ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

Alternative constructions of skewed multivariate distributions

BARRY C. ARNOLD AND ROBERT J. BEAVER

ABSTRACT. A review of the construction of skewed multivariate normal distributions is presented. The review considers construction via (1) hidden truncation, (2) threshold models, (3) additive components and (4) a location and scale change for k variables beginning with $k-1$ independent standard normal variates and one univariate skew normal density. Extensions to non-normal distributions have mainly used the hidden truncation approach. Unlike the normal case, the use of the three remaining techniques in constructing non-normal multivariate distributions leads to models distinct from those found using the hidden truncation approach. Examples of several tractable multivariate distributions using methods (1) and (3) are also presented.

1. Introduction

The Azzalini skew-normal density is a univariate density of the form

$$f(x;\lambda) = 2\varphi(x)\Phi(\lambda x) \tag{1}$$

where φ and Φ denote the standard normal density and distribution functions, respectively. A variety of univariate and multivariate extensions of this distribution have been considered in the literature. A recent survey may be found in Arnold and Beaver (2002a). The usual probabilistic genesis of variables with such skewed distributions involves a scenario in which random variables (and/or vectors) are observed only if they satisfy certain linear constraints or equivalently if some linear combination of the variables exceeds a given threshold. Arnold and Beaver (2000a) refer to those as distributions that are skewed via hidden truncation. In the normal case an equivalent representation of these models involves an additive component structure. The present paper is concerned with several alternative constructions of skewed distributions (including the additive component models) in

Received October 7, 2003.
2000 *Mathematics Subject Classification.* 62H05, 62H12.
Key words and phrases. Skewed distributions, hidden truncation, additive components.

both normal and non-normal cases. We will confirm their identification with hidden truncation models in the normal case and will observe that new families of skewed distributions (distinct from hidden truncation models) arise in non-normal settings.

2. Skew normal models

Begin with $k+1$ independent identically distributed (i.i.d.) standard normal random variables $W_1, W_2, \dots, W_k, U$. Consider the conditional density of $\mathbf{W}' = (W_1, W_2, \dots, W_k)$ given that for some λ_0 and $\boldsymbol{\lambda}_1$, $\lambda_0 + \boldsymbol{\lambda}_1'\mathbf{W} > U$. Here $\lambda_0 \in \mathbb{R}$ and $\boldsymbol{\lambda}_1 \in \mathbb{R}^k$. If we define the event A by

$$A = \{\lambda_0 + \boldsymbol{\lambda}_1'\boldsymbol{W} > U\} \tag{2}$$

then elementary computations yield

$$f_{\mathbf{W}|A}(\boldsymbol{w}) = [\prod_{i=1}^{k} \varphi(w_i)]\Phi(\lambda_0 + \boldsymbol{\lambda}_1'\boldsymbol{w})/P(A)\ . \tag{3}$$

It is easy in this case to evaluate $P(A)$ since $U - \boldsymbol{\lambda}_1'\boldsymbol{W} \sim N(0, 1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1)$. Thus

$$P(A) = P(U - \boldsymbol{\lambda}_1'\boldsymbol{W} < \lambda_0) = \Phi\left(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}}\right)\ . \tag{4}$$

Thus our family of k-dimensional densities skewed by hidden truncation takes the form

$$f(\boldsymbol{w}; \lambda_0, \boldsymbol{\lambda}_1) = [\prod_{i=1}^{k} \varphi(w_i)]\Phi(\lambda_0 + \boldsymbol{\lambda}_1'\boldsymbol{w})/\Phi\left(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}}\right)\ . \tag{5}$$

A few representative examples of densities (5) are displayed in Figure 1.

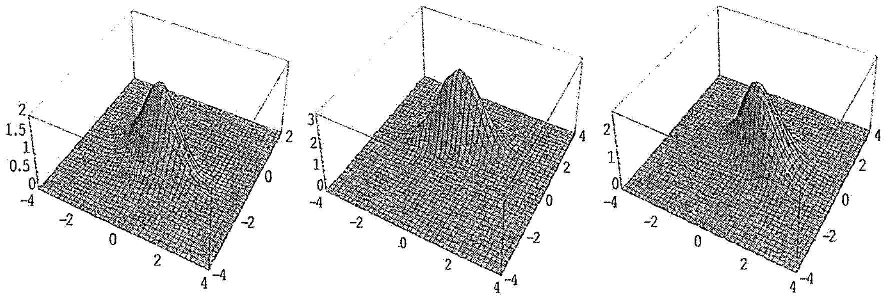

Figure 1. Skewed bivariate normal distributions via hidden truncation in (5). (a) $\lambda_0 = -2, \lambda_{12} = 3$; (b) $\lambda_0 = -2, \lambda_{11} = -1, \lambda_{12} = -3$; (c) $\lambda_0 = 1, \lambda_{11} = 2, \lambda_{12} = 3$.

The case in which $\lambda_0 = 0$ in (5) corresponds to the Azzalini and Dalla Valle (1996) skewed k-variate normal density. The more general threshold situation, i.e. the case in which λ_0 can be non-zero, was discussed in some detail in Arnold and Beaver (2000a).

The full k-dimensional skew normal model is obtained by introducing location and scale changes in (5). Thus we consider a random vector $\mathbf{X}$ which admits the representation

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{W} \tag{6}$$

where $\boldsymbol{\mu} \in \mathbb{R}^k, \Sigma^{1/2}$ is positive definite and $\boldsymbol{W}$ has density (5). It may be verified by elementary computations (see e.g. Arnold and Beaver (2002a)) that if $\boldsymbol{X}$ has the representation (6) then all of its marginal densities and all of its conditional densities are of the same type.

There are other routes available for arriving at the model (6). In the normal case, "all roads lead to Rome"; in the sense that all of the different modelling scenarios lead to the same class of distributions represented by (6). In non-normal cases the different modelling scenarios lead to interesting but distinct models as we shall see.

The second genesis for our hidden truncation model (6) begins with $k+1$ random variables $(Z_1, Z_2, \ldots, Z_k, V)$ which have a $(k+1)$-dimensional multivariate normal joint density. Now consider the joint density of $\boldsymbol{Z}$ given $V > v_0$. This clear hidden truncation construction will again lead to our model (6).

A novel third scenario involves an additive component and it is not so transparently obvious that we are again led to the model (6). For it, we begin with $X_0, X_1, \ldots, X_k$ i.i.d. $N(0,1)$ random variables. For an arbitrary $c \in \mathbb{R}$, we define $X_0(c)$ to be X_0 truncated below at c. Next define a random vector $\boldsymbol{Y}$ by:

$$Y_j = \delta_j X_0(c) + \sqrt{1-\delta_j^2}X_j \quad |\delta_j| < 1, \ j = 1, 2, \ldots, k \ . \tag{7}$$

(The special case of this model with $c = 0$ is equivalent to the model described in Azzalini (1986) since $X_0(0) \ \stackrel{d}{=} |X_0|$, where $\stackrel{d}{=}$ denotes equal in distribution.) It is readily verified that the moment generating function of $X_0(c)$ is given by

$$M_{X_0(c)}(t) = e^{t^2/2}\bar{\Phi}(c-t)/\bar{\Phi}(c), \tag{8}$$

where here and henceforth, for any distribution function F, we denote the corresponding survival function $1-F$ by $\bar{F}$. Consequently, since $M_{X_j}(t) = e^{t^2/2}, \quad j = 1, 2, \ldots, k$, we can write the joint moment generating function of

$\boldsymbol{Y}$ as

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= E\left(e^{\sum_{j=1}^{k} t_j(\delta_j X_0(c)+\sqrt{1-\delta_j^2}X_j)}\right) \\
&= M_{X_0(c)}(\sum_{j=1}^{k} t_j\delta_j)\prod_{j=1}^{k} M_{X_j}(t_j\sqrt{1-\delta_j^2}) \\
&= \exp\left[(\sum_{j=1}^{k} t_j\delta_j)^2 + \sum_{j=1}^{k} t_j^2(1-\delta_j^2)\right]\bar{\Phi}(c-\sum_{j=1}^{k} t_j\delta_j)/\bar{\Phi}(c) \\
&= \exp[\mathbf{t}'\mathbf{Q}\mathbf{t}]\Phi(-c+\sum_{j=1}^{k} t_j\delta_j)/\Phi(-c)\ . \qquad (9)
\end{aligned}$$

The elements of the matrix $\mathbf{Q}$ in (9) are given by

$$\begin{aligned}
q_{ii} &= 1, \qquad i = 1, 2, \ldots, k, \\
q_{ij} &= \delta_i\delta_j, \qquad i \neq j\ . \qquad (10)
\end{aligned}$$

It is then evident that a linear transformation will lead to a random variable $\mathbf{W}$ with joint density (5) and joint moment generating function

$$M_{\mathbf{W}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{t}/2}\Phi\left(\frac{\lambda_0+\boldsymbol{\lambda}_1'\mathbf{t}}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}}\right)/\Phi\left(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}}\right)\ . \qquad (11)$$

Consequently our additive component model (7) again eventually leads to the full family of k-variate skew normal variables defined in (6).

It may have become apparent to the reader by now that the skewing or hidden truncation can actually be applied to just one of the coordinate variables prior to making linear transformations to arrive at the family (6). This approach was apparently first explicitly stated by Jones (2002). (For earlier discussion see Azzalini and Capitanio (1999).) We sketch the development in the following paragraph.

Begin with k independent random variables $S_1, S_2, \ldots, S_k$ where S_1 has a univariate skew normal density of the form

$$f_{S_1}(s_1) = \varphi(s_1)\Phi(\nu_0+\nu_1 s_1)/\Phi(\frac{\nu_0}{\sqrt{1+\nu_1^2}}) \qquad (12)$$

and where $S_2, \ldots, S_k$ are standard normal random variables. Now consider the family of random variables of the form

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{S}\ . \qquad (13)$$

It is not difficult to verify that the family (13) coincides with the family (6) (though the parameters $\boldsymbol{\mu}$ and $\Sigma^{1/2}$ appearing in the two expressions are different but related). For completeness we will present the joint moment generating function of random variables of the form (6):

$$M_{\mathbf{X}}(\mathbf{t}) = \exp[\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}]\Phi(\frac{\lambda_0 + (\Sigma^{1/2}\boldsymbol{\lambda})'\mathbf{t}}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}})/\Phi(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}}) . \tag{14}$$

We thus have enumerated four alternative routes for arriving at the model (6) using normal components. It is perhaps surprising that all four routes lead eventually to the same model. As we shall see in the next section, this phenomenon is not encountered when we consider non-normal variants of the four constructions.

3. Non-normal skewed multivariate models

In our first (hidden truncation) development of a k-variate skew normal model we began with $k+1$ independent standard normal variables $W_1, W_2, \ldots, W_k, U$. We now consider the consequences of allowing these basic variables to have other distributions.

We thus now suppose that $W_1, W_2, \ldots, W_k$ and U are independent random variables with densities given by $\psi_1, \psi_2, \ldots, \psi_k$ and ψ_0, and distribution functions $\Psi_1, \Psi_2, \ldots, \Psi_k$ and Ψ_0 respectively. Again consider the joint density of $\mathbf{W}$ given $A = \{\lambda_0 + \boldsymbol{\lambda}_1'\mathbf{W} > U\}$ where $\lambda_0 \in \mathbb{R}$ and $\boldsymbol{\lambda}_1 \in \mathbb{R}^k$. We may verify that

$$f_{\mathbf{W}|A}(\mathbf{w}) = [\prod_{j=1}^{k} \phi_j(w_j)]\Phi_0(\lambda_0 + \boldsymbol{\lambda}_1'\boldsymbol{w})/P(A) . \tag{15}$$

The quantity $P(A)$ appearing in (15) will typically be difficult to evaluate. One case in which it can be easily evaluated is that in which all the joint densities $\phi_0, \phi_1, \ldots, \phi_k$ are symmetric about zero and $\lambda_0 = 0$. In that case $P(A) = 1/2$. The other scenario which will allow relatively easy computation of $P(A)$ is one in which the density of $\boldsymbol{\lambda}_1'\mathbf{W} - U$ is known and tractable. This would occur if all the ϕ_j's correspond to (possibly different) Cauchy densities. Such skew-Cauchy densities are discussed in some detail in Arnold and Beaver (2000c). The usual transformation

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{W} \tag{16}$$

can be used to extend the model (15) to enhance its flexibility. Sample graphs of densities corresponding to (15) (via hidden truncation) are given in Arnold and Beaver (2002b).

The second approach used to arrive at (6) began with $k+1$ multivariate normal random variables (not necessarily uncorrelated). We then considered the conditional distribution of k of these variables given that the other variable exceeded a threshold value. The key reason why this led to the same model (6) is that a general $(k+1)$-variate normal random vector is just a linear function of $k+1$ independent normal variables. In our more general non-normal setting we would begin with $(Z_1, Z_2, \ldots, Z_k, V)$ having an arbitrary $(k+1)$-dimensional joint distribution and consider the conditional distribution of $\mathbf{Z}$ given $\{V > v_0\}$. This will lead to models distinct from (15) (extended by transformations of the form (16)), unless the random vector $(Z_1, \ldots, Z_k, V)$ admits a representation in terms of linear functions of $k+1$ independent variables (as it does in the classical normal case). In more general cases the conditional density of $\mathbf{Z}$ given $\{V > v_0\}$ will depend in a complicated way upon the conditional distribution of V given $\mathbf{Z}$. Only very special cases can be expected to lead to tractable models.

Let us turn now to our third route (the additive component route) to the model (6). In our more general setting we will begin with $k+1$ independent random variables $Y_0, Y_1, \ldots, Y_k$ with corresponding densities $\psi_0, \psi_1, \ldots, \psi_k$. As in the normal case we will consider $Y_0(c)$ defined to be the random variable Y_0 truncated below at c. We then define the k dimensional random vector $\mathbf{Z}$ by

$$Z_j = Y_0(c) + \tau_j Y_j, \quad j = 1, 2, \ldots, k\ . \tag{17}$$

Since the density of $Y_0(c)$ is given by

$$f_{Y_0(c)}(y_0) = \psi_0(y_0) I(y_0 > c)/\bar{\Psi}_0(c)$$

where Ψ_0 is the distribution function corresponding to ψ_0, we can write the joint density of $\underline{Z}$ in the form

$$f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\tau}) = \int_c^\infty [\prod_{j=1}^k \frac{1}{\tau_j} \psi_j(\frac{z_j - y_0}{\tau_j})] \psi_0(y_0) dy_0 / \bar{\Psi}_0(c)\ . \tag{18}$$

Typically the integration in (18) will be difficult to perform. It can be done when the ψ_j's are normal, Cauchy, Laplace and logistic densities. In general, models obtained from (18) will be distinct from those obtained from (15). Some specific examples of such distributions skewed by an additive component (when $k = 1, 2$) can be found in Arnold and Beaver (2002b).

What if we consider non-normal variants of our fourth construction? For it we consider $k+1$ independent random variables $U_0, U_1, \ldots, U_k$ with common densities $\psi_0, \psi_1, \ldots, \psi_k$ and distribution functions $\Psi_0, \Psi_1, \ldots, \Psi_k$. We may then define

$$W_1 = U_1 + U_0(c) \tag{19}$$

where $U_0(c)$ denotes, as usual, U_0 truncated below at c, and for $i = 2, 3, \ldots, k$, define $W_i = U_i$. Alternatively we could define W_1 to be a random variable whose distribution is that of U_1 conditional on the event $\lambda_0 + \lambda_1 U_1 > U_0$. Finally we define

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{W}$$

to complete the construction. It is only in the case in which all ψ_i's are normal that we can expect these models to coincide with those derived using the previous 3 constructions.

An interesting alternative construction involving skewing only one of the coordinate variables was proposed by Jones (2002). He begins with a rather arbitrary joint distribution for $(W_1, W_2, \dots, W_k)$ and considers a new distribution obtained by replacing the marginal density of W_1 by a skewed version and retaining the original conditional structure of $W_2, \dots, W_k$ given W_1. This construction leads to old friends in the case in which $\mathbf{W}$ has a classical multivariate normal density but produces new models in other settings.

4. The Balakrishnan Extension

Motivated by order statistics concepts, Balakrishnan (2002) suggested that the Azzalini skew-normal density (1) can be extended to comprise the class of densities

$$f(x; \lambda, \alpha) \propto \varphi(x)[\Phi(\lambda x)]^{\alpha} \tag{20}$$

where $\lambda \in \mathbb{R}$ and $\alpha \in \mathbb{R}^+$. Analytic expressions for the normalizing constant needed in (20) are only available when $\alpha = 1, 2$ or 3. Nevertheless the family admits easy simulation (via a rejection algorithm, for example) and parameter estimation techniques, including maximum likelihood, can be implemented numerically. This idea can be also used to extend our hidden truncation k-variate normal model (5) or, for that matter, it can also be used to extend the non-normal model (15). The extended form of the density (15) will be:

$$f_{\mathbf{W}}(\mathbf{w}) \propto [\prod_{j=1}^{k} \phi_j(w_j)][\Phi_0(\lambda_0 + \boldsymbol{\lambda}_1' \mathbf{w})]^{\alpha} \tag{21}$$

where $\lambda_0 \in \mathbb{R}, \boldsymbol{\lambda}_1 \in \mathbb{R}^k$ and $\alpha \in \mathbb{R}^+$.

5. Multiple constraints and multiple additive components

Multiple hidden truncation can be envisioned. We also can envision more complicated additive component structures. Without going into details we merely describe how such models might be formulated. In most cases, they will involve a surfeit of parameters and this can be expected to limit their practical utility.

For a multiple constraint model we envision only retaining observations if linear combinations of the variables exceed certain random thresholds. The resulting k-dimensional density is of the form

$$f(\mathbf{w}) \propto [\prod_{i=1}^{k} \psi_i(w_i)][\prod_{j=1}^{\ell} \Psi_j(\lambda_0^{(j)} + \boldsymbol{\lambda}_1'^{(j)}\mathbf{w})] \tag{22}$$

where $\psi_1, \ldots, \psi_k$ are densities and $\Psi_1 \ldots \Psi_\ell$ are distribution functions. A Balakrishnan extension of such densities would take the form

$$f(\mathbf{w}) \propto [\prod_{i=1}^{k} \psi_i(w_i)][\prod_{j=1}^{\ell} \Psi_j^{\alpha_j}(\lambda_0^{(j)} + \boldsymbol{\lambda}_1'^{(j)}\mathbf{w})]\ . \tag{23}$$

The normalizing constants for these densities will usually be analytically intractable. An exception occurs if all densities and distributions are normal and if the skewness vectors, the $\boldsymbol{\lambda}_1^{(j)}$'s, are mutually orthogonal.

We will illustrate the concept of a multiple additive component model by considering a 3-dimensional example. For it we begin with 7 independent random variables $V_1, V_2, V_3, V_4, V_5, V_6, V_7$ with corresponding densities $\psi_1, \psi_2, \ldots, \psi_7$. For variables $V_4, \ldots, V_7$, we consider truncated versions $W_i = V_i(c_i), i = 4, 5, 6, 7$ (W_i is V_i truncated below at c_i). Now define the 3-dimensional random vector (X_1, X_2, X_3) by

$$\begin{array}{lllllllll} X_1 = & V_1 & + & W_1 & & & + & W_3 & + & W_4 \\ X_2 = & V_2 & + & W_1 & + & W_2 & & & + & W_4 \\ X_3 = & V_3 & & & + & W_2 & + & W_3 & + & W_4\ . \end{array} \tag{24}$$

Observe that W_1 contributes only to X_1 and X_2, W_2 contributes to X_2 and X_3, etc. while W_4 contributes to all X_i's. The analysis of models such as (24) will be easiest when all the component densities (the ψ_i's) are normal.

References

Arnold, B. C. and Beaver, R. J. (2000a). Hidden truncation models. *Sankhya* **62**, 23–25.

Arnold, B. C. and Beaver, R. J. (2000b). Some skewed multivariate distributions. *Amer. J. Math. Management Sci.* **20**, 27–38.

Arnold, B. C. and Beaver, R. J. (2000c). The skew-Cauchy distribution. *Statist. Probab. Lett.* **49**, 285–290.

Arnold, B. C. and Beaver, R. J. (2002a). Skewed multivariate models related to hidden truncation. *Test* **11**, 7–54.

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* **XLVI**, 201–208.

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

Balakrishnan, N. (2002). Discussion of: Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 37–39.

Jones, M. C. (2002). Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. *J. Multivariate Anal.* **81**, 85–99.

UNIVERSITY OF CALIFORNIA, RIVERSIDE, USA
E-mail address: barry.arnold@ucr.edu
E-mail address: robert.beaver@ucr.edu