models

The possibility of "multiple" hidden truncations cannot be ignored. We can easily imagine situations in which observations are made only if all of several covariables attain certain critical levels. The normal multiple constraint model is most easily visualized and most tractable but, without going into details, we can just as well consider more general models involving non-normal not necessarily identically distributed basic distributions (as in Section 5).

Thus we begin with $k+\ell$ independent random variables $W_1, W_2, \ldots, W_k$, $U_1, U_2, \ldots, U_\ell$. Then we consider the conditional distribution of $\underline{W}$ given that for $j = 1, 2, \ldots, \ell$, $\lambda_0^{(j)} + \underline{\lambda}_1^{(j)} \underline{W} > U_j$. Denote the density (distribution) functions of the W_i's by $\psi_i(\Psi_i)$ respectively and the density (distribution) functions of the U_j's by $\phi_j(\Phi_j)$ respectively. Next define events $A_j, j = 1, 2, \ldots, \ell$ by

$$A_j = \{\lambda_0^{(j)} + \underline{\lambda}_1^{(j)} \underline{W} > U_j\}. \tag{9.1}$$

We next consider the conditional density of $(\underline{W}, \underline{U})$ given $A^* = \cap_{j=1}^{\ell} A_j$ and integrate out unwanted variables to obtain the desired conditional density of $\underline{W}$ given A^* of the form

$$f_{\underline{W}|A^*}(\underline{w}) = \frac{\left[\prod_{i=1}^{k} \psi_i(w_i)\right] \prod_{j=1}^{\ell} \Phi_j(\lambda_0^{(j)} + \underline{\lambda}_1^{(j)\prime} \underline{w})}{P\left(A^*\right)} \tag{9.2}$$

As in the single constraint case, it is the denominator in (9.2) which generally will be difficult to deal with. The quantity $P(A^*)$ will be computable in the following special case. Suppose that all the ψ_i's and all the ψ_j's are standard normal densities (denoted as usual by φ with corresponding distribution function Φ). In addition assume that the "skewness" vectors $\underline{\lambda}_1^{(j)}$, $j = 1, 2, \ldots, \ell$ are mutually orthogonal. With these assumptions, the A_j's are independent events and our density (9.2) takes the specific form

$$f_{\underline{W}|A^*}(\underline{w}) = \frac{\left[\prod_{i=1}^{k} \varphi(w_i)\right] \prod_{j=1}^{\ell} \Phi\left(\lambda_0^{(i)} + \underline{\lambda}_1^{(j)\prime} \underline{w}\right)}{\prod_{j=1}^{\ell} \Phi\left(\frac{\lambda_0^{(j)}}{\sqrt{1+\underline{\lambda}_1^{(j)\prime}\underline{\lambda}_1^{(j)}}}\right)}. \tag{9.3}$$

If in addition we have $\lambda_0^{(j)} = 0$, $j = 1, 2, \ldots, \ell$ then this simplifies to yield

$$f_{\underline{W}|A^*}(\underline{w}) = 2^\ell \left[\prod_{i=1}^{k} \varphi(w_i) \right] \prod_{j=1}^{\ell} \Phi \left(\underline{\lambda}_1^{(j)\prime} \underline{w} \right), \tag{9.4}$$

recognizable as a natural extension of the density (4.4).

10 Skew-elliptical distributions

Azzalini and Capitanio (1999) suggested the possibility of extending the skew-normal model by replacing the multivariate normal component of the model by a more general elliptically contoured distribution. Branco and Dey (2001) also provide a detailed discussion of such skew-elliptical models. The present section draws on both sources but, for further detail, the original papers should be consulted. We will say that a k-dimensional random vector $\underline{W}$ has an elliptically contoured distribution if it admits the representation

$$\underline{W} = \underline{\mu} + \Sigma^{1/2} \underline{U} \tag{10.1}$$

where $\underline{U}$ is a spherically symmetric random vector. Such a random vector $\underline{U}$ may be represented as

$$\underline{U} = Z(R_1, \ldots, R_k), \tag{10.2}$$

where Z and $\underline{R}$ are independent random variables with $Z > 0$ with probability 1 and with distribution function F_Z and $\underline{R}$ is uniformly distributed over the unit k-sphere. The choice of distribution function F_Z will determine the distribution of $\underline{W}$. Generally speaking the coordinates of $\underline{W}$ will be dependent as will the coordinates of $\underline{U}$. As is well known, the coordinates of $\underline{U}$ will be independent only if Z has a particular chi-distribution (i.e. Z is a constant multiple of the square root of a χ^2 distribution with k degrees of freedom). In that case $\underline{U}$ will be $N^k(\underline{0}, \sigma^2 I)$ and $\underline{W}$ will have also independent coordinates if $\Sigma^{1/2}$ in (10.1) is orthogonal.

To construct a k-dimensional skew-elliptical distribution we can begin with a $(k+1)$-dimensional elliptically contoured random vector of the form (10.1), say $(W_0, W_1, W_2, \ldots, W_k)$. Then the conditional distribution of $(W_1, \ldots, W_k)$ given that $W_0 > c$ where c is some prescribed threshhold (Branco and Dey (2001) fix $c = 0$), can be reasonably called a k-dimensional skew-elliptical density. In the special case in which Z (in (10.2)) has an appropriate chi-distribution, this skewed model will reduce to the model obtained by applying affine transformations to the density (4.4). If we chose Z such that $\underline{U}$ in (10.2) has Cauchy marginals, the resulting skewed-Cauchy distribution will be distinct from the one introduced by Arnold and Beaver (2000b) and discussed in Section 5.

Many spherically symmetric random vectors (but not all) can be represented as scale mixtures of $N(0, \sigma^2 I)$ random vectors. For such distributions and their skewed counterparts expressions for moments and moment generating functions can be obtained by suitable integration of the corresponding expressions for their normal and skew-normal counterparts (for details see Branco and Dey (2001)).

It is interesting to observe that the skew-elliptical distributions derived by Branco and Dey (2001) using this paradigm (with $c = 0$) turn out to be special cases of the quite general mechanism for "skewing" an arbitrary joint density exhibited in equation (5.5). In it, recall that $\psi(\underline{w})$ was a general k-dimensional density and Ψ_0 was an arbitrary one dimensional distribution function. The resulting skewed density took the form

$$f(\underline{w}) \propto \psi(\underline{w})\Psi_0(\lambda_0 + \underline{\lambda}_1'\underline{w}). \tag{10.3}$$

If we begin with $\psi(\underline{w})$ an elliptically contoured density then (10.3) includes the Branco-Dey models as special cases. Azzalini and Capitanio (1999) describe a different kind of skew-elliptical distribution. It also can be viewed as a special case of (10.3). See Branco and Dey (2001) and Azzalini and Capitanio (1999) for more detail on the missing normalization constant in (10.3) appropriate for their models.

11 Inference

The skewed normal models (2.1), (2.3), (4.2), (4.4) do not constitute exponential families of densities. As a consequence, essentially no reduction in complexity of the data can be obtained by invoking sufficiency arguments. Maximum likelihood estimation will necessarily be performed numerically and, unfortunately, the likelihood functions associated with these models often do not have easily identified modes. With those warnings in mind, let us see what can be (or has so far been) done with regard to estimation, testing and fitting such models.

We remarked earlier that Roberts (1966) in his "model for twin data" was led to consideration of a skew normal model (2.1) with

$$\lambda = -\sqrt{(1-\rho)/(1+\rho)}.$$

In this setting he proposed a method of moments estimate of ρ. In our notation, he assumed that a sample $Z_1, Z_2, \ldots, Z_n$ was available from the density (2.1). He then equated $(\frac{1}{n}\sum_{i=1}^n Z_i)^2$ to its expectation and solved for λ. From this, he then estimated ρ, the parameter of interest in his paper. These estimates of λ and ρ are consistent and asymptotically normal.

For a sample $X_1, X_2, \ldots, X_n$ from the univariate skewed normal (2.3) we can estimate the 4 parameters μ, σ, λ_0 and λ_1 using the method of moments. Note however that Arnold et al. (1993) report that in some simulation studies, even under the simplifying assumption that $\lambda_0 = 0$, the moment equations will for some samples not have a solution.

It turns out that even for the basic Azzalini model with 3 free parameters μ, σ and λ_1 (with λ_0 set equal to 0), the likelihood surface is frequently anomalous. The likelihood equations for this model are easily written down, viz:

$$\hat{\mu}_1 = \bar{x} - \left(\frac{\hat{\lambda}\hat{\sigma}_1}{n}\right)\sum_{i=1}^n W(x_i), \tag{11.1}$$

$$\hat{\sigma}_1^2 = \frac{1}{n}\left[\sum_{i=1}^n (x_i - \hat{\mu}_1)^2\right], \tag{11.2}$$

and

$$0 = \sum_{i=1}^n \left(\frac{x_i - \hat{\mu}_1}{\hat{\sigma}_1}\right) W(x_i), \tag{11.3}$$

in which

$$W(x_i) = \frac{\phi\left(\hat{\lambda}\frac{(x_i - \hat{\mu}_1)}{\hat{\sigma}_1}\right)}{\Phi\left(\hat{\lambda}\frac{(x_i - \hat{\mu}_1)}{\hat{\sigma}_1}\right)}. \tag{11.4}$$

Arnold et al. (1993) following the suggestion of Azzalini (1985) are able to solve these equations in many cases by using the profile likelihood of λ. Azzalini provides the Fisher information matrix for this 3 parameter model and notes that it becomes singular as $\lambda \to 0$. As a consequence he recommends a specific reparameterization prior to maximizing the likelihood.

He recommends parameterizing in terms of mean, standard deviation and skewness. More detailed discussion of these issues and certain remaining anomalies associated with the likelihood surface (even after reparameterizing) may be found in Section 5 of Azzalini and Capitanio (1999). See also Pewsey (2000).

For the 4-parameter model (2.3) a direct search of the likelihood surface may have to be made. In higher dimensions such numerical methods will be unavoidable. For the k-dimensional model (4.11) with $\delta_0 = 0$ (equivalently $\lambda_0 = 0$ in (4.3)), Azzalini and Capitanio (1999) recommend picking initial values using method of moments estimates and then use these in a variation of the Newton-Raphson approach to finding the maximum of the surface. In some cases they need to use an EM algorithm, but they do report final convergence in all examples tried.

Arnold and Beaver (2000a) considered the full multivariate model (4.11) but with $k = 2$, so that the model has 8 parameters. They fitted this model to a particular height-weight data set. In addition, they fitted a 7 parameter model (obtained from (4.11) by setting $\lambda_0 = 0$) and a 5 parameter classical bivariate normal model. Estimates for the 7 and 8 parameter models were successfully obtained using a genetic algorithm and simulated annealing. By considering likelihood ratios associated with these 3 nested models it was determined that for the data set in question the 7 parameter skewed normal model fit the data significantly better than a classical normal model. The improved fit obtained by introducing the 8th parameter λ_0 was only marginally significant. The same data set was used for illustrative purposes in Arnold and Beaver (2000b). In that paper a bivariate skew-Cauchy model was fitted by maximum likelihood again using a genetic algorithm and simulated annealing.

What about goodness of fit tests for univariate or k-variate skewed distributions? Only one paper, that we are aware of, specifically addreses this problem. Gupta and Chen (2001) present tables of the distribution function for the skew-normal density (2.1) with only one parameter λ. With these tables they describe implementation of the Pearson χ^2 test and the Kolmogorov-Smirnov test of goodness of fit to the skew-normal density (2.1) with a given value of λ. Presumably one could use this approach in combination with an estimated value of λ to construct an overall goodness of fit test to the model (2.1), or one could estimate μ, σ and λ_1 in the model (2.4) (with $\lambda_0 = 0$) using the data and then use Gupta and Chen's tables to approximately test goodness of fit of this 3 parameter model. So only the surface of the goodness of fit problem has been scratched; and only in one dimension.

12 A continuing saga

Data sets involving hidden truncation, or if you will, data modelled by variants of the skew-normal distribution, are surely endemic in the world. Any statistical analysis involving linear normal models (analysis of variance, regression, principal components etc. etc.) can be considered to be potentially a setting in which hidden truncation might have occurred and some effort to investigate the possibility can be clearly justified. Azzalini and Capitanio (1999) provide discussion of some regression and graphical models involving skew-normal variates. A subsequent paper by Capitanio et al. (2002) explores the graphical model theme in more depth.

Gupta and Brown (2001) explored the application of the skew-normal as a unimodal distribution from the point of view of reliability analysis. They show that the skew-normal density is log-concave, and that a skew normal variable, $SN(\lambda)$, has an increasing failure rate, and therefore a decreasing mean residual life.

Normal distributions are frequently used as prior densities in Bayesian analyses. A possible role for skewed-normal priors clearly exists here (see the earlier cited paper by O'Hagan and Leonard (1976) and also Mukhopadhyay and Vidakovic (1995) for some related discussion). Additionally more progress can be expected on the goodness of fit problems discussed in Section 11 and on multiple constraint models (Section 9). One multiple constraint scenario that quite frequently might be encountered is one in which a single constraint (or several) of the form $\{\lambda_0 + \underline{\lambda}_1 \underline{W} > U\}$ is replaced by a two sided constraint such as $\{\gamma_1 < \underline{\lambda}_1 W - U < \gamma_2\}$. Indeed, Arnold et al. (1993), in the univariate case, did consider such two sided constraints.

References

Arnold, B. C. and Beaver, R. J. (2000a). Hidden truncation models. *Sankhya Series A*, 62:23–35.

Arnold, B. C. and Beaver, R. J. (2000b). The skew-Cauchy distribution. *Statistics and Probability Letters*, 49:285–290.

Arnold, B. C. and Beaver, R. J. (2000c). Some skewed multivariate distributions. *American Journal of Mathematical and Management Sciences*, 20:27–38.

Arnold, B. C. and Beaver, R. J. (2002a). An alternative construction of skewed multivariate distributions. Technical Report 270, Department of Statistics, University of California, Riverside.

Arnold, B. C. and Beaver, R. J. (2002b). Multivariate survival models incorporating hidden truncation. Technical Report 269, Department of Statistics, University of California, Riverside.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58:471–188.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2002). Conditionally specified multivariate skewed distributions. *Sankhya Series A (to appear)*.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–17.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199–208.

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

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

Balakrishnan, N. and Ambagaspitiya, R. S. (1994). On skew-Laplace distributions. Technical report, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *The Annals of Mathematical Statitics*, 21:272–279.

Branco, C. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79:99–113.

Cain, M. (1994). The moment-generating function of the minimum of bivariate normal random variables. *The American Statistician*, 48:124–125.

Capitanio, A., Azzalini, A., and Stanghellini, E. (2002). Graphical models for skew normal variates. *Scandinavian Journal of Statistics*, 29 (to appear).

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *Journal of the Italian Statistical Society*, 7:1–13.

Genton, M. G., Li, H., and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics and Probability Letters*, 51:319–325.

Gupta, A. K. and Chen, T. (2001). Goodness-of-fit tests for the skew-normal distribution. *Communications in Statistics, Part C: Computing and Simulation*, 30:907–930.

Gupta, R. C. and Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Communication in Statistics: Theory and Methods*, 30:2427–2445.

Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scandinavian Journal of Statistics*, 13:271–275.

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statistics and Probability Letters*, 54:381–387.

Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics and Probability Letters*, 56:13–22.

Mukhopadhyay, S. and Vidakovic, B. (1995). Efficiency of linear Bayes rules for a normal mean: skewed prior class. *Journal of the Royal Statistical Society D*, 44:469–471.

Nelson, L. S. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6:469–470.

O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63:201–203.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27:859–870.

Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61:1184–1190.

Weinstein, M. A. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6:104–105.

DISCUSSION

A. Azzalini
Università di Padova, Italy

Professors Arnold and Beaver must be congratulated on their presentation of the literature on skewed distribution arising from hidden truncation and other related schemes. The coverage of material and the depth of discussion make this a most interesting article, which surely will contribute to draw the attention of more people to this fast developing area.

My only addition to the discussion is to draw attention to connections with two other streams of literature. One is the theme of biased sampling, which has been presented by Copas and Li (1997), highlighting in turn connections in the econometric literature to the Heckman model. The other theme is represented by the literature on stochastic frontier models, pioneered by Aigner et al. (1977); a fairly recent account is Coelli et al. (1998). The stochastic frontier model is closely related to representation (3.3) here.

Another, very general, comment which I would like to make is not really related to the paper itself, but to the literature on "skewed models" as a whole. This has developed quite fast in the last five years or so, but the body of theoretical results which have been developed has not yet been widely applied to real problems. The practical usefulness of these stochastic models is the real benchmark to test whether these results are going to have actual relevance in the statistical practice. This is a plea for more applied work to make use of the available theoretical tools.

References

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model. *Journal of Econometrics*, 12:21–37.

Coelli, T., Prasada Rao, D. S., and Battese, G. E. (1998). *An introduction to efficiency and productivity analysis, Chapter 8-9*. Kluwer Academic Publishers, Boston, Dordrecht, London.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society B*, 59:55–95.

N. Balakrishnan
McMaster University, Canada

First of all, I would like to congratulate Professors Arnold and Beaver for providing an excellent discussion on skew-normal distribution, different derivations, various generalizations, and multivariate extensions of it. The authors' idea of using *hidden truncation* to propose univariate generalizations as well as natural multivariate extensions of skew-normal distributions is quite interesting. The approach is flexible enough to derive general forms as well. As the authors have pointed out at the end of Section 2, the skew-normal distribution was present either explicitly or implicitly in the early works of Birnbaum (1950), Nelson (1964), Weinstein (1964), and O'Hagan and Leonard (1976). Yet, it is the article of Azzalini (1985) that generated a lot of activity on this family of distributions. Considerable attention seems to have been paid to deriving generalizations and multivariate forms. It appears to me that more attention needs to be paid to the development of efficient inferential methods for these models and also on discrimination between these models and formal tests of fit. If significant progress is made on these aspects, then the practical applicability of these distributions will be greatly enhanced.

In what follows, I will present a new motivation of the skew-normal distribution and point a close connection of it to normal order statistics. I shall then use this approach to derive two more general families of skew-normal distributions (as they are similar in form to Azzalini's skew-normal distribution in (2.1)). For $n = 0, 1, 2, \cdots$, let us consider the integral

$$L_n(\lambda) = \int_{-\infty}^{\infty} \{\Phi(\lambda z)\}^n \;\; \varphi(z) dz, \quad z \in \mathbb{R}. \tag{1}$$

Clearly, we have $L_0(\lambda) = 1,\ \lambda \in \mathbb{R}$. Then, since

$$\int_{-\infty}^{\infty} \left\{\Phi(\lambda z) - \frac{1}{2}\right\}^{2n+1} \;\; \varphi(z) dz = 0,$$

because the integrand is an odd function of z, we readily have

$$L_{2n+1}(\lambda) = \sum_{i=1}^{2n+1} (-1)^{i+1} \binom{2n+1}{i} \frac{1}{2^i} L_{2n+1-i}(\lambda), n = 0, 1, 2, \cdots, \lambda \in \mathbb{R}. \tag{2}$$

Equation (2) implies (for $n = 0$)

$$L_1(\lambda) = \int_{-\infty}^{\infty} \Phi(\lambda z)\ \varphi(z) dz = \frac{1}{2} L_0(z) = \frac{1}{2}, \quad \lambda \in \mathbb{R},$$

which leads to the skew-normal density (cf. (2.1))

$$f_1(z;\lambda) = 2\Phi(\lambda z)\ \varphi(z), \quad z \in \mathbb{R}, \quad \lambda \in \mathbb{R}. \tag{3}$$

Next, from

$$L_2(\lambda) = \int_{-\infty}^{\infty} \{\Phi(\lambda z)\}^2\ \varphi(z) dz, \quad \lambda \in \mathbb{R}, \tag{4}$$

we find

$$\begin{aligned}
\frac{dL_2(\lambda)}{d\lambda} &= 2\int_{-\infty}^{\infty} z\Phi(\lambda z)\ \varphi(\lambda z)\ \varphi(z) dz \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \Phi(\lambda z)\ z\ \exp\left\{-\frac{1}{2} z^2 (1+\lambda^2)\right\} dz \\
&= -\frac{1}{\pi(1+\lambda^2)} \int_{-\infty}^{\infty} \Phi(\lambda z)\ \frac{d}{dz}\left[\exp\left\{-\frac{1}{2} z^2(1+\lambda^2)\right\}\right] \\
&= \frac{\lambda}{\pi(1+\lambda^2)} \int_{-\infty}^{\infty} \varphi(\lambda z)\ \exp\left\{-\frac{1}{2} z^2(1+\lambda^2)\right\} dz \\
&= \frac{\lambda}{\pi(1+\lambda^2)\sqrt{1+2\lambda^2}},
\end{aligned}$$

which yields

$$L_2(\lambda) = \frac{1}{\pi}\ \tan^{-1}\sqrt{1+2\lambda^2}, \quad \lambda \in \mathbb{R}.$$

This, along with (4), then leads to another skew-normal density

$$f_2(z;\lambda) = \frac{\pi}{\tan^{-1}\sqrt{1+2\lambda^2}} \{\Phi(\lambda z)\}^2 \varphi(z), \quad z \in \mathbb{R}, \quad \lambda \in \mathbb{R}. \tag{5}$$

Note that this family of distributions also includes the standard normal distribution when $\lambda = 0$ and the half normal distribution when $\lambda \to \infty$ (just as f_1 does).

Next, (2) implies (for $n = 1$)

$$\begin{aligned} L_3(\lambda) &= \frac{3}{2}L_2(\lambda) - \frac{3}{4}L_1(\lambda) + \frac{1}{8}L_0(\lambda) \\ &= \frac{3}{2}L_2(\lambda) - \frac{1}{4} = \frac{3}{2\pi}\ \tan^{-1}\sqrt{1+2\lambda^2} - \frac{1}{4}. \end{aligned}$$

This leads to yet another skew-normal density

$$f_3(z;\lambda) = \frac{1}{\frac{3}{2\pi}\tan^{-1}\sqrt{1+2\lambda^2} - \frac{1}{4}} \{\Phi(\lambda z)\}^3\ \varphi(z), \quad z \in \mathbb{R}, \quad \lambda \in \mathbb{R}. \quad (6)$$

This family of distributions also includes standard normal distribution when $\lambda = 0$ and the half normal distribution when $\lambda \to \infty$.

Evidently, in the special case when $\lambda = 1$, the densities f_1, f_2 and f_3 in (3), (5) and (6) become the densities of the largest order statistic in a sample of size 2, 3 and 4, respectively, from the standard normal distribution. In addition to this obvious connection to order statistics, the integral $L_n(\lambda)$ in (1) also plays a role in the derivation of means of standard normal order statistics as shown by Bose and Gupta (1959). For example, with $\mu_{m:m}$ denoting the mean of the largest order statistic in a sample of size m from standard normal distribution, Bose and Gupta (1959) have shown that

$$\mu_{2:2} = \frac{1}{\sqrt{\pi}}L_0,\ \mu_{3:3} = \frac{3}{\sqrt{\pi}}L_1(1),\ \mu_{4:4} = \frac{6}{\sqrt{\pi}}L_2(1), \ \text{ and }\ \mu_{5:5} = \frac{10}{\sqrt{\pi}}L_3(1);$$

for pertinent details, one may also refer to David (1981) and Arnold et al. (1992).

The densities $f_2(z;\lambda)$ and $f_3(z;\lambda)$ in (5) and (6) may be further studied in terms of properties, interpretations, inferential issues, etc.

References

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. John Wiley and Sons, New York.

Bose, R. C. and Gupta, S. S. (1959). Moments of order statistics from a normal population. *Biometrika*, 46:433–440.

David, H. A. (1981). *Order Statistics. Second Edition.* John Wiley and Sons, New York.

A. Bhaumik and D. K. Dey
University of Connecticut, USA

First let us compliment Professor Barry Arnold and Professor Robert Beaver on their excellent paper on Skewed Multivariate distributions. Though it was primarily devoted to the scope of application of skewed multivariate distributions, it is a useful summary of all the recent developments in skewed distributions. It is very unfortunate that the development of this class of distribution has progressed very slowly after its introduction by Azzalini (1985). Some of the recent papers including Arnold and Beaver (2000a,b,c), Azzalini and Dalla Valle (1996), Branco and Dey (2001), Gupta and Chen (2001), Sahu et al. (2001) showed various scopes of this distribution. This paper not only demonstrates its practical applicability but also gears towards the development of non normal skewed distributions. It is really hard to add something interesting to this complete review. In view of that we would like to add three recent applications of skewed distributions and one possible area where it's application is very meaningful.

Regression with skew error

Sahu et al. (2001) has proposed a Bayesian model with skewed elliptical error which can eliminate the need of ad-hoc transformation for symmetry of skewed data. This model is also capable of taking care of unobserved random effects. Formulation of this new skewed elliptical family is done first by defining general elliptical distribution and then conditioning on a given linear constraint of two independent elliptical distributions. Following is the complete formulation,

$\mathbf{X}$, a m-dimensional random vector is said to be elliptically distributed if its pdf is of the form

$$f\left(\mathbf{x}|\theta, \mathbf{\Omega}; g^{(m)}\right) = |\mathbf{\Omega}|^{-\frac{1}{2}} g^{(m)}\left[(\mathbf{x}-\theta)'\mathbf{\Omega}^{-1}(\mathbf{x}-\theta)\right], \quad \mathbf{x}\epsilon\mathbb{R}^m,$$

where $g^{(m)}(u)$, the density generator of $\mathbf{X}$, is a non-increasing function from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that the integral $\int_0^\infty r^{m/2-1} g(r;m)dr$ exists. We will denote this distribution by $EL(\theta, \Omega; g^{(m)})$. Different choices of it yields well known distributions like multivariate normal, multivariate-t etc. Let

$$\mathbf{Y} = \mathbf{DZ} + \epsilon$$

where

$$\begin{pmatrix} \epsilon \\ \mathbf{Z} \end{pmatrix} \backsim EL\left(\theta = \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \mathbf{\Omega} = \begin{pmatrix} \mathbf{\Sigma} & 0 \\ 0 & \mathbf{I} \end{pmatrix}; g^{(2m)}\right) \tag{1}$$

and $\mathbf{D}$ is a diagonal matrix with elements $\delta_1, \delta_2, ..., \delta_m$. This new skewed elliptical class of distributions is developed by obtaining the distribution of the random variable $[\mathbf{Y}|\mathbf{Z} > 0]$, where $\mathbf{Z} > 0$ means $\mathbf{Z}_i > 0$, for $i = 1, 2, ..., m$.

The proposed regression model is a usual regression model except the error term is skewed elliptically distributed. Let $\mathbf{Y}_i \backsim SE(\mu_i, \mathbf{\Sigma}, \mathbf{D}; g^{(m)})$ for $i = 1, 2, ..., n$. For each data point with the covariate information assumed in a $p \times m$ matrix $\mathbf{X}_i$, the linear model is specified as

$$\mu_i = \mathbf{X}_i^{'}\beta.$$

Using a Bayesian viewpoint, usual conjugate priors on β and $\mathbf{\Sigma}$ are taken to get the tractable expressions of the posterior distribution. For the ease of MCMC computation hierarchical setup of $f(\mathbf{y}|\mathbf{z})$ and $f(\mathbf{z})I(\mathbf{z} > \mathbf{0})$ are used. The form of these two distributions are same as (1) which is,

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \backsim EL\left(\theta = \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \mathbf{\Omega} = \begin{pmatrix} \mathbf{\Sigma} + \mathbf{D}^2 & \mathbf{D} \\ \mathbf{D} & \mathbf{I} \end{pmatrix}; g^{(2m)}\right) \tag{2}$$

which implies

$$\mathbf{Y}|\mathbf{Z} = \mathbf{z} \backsim EL(\mu + \mathbf{Dz}, \mathbf{\Sigma}; g_{q(z)}^{(2m)}).$$

with $q(z) = \mathbf{z}^{'}\mathbf{z}$. Four models, normal, skewed-normal, t and skewed-t are studied with four real life data set and are compared with Bayes factor. One important observation in these studies is that the skewness parameter is prior sensitive whereas the regression parameter is not. The diffuse prior allows the regression coefficients to wander around with larger variation. This influences the skewness parameter considerably. Under diffuse prior it takes much higher value. The estimate of the variance also changes under different error distributions. Under the skewed setup it is a little smaller, which indicates that the skewness parameter is taking care of some variability present in the data.

The above formulation of the model is quite simple but very general in nature. It is also quite extendable towards different directions. As mentioned by Arnold and Beaver the linear constraint of $\mathbf{Z} > 0$ can be replaced by $\mathbf{Z} > \mathbf{c}$ for some non zero constant $\mathbf{c}$. Variable selection and model averaging are two other possible issues that can also be explored under skewed situation.

Binary response data with skew link

Chen et al. (1999) showed the applicability of skewed link function to model binary response data in particular when the probability of a given binary response approaches 0 at a different rate than it approaches to 1. Symmetric link functions like logit, probit are inappropriate in this situation. There are many proposals of the use of different kinds of asymmetric link functions. One very general link function is Stukel's generalized logistic link function (see Stukel (1988)). But one problem with this link function is that it yields improper posterior distribution in the presence of covariates for many types of noninformative improper priors. The method introduced by Chen et al. (1999) not only yields proper posterior under a wide class of noninformative priors, but also it includes many different types of commonly used symmetric and asymmetric link functions. Also it is computationally very attractive because of its formulation using latent variables. The formulation is very simple and straightforward. The proposed model is

$$y_i = \begin{cases} 0 & \text{if } w_i < 0 \\ 1 & \text{if } w_i \geq 0 \end{cases},$$

where $\mathbf{w} = (w_1, w_2, ..., w_n)'$ is a vector of independent latent variables with the following setup,

$$w_i = \mathbf{x}_i'\beta + \delta z_i + \epsilon_i, \qquad z_i \backsim G,$$

and

$$\epsilon_i \backsim F,$$

z_i and ϵ_i are independent, G is the cdf of a skewed distribution, and F is the cdf of a symmetric distribution. Using different examples with real life data and empirical study, it was shown that the asymmetric link model is more appropriate than symmetric models when the number of 0's are much different than number of 1's in the data. One important issue in this model is the choice of G and F. The choices used in that article was normal or logistic for F and standard half-normal for G. Nonparametric mixture models such as Dirichlet process mixture model and more general class of skewed link models are a few directions to be explored.

Modelling frailty distribution in survival analysis

Sahu and Dey (2001) proposed the use of a class of log-skew-t distributions to model the dependence in multivariate survival data. Gamma distribution is most popular and mathematically simple to model frailty. But one disadvantage with this distribution is that it weakens the effect of the covariates. Hougaard (1986), Qiou et al. (2000) and Ravishanker and Dey (2000) discussed the use of heavy tailed positive stable distributions to overcome this problem. The proposed class of distributions often have heavier tails than the gamma and positive stable distribution and include many other heavy tailed distribution like log-Cauchy, log-t etc. So, these are more suitable for studying robustness. A correlated prior process is proposed for the baseline hazard function, which imposes smoothness on the baseline hazard function in adjacent intervals. In addition to this the endpoints of the interval themselves are assumed to form a time homogeneous Poisson process. So the number and position of the jump can be estimated using data and prior assumptions. The formulation of the model follows :

Let the frailty parameters $b_i, (= \log w_i)$ $i = 1, 2, ..., n$ be independent and identically distributed for every group with the following density,

$$\begin{aligned} f_b(b|\rho, \delta, \nu) &= 2(\rho^2+\delta^2)^{-1/2}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)(\nu\pi)^{1/2}}\left[1+\frac{b^2}{\nu(\rho^2+\delta^2)}\right]^{-(\nu+1)/2} \\ &\quad \times \mathbf{T}_{\nu+1}\left[\sqrt{q(b)}\frac{\delta}{\rho}\frac{b}{\sqrt{\rho^2+\delta^2}}\right] \qquad (3) \end{aligned}$$

where

$$q(b) = \frac{\nu+1}{\nu+b^3/(\rho^2+\delta^2)}$$

and $\mathbf{T}_m(.)$ is the cumulative distribution function of the standard t-distribution with m degrees of freedom. The parameters ρ, δ and ν influence the shape of the distribution. With $\delta = 0$ the above distribution reduces to a standard t-distribution with ν df. In addition if $\nu \longrightarrow \infty$ then it approaches the normal distribution with ρ^2 as scale parameter. With $\delta = 0$ and $\nu = 1$ (3) is the pdf of the Cauchy distribution. For $\delta \neq 0$ a skewed distribution is obtained. The mean of this distribution exists if $\nu > 1$ and variance exists if $\nu > 2$. Mean and variance of b are of the form

$$E(b) = \left(\frac{\nu}{\pi}\right)^{1/2}\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\delta,$$

and

$$\text{Var}(b) = (\rho^2 + \delta^2)\frac{\nu}{\nu - 2} - \frac{\nu}{\pi}\left(\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\right)^2 \delta^2.$$

The density of $w = \exp(b)$ implied by (3) is given by

$$f_w(w|\rho, \delta, \nu) = \frac{1}{w} f_b(\log(b)|\rho, \delta, \nu) \tag{4}$$

where $f_b(.|\rho, \delta, \nu)$ is given in (3).

The assumption of correlated processes for the baseline hazard function is very common in survival analysis. References are due to Gamerman (1991), Arjas and Gasbarra (1994), Sinha and Dey (1997) and Sahu et al. (1997). Here constant baseline hazard is assumed within each interval. More general models are discussed in Gamerman (1991) and Arjas and Gasbarra (1994).

Cross-ratio function is used to compare local dependence structure among the frailty models. Unfortunately this function does not have analytical closed form under log-skew-t frailty distribution which limits the scope of the theoretical comparison. Correlation of log-survival times is used to study frailty models using a global measure of dependance. Since moments of the frailty distribution w do not always exist the correlation between survival times themselves, is not meaningful in this setup. One interesting observation was that the tail of the frailty distribution plays an important role in dictating the dependance structure.

Using various graphical methods the proposed model is shown to be a viable alternative to the gamma and stable frailty models. Though the introduction of skewed distributions made the formulation little more complicated, it is a more flexible approach to model frailty.

Skewed distributions in time series analysis

Time series analysis probably is another topic where skewed distribution will give more flexibility. Azzalini (1985) formulated a dynamic model with normality in observational error and skewed normal error in parameter which is changing with time. The formulation of his model is very straight forward:

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{W}_t + \eta_t \\ \mathbf{W}_t &= \varrho \mathbf{W}_{t-1} + \varepsilon_t, \end{aligned}$$

where η_t is white noise $N(0, \sigma^2)$. $\{\eta_t\}$ and $\{\varepsilon_t\}$ are independent. Under normality assumption on the initial prior on $\mathbf{W}_0$ all the conditional distributions of $\mathbf{W}_t$ given $(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_t)$ are normal. What Azzalini (1985) has shown is that the similar property holds if we change the initial prior on $\mathbf{W}_0$ from normal to skewed normal distribution. The advantage of this formulation is that it is computationally as intensive as usual normal model but more flexibility will be gained through skewness parameter. Azzalini used most general version of skewed normal distribution. It is of the form

$$\phi(t; \lambda_1, \lambda_2, \lambda, \xi) = \frac{\phi\left(\frac{t-\lambda_1}{\lambda_2}\right) \Phi\left\{\lambda\left(\frac{t-\lambda_1}{\lambda_2}\right) + \xi\right\}}{\left\{\lambda_2 \Phi\left(\frac{\xi}{\sqrt{(1+\lambda^2)}}\right)\right\}}.$$

The above model can be extended by taking skewed distributions in observational error which will be able to accommodate skewed time series data. Any available covariate information can easily be included in the observation equation. Also further generalization can be obtained through the introduction of skewed elliptical distributions.

As an overall conclusion, we could say that multivariate skewed distributions have a tremendous scope of application in different variety of statistical modeling. Professors Arnold and Beaver's paper has done a great job to motivate the use of skewed distributions.

References

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference for right-censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4:505–524.

Chen, M., Dey, D. K., and Shao, Q. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94:1172–1186.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics*, 40:63–79.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73:671–678.

Qiou, Z. Q., Ravishanker, N., and Dey, D. K. (2000). Multivariate survival analysis with positive stable frailties. *Biometrika*, 55:637–644.

Ravishanker, N. and Dey, D. K. (2000). Multivariate survival models with a mixture of positive stable frailties. *Methodology and Computing in Applied Probability*, 2:293–308.

Sahu, S. K. and Dey, D. K. (2001). Regression with robust frailty distribution in survival analysis. Technical report, Department of Statistics, University of Connecticut, Connecticut.

Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D. (1997). A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis*, 6:207–228.

Sahu, S. K., Dey, D. K., and Branco, C. D. (2001). A new class of multivariate skew distribution with application to Bayesian regression models. Technical report, Department of Statistics, University of Connecticut, Connecticut.

Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1192–1212.

Stukel, T. (1988). Generalized logistic models. *Journal of the American Statistical Association*, 83:426–431.

C. M. Cuadras

Universitat de Barcelona, Spain

Arnold and Beaver present a useful and well written overview on skew multivariate models, including skew-normal, with new results on inference. I enjoyed reading this article and I only would like to add an aspect concerning mixtures of distributions.

The starting point of this work is the Azzalini distribution with density (2.1), which has nice properties. This distribution generalizes the fact that $\Phi^2(z)$ is a cdf, the pdf being $2\varphi(z)\Phi(z)$, the density of $\max(W_1, W_2)$, where

W_1, W_2 are i.i.d. $N(0,1)$. Azzalini introduces a parameter and considers the family $SN(\lambda)$ with pdf $2\varphi(z)\Phi(\lambda z)$, which is $N(0,1)$ for $\lambda = 0$ and converges to the half-normal as $\lambda \to \infty$. Thus λ is a skewness parameter and $SN(\lambda)$ is useful for modelling skew data.

The family $SN(\lambda)$ contains the normal and is different from normal if $\lambda \neq 0$. However, the normal distribution can appear by mixture. First note that

$$\int_R 2\varphi(z)\Phi(\lambda z)\varphi(\lambda)d\lambda = \varphi(z),$$

and the mixture of $SN(\lambda)$ when λ is $N(0,1)$ is also $N(0,1)$. More generally, if the pdf $f(\lambda)$ is symmetric, $f(\lambda) = f(-\lambda)$, then

$$\begin{aligned} \textstyle\int_R 2\varphi(z)\Phi(\lambda z)f(\lambda)d\lambda &= \textstyle\int_0^\infty 2\varphi(z)[\Phi(\lambda z) + \Phi(-\lambda z)]f(\lambda)d\lambda \\ &= 2\varphi(z)\textstyle\int_0^\infty f(\lambda)d\lambda = \varphi(z). \end{aligned}$$

Therefore, if W_1, W_2, Y_1, Y_2, Z, Y are $N(0,1)$, and λ is random with symmetric distribution, then: a) if W_1, W_2 have correlation $\rho = (1-\lambda^2)/(1+\lambda^2)$, the mixture distribution of $\max(W_1, W_2)$ is $N(0,1)$, b) if Z, Y have correlation ρ such that $\lambda = \rho/\sqrt{1-\rho^2}$, the mixture of Z given $(Y > 0)$ is $N(0,1)$, c) if Y_1, Y_2 are independent and λ is distributed on $(-1,1)$, the mixture distribution of $\lambda|Y_1| + \sqrt{1-\lambda^2}Y_2$ is again $N(0,1)$.

These mixture properties do not apply for the extended family (2.3) proposed by Arnold and Beaver, except for $\lambda_0 = 0$, so providing a more general family. However, the mixture of (2.3) for λ_1 symmetrically distributed may give, practically, the normal distribution. Thus the mixture

$$p(z) = \int_R 2\varphi(z)\frac{\Phi(\lambda_0 + \lambda_1 z)}{\Phi\left(\lambda_0/\sqrt{1+\lambda_1^2}\right)}f(\lambda_1)d\lambda_1$$

is practically $N(0,1)$ when λ_1 is symmetric normal, Laplace or Cauchy for $\lambda_0 \geq 4$, or when λ_1 is symmetric beta on $(-1/2, 1/2)$, (i.e., with parameters $\alpha = \beta$ and subtracting $1/2$ from the beta variable) for $\lambda_0 \geq 0$. The difference may be appreciable for $\lambda_0 < 0$, but even there is a quite small distinction between $p(z)$ and $N(0,1)$ in the normal, Laplace or Cauchy distribution for λ_1 when λ_0 ranges in the interval $(0,4)$. Statistically speaking, it is quite difficult to reject the normality of a sample following the distribution (2.3) when $\lambda_0 \geq 0$ and λ_1 is random and symmetrically distributed.

The multivariate skew-normal (4.4) has a similar property: if $\underline{\lambda} = (\lambda_1, \ldots, \lambda_k)'$ is a random vector of i.i.d. $N(0,1)$, then the mixture is also normal

$$\int_{R^k} 2\prod_{i=1}^{k}\varphi(w_i)\Phi(\underline{\lambda}'\underline{w})\prod_{i=1}^{k}\varphi(\lambda_i)d\lambda_i = \prod_{i=1}^{k}\varphi(w_i).$$

Even, if $\lambda_1, \ldots, \lambda_k$ are independent and the pdf of each λ_i is symmetric, $f_i(\lambda_i) = f_i(-\lambda_i)$, then the mixture is

$$\begin{aligned}
&\int_{R^k} 2\prod_{i=1}^{k}\varphi(w_i)\Phi(\underline{\lambda}'\underline{w})\prod_{i=1}^{k}f_i(\lambda_i)d\lambda_i \\
&= 2\prod_{i=1}^{k}\varphi(w_i)\int_{\underline{\lambda}>0}[\Phi(\underline{\lambda}'\underline{w}) + \Phi(-\underline{\lambda}'\underline{w}) + \ldots]\prod_{i=1}^{k}f_i(\lambda_i)d\lambda_i \\
&= 2\prod_{i=1}^{k}\varphi(w_i)2^{k-1}\int_{\underline{\lambda}>0}\prod_{i=1}^{k}f_i(\lambda_i)d\lambda_i = \prod_{i=1}^{k}\varphi(w_i),
\end{aligned}$$

again multivariate normal.

The family (4.3) proposed by Arnold and Beaver has nice features, as marginals and conditionals distributions belong to the same family, and specific quadratic forms may be chi-square. The above mixture property does not hold except for $\lambda_0 = 0$. However, as in the univariate case, the mixture may give approximately a multivariate normal distribution.

As a conclusion, the skew-normal distribution may model the data with hidden truncation when λ_1 is fixed, but the normal distribution may yet be correct when λ_1 is random.

J. M. Sarabia
University of Cantabria, Spain

I would like to congratulate Professors Arnold and Beaver for this nice paper about skewed multivariate models, and their relation with truncation and selective reporting. Many of the ideas developed by Arnold and Beaver should have impact in the field of data analysis. For example, if we think about modeling observations (univariate or multivariate) taken from psychological, biometrical or economical sciences, on individuals with some type of previous selection, we should work with this type of distributions.

Arnold and Beaver present several interesting examples of this situation. This fact has already been noted by Azzalini and Dalla Valle (1996) using data of weight and height in australian athletes. I shall divide my comments in two parts. The first part is devoted to the stochastic frontier models. The second part is dedicated to the study of a class of multivariate skewed distributions based on conditional specification.

Stochastic frontier models

The frontier models together with the associated evaluation of efficiency of the economic units are an active research field in econometrics. The production or cost frontier itself is not supposed to be known exactly and usually some symmetric error is attached to it. A second type of error is introduced to represent deviations of the individual units from the frontier (inefficiency) which is by definition of a one-sided nature. Then, the stochastic frontier model takes the form:

$$y_i = f(x_i; \beta) + v_i - |u_i|,$$

with v_i symmetric. Among the proposals for $|u_i|$, one encounters the exponential distribution of Meeusen and van den Broeck (1977), the half-nomal of Aigner et al. (1977), the truncated normal proposed by Stevenson (1980), the gamma distribution of Greene (1990) and several proposals of Broeck et al. (1994) from a Bayesian framework. The most common specification consists of assuming $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim N(0, \sigma_u^2)$, to obtain the pdf:

$$f(\varepsilon_i; \lambda, \sigma) = \frac{2}{\sigma} \phi\left(\frac{\varepsilon_i}{\sigma}\right) \Phi\left(-\lambda \frac{\varepsilon_i}{\sigma}\right),$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u / \sigma_v$. Several generalizations of the previous models immediately arise based on Arnold and Beaver ideas. The following robust model can be obtained by applying (3.2) to the classical t-distribution:

$$f(x; \lambda, \nu, \sigma) = \frac{2}{\sigma} f_{t_\nu}\left(\frac{x}{\sigma}\right) F_{t_\nu}\left(\lambda \frac{x}{\sigma}\right), \tag{1}$$

where f_{t_ν} and F_{t_ν} represent the pdf and cdf of a t-distribution with ν degrees of freedom, respectively. If $\nu \to \infty$ the classical skewn-normal distribution is obtained. The even moments of (1) coincide with the even moments of a t-distribution, and in consequence $\mathrm{E}(X^2) = \nu(\nu-2)^{-1}\sigma^2$, if $\nu > 2$. The odd moments don't seem easy to obtain. For example, for $\nu > 1$ one gets:

$$\mathrm{E}(X) = \frac{2\nu\lambda\sigma}{(\nu-1)k_\nu}\mathrm{E}\left\{\left(1+Z_\nu^2/\nu\right)\left(1+\lambda^2 Z_\nu^2/\nu\right)^{-(\nu+1)/2}\right\},$$

where $Z_\nu \sim t_\nu$ and $k_\nu = \sqrt{\nu}B(\frac{1}{2},\frac{\nu}{2})$. A simple approximation is given by:

$$\mathrm{E}(X) \approx \frac{2\nu\lambda\sigma}{(\nu-1)k_\nu}\left\{1+\left(1-(\nu+1)\lambda^2/2\right)\frac{1}{\nu-2}\right\}.$$

This basic model and the different multivariate extensions require a more detailed analysis.

A class of multivariate skewed distributions based on conditional specification

In this section we present a general class of multivariate skewed distributions from the conditional perspective included in Section 8 of the paper and largely exposed in Arnold et al. (1999, 2001). Let $X \sim SD(\psi_1, \Psi_2; \lambda)$ be a random variable with pdf (3.2), where ψ_1 and Ψ_2 are symmetric. Let us assume one is interested in obtaining the most general k-dimensional random variable with univariate conditional distributions such that:

$$X_i|\underline{X}_{(i)} = \underline{x}_{(i)} \sim SD(\psi_i; \Psi; \lambda_i(\underline{x}_{(i)})), \quad i = 1,2,\ldots,k \tag{2}$$

for some functions $\lambda_i(\cdot)$. Note that the cdf Ψ is common to all the conditional distributions. Two types of solutions appear in (2). The first one corresponds to the independence case, where λ_i is constant, $i = 1,2,\ldots,k$. After introducing location and scale parameters, the second leads to the k-dimensional density.

$$f(x_1,\ldots,x_k;\lambda,\underline{\mu},\underline{\sigma}) = 2\prod_{i=1}^{k}\left[\frac{1}{\sigma_i}\psi_i\left(\frac{x_i-\mu_i}{\sigma_i}\right)\right]\Psi\left(\lambda\prod_{i=1}^{k}\frac{x_i-\mu_i}{\sigma_i}\right). \tag{3}$$

The family (3) has some unusual properties. First at all, the normalizing constant is known (note that in many models of the type (2) the normalizing constant should be evaluated numerically). The marginal distributions of X_i have pdf $\frac{1}{\sigma_i}\psi_i\left(\frac{x_i-\mu_i}{\sigma_i}\right)$ and the univariate conditional distributions are of the form (2) with $\lambda_i(\underline{x}_{(i)}) = \lambda \prod_{j \neq i} (x_j - \mu_j)/\sigma_j$, $i = 1, 2, \ldots, k$. A non expected property is the multimodality. In the normal case with $\psi_i = \phi$, $i = 1, 2, \ldots, k$ and $\Psi = \Phi$, the modes are the solutions of the system:

$$\lambda \prod_{j \neq i} x_j \phi \left(\lambda \prod_{i=1}^{k} x_i \right) - x_i \Phi \left(\lambda \prod_{i=1}^{k} x_i \right) = 0, \;\; i = 1, 2, \ldots, k.$$

According to Arnold et al. (2002), in the bidimensional case we have bimodality if $|\lambda| > \sqrt{\pi/2}$. As a consequence, this class of models are a reasonable alternative to the finite mixtures of distributions, for modelling heterogeneity. This implies an important reduction in the number of parameters. More versatility could be obtained by considering the pdf (3.1) or (3.2) with location and scale parameters. For example, the model (8.3) in the bidimensional case contains as submodels: the product of independent normal densities, a model with skewed normal conditionals and normal marginals, a model with skewed normal marginals and linearly skewed normal conditionals and the model (4.3) with linearly skewed normal marginals and conditionals.

Acknowledgment

I am grateful to Ministerio de Ciencia y Tecnología for partial support, project BEC2000-1186.

References

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999). *Conditional specification of statistical models*. Springer Series in Statistics, Springer Verlag, New York.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2001). Conditionally specified distributions: an introduction (with discussion). *Statistical Science*, 16:249–274.

Broeck, J. van den, Koop, J. O., and Steel, M. (1994). Stochastic frontier models. A Bayesian perspective. *Journal of Econometrics*, 61:273–303.

Greene, W. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46:141–163.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 8:435–444.

Stevenson, R. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13:57–66.

Rejoinder by Barry C. Arnold and Robert J. Beaver

We are grateful to all those who have contributed to the discussion of our article. Several common themes can be noted in their remarks. The skew-normal distribution (and many of its variants and relatives) has surfaced in the quantitative literature related to a wide variety of scientific disciplines. We were aware of some modelling involving hidden and/or unobserved truncation in economic settings. It is however quite appropriate that the close parallel between hidden truncation models as discussed in our paper and stochastic frontier models be highlighted. Both Professor Azzalini and Professor Sarabia provide useful links, discussion and references related to the stochastic frontier avatar of hidden truncation. Professor Azzalini also reminds us of the relation to biased sampling (a commonly postulated genesis of weighted distributions (see, e.g., Rao (1985)), of which our hidden truncation models are indeed a special case.

Professor Cuadras draws attention to models obtained by mixing skewed distributions of the Azzalini type. Typically such mixture models are more complicated (involving more parameters) than are the component distributions. But as Professor Cuadras points out, this is not always the case. Mixing can take us back to the simple symmetric normal model. Indeed an analogous phenonenon can be encountered in the more general (non-normal) model (3.2) provided that the distribution function Ψ_2 in (3.2) is symmetric and the mixing distribution for λ_1 is also symmetric. It is intriguing to speculate on the consequences of such "simplification by mixing" in a Bayesian context where the role of the mixing distribution for the skewness parameters can be played by a subjective or objective prior distribution. It is also interesting that, even in the extended family of densities (2.3) (or (3.1)), mixtures can be hard to distinguish from the standard normal model (or from Ψ_2 in the case of the more general model (3.1).

Professor Sarabia has provided a good introduction to non-normal conditionally specified distributions involving hidden truncation. There are still interesting open questions regarding how to characterize all multivariate densities with given skewed conditionals. Hopefully, Professor Sarabia's comments will prompt someone to make progress in this direction.

It is not surprising that Professor Balakrishnan would find order statistics lurking in the background of our skew-normal discussion. It has been observed by several researchers that the skew-normal family (2.1) includes as special cases the standard normal density when ($\lambda = 0$), the density of the minimum of two i.i.d. standard normal variables $Z_{1:2}$ (when $\lambda = -1$), and the density of the maximum of two standard normal variates $Z_{2:2}$ (when $\lambda = 1$). The form of the density (2.1) evokes images of the density of $Z_{2:2}$ to anyone who has more than a minimal amount of experience with order statistics. Professor Balakrishnan invites us to consider models in which we consider a weight function of the form $[\Phi(\lambda z)]^n$ which will be related (when $\lambda = 1$) to the distribution of maxima of standard normal samples of size greater than 2. He is able to obtain the explicit analytic form of the required normalizing constant in the cases $n = 1$ (already known), $n = 2$ and $n = 3$ (not previously known). For other values of n, the normalizing constant will most likely need to be obtained by numerical integration for various values of λ. Professor Balakrishnan's comments lead naturally to consideration of the following very general family of skewed normal distributions:

$$f(z; \lambda_0, \lambda_1, \alpha) \propto [\Phi(\lambda_0 + \lambda_1 z)]^\alpha \varphi(z),$$

where $\lambda_0, \lambda_1 \in \mathbb{R}$ and $\alpha \in \mathbb{R}^+$. Except in special cases, the corresponding normalizing constant will be unavailable in closed form but, for example, numerical implementation of maximum likelihood estimation should be feasible.

But, as Professors Azzalini and Balakrishnan point out, we are less in need of models with more bells and whistles such as these, than we are of some illustrations of practical utility of all these skewed models. They both call for more practical examples.

It is consequently more serendipity than we deserve that Professors Bhaumik and Dey chose to provide us with a catalog of settings in which such skewed models might be profitably utilized. They point out that such skew models can provide added flexibility when used as error distributions, link functions, frailty structures and shock processes in time series settings.

There is undoubtedly still much scope for development of improved estimation and inference procedures associated with all these complicated models. Misbehaving likelihood functions will abound but progress in overcoming this difficulty for the basic Azzalini model has already been documented. We are hopeful that advances will soon be made in more general settings as well.

The constructive and creative comments made by the discussants will surely help to kindle interest in hidden truncation models and to encourage progress in developing needed inferential techniques. Again, we thank them for sharing their ideas with us.

References

Rao, C. (1985). Weighted distributions arising out of methods of ascertainment: what population does a sample represent? In *A Celebration of Statistics. The ISI Centenary Volume*, pp. 543–570. Springer-Verlag, New York.