

Multiple and Non-linear Constraint Models

Barry C. Arnold

Abstract Since the basic Azzalini (1985) skew-normal model can be viewed as a hidden truncation model with normal components and one linear constraint, it has been natural to extend the scope of the model by considering univariate and multivariate models with possibly non-normal components, involving single and multiple constraints that are not necessarily linear. Some general features of such models are discussed in this contribution.

Key words: skew-normal, additive components, hidden truncation, latent variable models.

1 Introduction

Begin with $Z_1, Z_2, \dots, Z_k$ and $W_1, W_2, \dots, W_m$, which are $k + m$ independent random variables, and use the notation $\mathbf{Z}$ and $\mathbf{W}$ to denote the corresponding k and m dimensional random vectors. Initially we allow the Z_i 's and W_j 's to have possibly different densities denoted by $\phi_1, \phi_2, \dots, \phi_k$ and $\psi_1, \psi_2, \dots, \psi_m$ with corresponding distribution functions denoted by appropriate upper case letters (Φ_i 's and Ψ_j 's). Now for a given set of m constraint functions $g_1(\mathbf{z}), g_2(\mathbf{z}), \dots, g_m(\mathbf{z})$ we consider the conditional density of $\mathbf{Z}$ given $\bigcap_{j=1}^m [W_j \leq g_j(\mathbf{Z})]$. This density will be of the form

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\prod_{i=1}^k \phi_i(z_i) \prod_{j=1}^m \Psi_j(g_j(\mathbf{z}))}{P\{\bigcap_{j=1}^m [W_j \leq g_j(\mathbf{Z})]\}} \quad (1)$$

More general models will allow $\mathbf{Z}$ to have dependent components and also allow $\mathbf{W}$ to have dependent components, but we will not pursue this initially.

Barry C. Arnold
University of California, Riverside CA, USA, e-mail: barry.arnold@ucr.edu

It is natural initially to consider linear functions, as is done in many univariate and multivariate variations of the basic Azzalini (1985) univariate skew normal theme. The model (1) with linear constraint functions will assume the form:

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\prod_{i=1}^k \phi_i(z_i) \prod_{j=1}^m \Psi_j(\alpha_j + \beta(j)' \mathbf{z})}{P\{\cap_{j=1}^m [W_j \leq \alpha_j + \beta(j)' \mathbf{Z}]\}} \quad (2)$$

Notation. ϕ and Φ denote the standard normal d.f. and c.d.f. respectively.

The functions $\{g_j\}_{j=1}^m$ appearing in (1) will typically not be completely known. They will usually be assumed to be known except for some unknown parameters which appear in their definitions. Taking this into account, our model should be written in the form

$$f_{\mathbf{Z}}(\mathbf{z}; \lambda^{(1)}, \dots, \lambda^{(m)}) = \frac{\prod_{i=1}^k \phi_i(z_i) \prod_{j=1}^m \Psi_j(g_j(\mathbf{z}); \lambda^{(j)})}{P\{\cap_{j=1}^m [W_j \leq g_j(\mathbf{Z}; \lambda^{(j)})]\}}. \quad (3)$$

If the dimension, k , and/or the number of constraints, m , is large, then the number of parameters in the model is evidently excessive, and simplified sub-models will necessarily need to be considered, unless there are compelling stochastic modeling arguments which require retaining the full panoply of parameters.

As remarked by Arnold and Beaver (2002), it is only under very special circumstances (e.g., normal component densities and distributions and linear constraints) that the normalizing constant appearing in the denominator of (3) can be evaluated in reasonably simple form. In a particular case, (2) with normal components and orthogonal $\alpha^{(j)}$'s the density can be expressed in the form

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\prod_{i=1}^k \phi(z_i) \prod_{j=1}^m \Phi(\alpha_j + \beta(j)' \mathbf{z})}{\prod_{j=1}^m \Phi\left(\frac{\alpha_j}{\sqrt{1 + \beta(j)' \beta(j)}}\right)}. \quad (4)$$

This expression simplifies gratifyingly if $\alpha_j = 0$, $j = 1, 2, \dots, m$, reducing to

$$f_{\mathbf{Z}}(\mathbf{z}) = 2^m \prod_{i=1}^k \phi(z_i) \prod_{j=1}^m \Phi(\beta(j)' \mathbf{z}). \quad (5)$$

When $m = 1$, this is recognizable as the Azzalini-Dalla Valle (1996) single constraint multivariate skew-normal model. See Section 4 for discussion of more general models involving multiple linear constraints.

In practice it is reasonable to expect that the number of constraints in the models (3) will be small, perhaps one, two or three. The resulting models are still remarkably flexible. Specific choices for the densities $\{\phi_i\}$ and $\{\psi_j\}$ and specific parametric forms for the functions $\{g_j\}$ merit separate investigation. Typically the g_j 's might be taken to be polynomial functions of the z_i 's, but other choices have ap-

peared in the literature. For example, with $k = m = 1$ and with normal component densities (i.e., $\phi_i = \psi_i = \varphi$), Arellano-Valle, Gomez and Quintana (2004) consider the non-linear choice

$$g(z; \alpha, \beta) = \frac{\alpha z}{\sqrt{1 + \beta z^2}}. \quad (6)$$

Kim (2005), also in the normal setting, used a piecewise linear constraint function

$$g(z; \alpha) = \alpha|z|, \quad (7)$$

in the construction of his two-piece skew-normal density.

With regard to analyses involving multiple constraints that are not all linear, the surface has only been scratched. Some examples will be reviewed in Section 3. But, first we digress to discuss different manners of arriving at models involving constraints on unobserved covariates.

2 More than one road leads to Rome

A well-known additive component representation is available for the basic one dimensional skew normal model. This basic model, a slight extension of the Azzalini (1985) model, can be constructed by beginning with a two dimensional normal random vector

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N^{(2)} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

and considering the distribution of X to be that of Z_1 given that $Z_2 \leq c$. Elementary computations then yield

$$f_X(x) = \frac{\varphi(x)P\{Z_2 \leq c|Z_1 = x\}}{P\{Z_2 \leq c\}} = \varphi(x) \frac{\Phi\left(\frac{c - \rho x}{\sqrt{1 - \rho^2}}\right)}{\Phi(c)} = \varphi(x) \frac{\Phi(\lambda_0 + \lambda_1 x)}{\Phi\left(\frac{\lambda_0}{\sqrt{1 + \lambda_1^2}}\right)}, \quad (8)$$

after reparameterization.

If we use the notation $Z(c)$ to denote a random variable with a standard normal distribution truncated above at c , then its moment generating function (m.g.f.) is given by

$$M_{Z(c)}(t) = e^{t^2/2} \Phi(c - t) / \Phi(c). \quad (9)$$

Now assume that Z_1, Z_2 are i.i.d. standard normal random variables. It follows that, for $\delta \in (-1, 1)$, the random variable

$$X = \sqrt{1 - \delta^2} Z_1 + \delta Z_2(c) \quad (10)$$

has a m.g.f. of the form

$$M_X(t) = e^{t^2/2} \frac{\Phi(c - \delta t)}{\Phi(c)} \quad (11)$$

which can be verified to be the m.g.f. of the skew-normal density (8), where $\lambda_0 = -c/\sqrt{1-\delta^2}$ and $\lambda_1 = \delta/\sqrt{1-\delta^2}$. Thus confirming that an additive component representation of the form (10) is available for skew-normal densities given by (8). [This observation, when $c = 0$, was first made by Henze (1986).] So either formulation, a hidden truncation formulation (Z_1 given $Z_2 \leq c$) or an additive component formulation (10) leads to the same model. One might conjecture that this phenomenon is only encountered when the Z_i 's are normally distributed. See Arnold and Gómez (2009) for related discussion.

In the normal scenario, we may ask if a similar phenomenon occurs with a more general form of truncation. Remarkably, the answer is yes.

To verify this claim, it is convenient to introduce notation for a standard normal measure defined on the class $\mathcal{B}$ of Borel subsets of the real line, $\mathcal{R}$. With little risk of confusion we will use the same letter Φ to denote this measure as is used to denote the standard normal distribution function. If the argument of Φ is a set then Φ is the measure, if the argument is a real number then Φ is the distribution function. The measure Φ is defined by

$$\Phi(B) = \int_B \varphi(x) dx, \quad B \in \mathcal{B}. \quad (12)$$

With this notation, observe that $\Phi(x) = \Phi((-\infty, x])$ for each $x \in \mathcal{R}$.

Another notational convention will prove to be convenient in the subsequent discussion. For any $B \in \mathcal{B}$, any $c \in \mathcal{R}$, and any $d \in \mathcal{R}_+$, define the set $d(B+c)$ to be

$$d(B+c) = \{y : y \in \mathcal{R}, \frac{y-c}{d} \in B\}. \quad (13)$$

Using these notational conventions we wish to consider hidden truncation to an arbitrary Borel set $A \subset \mathcal{R}$, instead of just a one or two sided interval. We begin, as in the interval case, with a bivariate normal model. i.e.,

$$Z_1 \sim N(0, 1)$$

and, for $z_1 \in \mathcal{R}$,

$$Z_2|Z_1 = z_1 \sim N(\rho z_1, 1 - \rho^2),$$

i.e., assume that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N^{(2)} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Thus Z_2 also has a standard normal distribution. For a given Borel set A , consider the conditional distribution of Z_1 given $Z_2 \in A$. It will be of the form

$$f_X(x) = \varphi(x) \frac{P\{Z_2 \in A | Z_1 = x\}}{P\{Z_2 \in A\}} = \varphi(x) \frac{\Phi\left(\frac{A - \rho x}{\sqrt{1 - \rho^2}}\right)}{\Phi(A)} \quad (14)$$

We turn to the question of whether an additive component representation of the density (14) is available, as it was in the case in which A was a one-sided interval. To resolve this question, begin with Z_1, Z_2 i.i.d. standard normal random variables and define $Z_2(A)$ to be the random variable Z_2 truncated to the set A , i.e., a random variable with distribution function

$$F_{Z_2(A)}(z) = \frac{\Phi(A \cap (-\infty, z])}{\Phi(A)}, \quad z \in \mathcal{R}. \quad (15)$$

Then, for $\delta \in (-1, 1)$, as before, define

$$X = \sqrt{1 - \delta^2} Z_1 + \delta Z_2(A). \quad (16)$$

To determine whether (14) and (16) are equivalent, consider their corresponding moment generating functions. The m.g.f. of $Z_2(A)$ is given by

$$M_{Z_2(A)}(t) = e^{t^2/2} \frac{\Phi(A - t)}{\Phi(A)}. \quad (17)$$

It then follows that the m.g.f. of X defined in (16) is of the form

$$M_X(t) = e^{t^2/2} \frac{\Phi(A - \delta t)}{\Phi(A)}. \quad (18)$$

Next consider the m.g.f. of a random variable Y with density (14).

$$\begin{aligned} M_Y(t) &= \int_{-\infty}^{\infty} e^{ty} \varphi(y) \frac{\Phi\left(\frac{A - \rho y}{\sqrt{1 - \rho^2}}\right)}{\Phi(A)} dy = e^{t^2/2} \int_{-\infty}^{\infty} \varphi(y - t) \frac{\Phi\left(\frac{A - \rho y}{\sqrt{1 - \rho^2}}\right)}{\Phi(A)} dy \\ &= e^{t^2/2} \int_{-\infty}^{\infty} \varphi(u) \frac{\Phi\left(\frac{A - \rho t - \rho u}{\sqrt{1 - \rho^2}}\right)}{\Phi(A)} dy. \end{aligned}$$

However

$$\begin{aligned} \Phi\left(\frac{A - \rho t - \rho u}{\sqrt{1 - \rho^2}}\right) &= \int_{-\infty}^{\infty} I(v \in \frac{A - \rho t - \rho u}{\sqrt{1 - \rho^2}}) \varphi(v) dv \\ &= \int_{-\infty}^{\infty} I(w \in A - \rho t) (1 - \rho^2)^{-1/2} \varphi\left(\frac{w - \rho u}{\sqrt{1 - \rho^2}}\right) dw \end{aligned}$$

Consequently

$$M_Y(t) = \frac{e^{t^2/2}}{\Phi(A)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(w \in A - \rho t) (1 - \rho^2)^{-1/2} \varphi(u) \varphi\left(\frac{w - \rho u}{\sqrt{1 - \rho^2}}\right) du dw$$

$$= \frac{e^{t^2/2}}{\Phi(A)} P\{Z_2 \in A - \rho t\} = e^{t^2/2} \frac{\Phi(A - \rho t)}{\Phi(A)} \quad (19)$$

Comparison of (18) and (19) confirms the fact that the equivalence of hidden truncation and additive component formulations continues to hold for general forms of truncation, i.e., truncation to an arbitrary Borel set A . In fact, as will be shown in Section 5, a multivariate version of this equivalence relationship can be documented.

3 Specific examples of non-linear constraint functions

In this section, we will focus on examples of model (1) in the univariate case, i.e., when $k = 1$. Initially only one constraint will be imposed; subsequently more constraints will be considered. A one-dimensional single constraint version of (1) will be of the form

$$f_Z(z) \propto \phi(z) \Psi(g(z)) \quad (20)$$

where ϕ is a density function, Ψ is the distribution function corresponding to the density ψ (which may or may not differ from ϕ), and g is a “constraint” function. The required normalizing constant in (20) will frequently be difficult to evaluate.

Most of the discussion of such models in the literature assume that $\phi = \psi = \varphi$, i.e., normal components. Arellano-Valle, Gómez and Quintana (2004) assume this and begin with a basic Azzalini skew-normal model

$$f(z; \lambda) = 2\varphi(z) \Phi(\lambda z). \quad (21)$$

By considering a mixture of such distributions obtained by assuming that λ has a normal distribution with mean α and variance β , they arrive at the model

$$f(z; \alpha, \beta) = 2\varphi(z) \Phi\left(\frac{\alpha z}{\sqrt{1 + \beta z^2}}\right), \quad (22)$$

corresponding to the non-linear constraint function $g(z) = \alpha z / \sqrt{1 + \beta z^2}$.

The key feature that permitted straightforward evaluation of the mixed distribution in this case, was the fact that the normalizing constant in (21), namely 2, does not depend on the parameter λ of the density.

Remark. It is of course possible to mimic the development of (22) using Cauchy components to obtain a skew-Cauchy model:

$$f(z; \alpha, \beta) = 2\psi_C(z) \Psi_C\left(\frac{\alpha z}{1 + \beta|z|}\right), \quad (23)$$

where ψ_C and Ψ_C denote the standard Cauchy density and distribution function.

Kim (2005), beginning with W_0 and W_1 being independent random variables with symmetric distributions F_0 and F_1 respectively, considered the distribution of W_0 given $W_1 \leq \alpha|W_0|$. In this manner one arrives at a symmetric density of the form:

$$f(z; \alpha) \propto f_0(z)F_1(\alpha|z|). \quad (24)$$

In the special case in which $F_0 = F_1 = \Phi$, the required normalizing constant in (24) may be evaluated. It is

$$c_\alpha = 2\pi/(\pi + 2 \arctan \alpha). \quad (25)$$

Kim's (2005) two-piece skew-normal (TN) density is thus of the form

$$f(z; \alpha) = c_\alpha \varphi(z) \Phi(\alpha|z|). \quad (26)$$

where c_α is as in (25).

Because the TN density (26) is symmetric, one can apply the basic Azzalini constraint to it and in this way arrive at a two constraint model of the form

$$f(z; \alpha, \beta) = 2c_\alpha \varphi(z) \Phi(\alpha|z|) \Phi(\beta z). \quad (27)$$

Arnold, Gómez and Salinas (2009) call this an extended two-piece skew-normal (ETN) distribution. Note that the ETN normalizing constant is no more complicated than the normalizing constant for the symmetric TN distribution. They demonstrate the flexibility of this family of distributions and provide an example of a data set which is better fitted by the ETN model (27), with added location and scale parameters, than by competing location and scale models involving basic distributions of the skew-normal form (21), the two piece skew-normal (TN) form (26) and the extended skew-normal form (22). A key feature of the ETN family of distributions is that it includes distributions that have densities that are symmetric, asymmetric, unimodal and bimodal.

If one is willing to deal with an analytically intractable normalizing constant, the model (27) might be extended to the form

$$f(z; \alpha_0, \alpha_1, \beta_0, \beta_1) \propto \psi_1(z) \Psi_2(\alpha_0 + \alpha_1|z|) \Psi_3(\beta_0 + \beta_1 z). \quad (28)$$

Models in which the densities and distributions in (28) are all standard Cauchy or all standard logistic might be considered to be suitable competitors for the corresponding normal component model.

In applications, there may be features of the data generating process that will suggest appropriate choices for the constraints, the $g_j(z)$'s, in single or multiple constraint models. Absent such information, it is to be hoped that choices of low order polynomials in z and $|z|$ for the constraint functions will provide adequate modeling flexibility. Multiple linear constraints as in, for example, (4) have received considerable attention, especially those involving normal component distributions.

Such models in both univariate and multivariate settings will be discussed in Sections 4 and 5. It will be seen that many of the univariate results, even those involving arbitrary truncation can be extended quite readily to multivariate normal scenarios.

Remark: Hidden truncation in pseudo-normal models. A quite general approach to the construction of normal distributions skewed by hidden truncation involves bivariate distributions of the pseudo-normal form in the sense introduced by Filus and Filus (2007). We will say that a random vector $(X_1, X_2)'$ has a pseudo normal distribution if

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

and for each $x_1 \in \mathcal{R}$,

$$X_2|X_1 = x_1 \sim N(\mu_2(x_1), \sigma_2^2(x_1)),$$

for some functions $\mu_2(\cdot)$ and $\sigma_2^2(\cdot)$.

If we apply hidden truncation to a pseudo-normal random vector with $\mu_1 = 0$ and $\sigma_1^2 = 1$, i.e., if we consider the density of X_1 given $X_2 \leq c$, we obtain the model

$$f_Y(y) = \varphi(y) \frac{P\{X_2 \leq c | X_1 = y\}}{P\{X_2 \leq c\}} = \varphi(y) \frac{\Phi\left(\frac{c - \mu_2(y)}{\sigma_2(y)}\right)}{P\{X_2 \leq c\}}, \quad (29)$$

which can be recognized as being of the form (1) with $k = m = 1$, $\phi = \varphi$, $\Psi = \Phi$ and $g_1(y) = \frac{c - \mu_2(y)}{\sigma_2(y)}$.

More generally we could apply hidden truncation to an arbitrary Borel set A , to arrive at the model

$$f_Y(y) = \varphi(y) \frac{\Phi\left(\frac{A - \mu_2(y)}{\sigma_2(y)}\right)}{P\{X_2 \in A\}}, \quad (30)$$

Note that evaluation of $P\{X_2 \leq c\}$ or of $P\{X_2 \in A\}$, will often be problematic in the pseudo-normal setting.

4 Extensions to multiple constraint multivariate cases

Notation. If $\mathbf{X} \sim N^{(m)}(\mu, \Sigma)$ then its distribution function is denoted by $\Phi^{(m)}(\mathbf{x}; \mu, \Sigma)$ and its corresponding density by $\varphi^{(m)}(\mathbf{x}; \mu, \Sigma)$.

The basic skew-normal distribution (21) has been extended from one to k dimensions and the number of linear constraints has been increased to m in contributions by several authors. Most of the models can be subsumed by a model of the following type.

Begin with $\begin{pmatrix} \mathbf{Z} \\ \mathbf{W} \end{pmatrix}$, a $k + m$ -dimensional random vector, in which $\mathbf{Z}$ is of dimension k and $\mathbf{W}$ is of dimension m . Assume that

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{W} \end{pmatrix} \sim N^{(k+m)} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \right). \quad (31)$$

where $\mathbf{R}$ is a correlation matrix. Analogous to the development in the one dimensional case, we consider the conditional distribution of $\mathbf{Z}$ given $\mathbf{W} \leq \mathbf{a} + \mathbf{BZ}$. If $\mathbf{Y}$ has this distribution and if we introduce “location” and “scale” parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by defining

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{Y}, \quad (32)$$

then we obtain (after reparameterization) the general skew-normal model with m.g.f. given by

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \frac{\Phi^{(m)}(\boldsymbol{\gamma} + \boldsymbol{\Gamma}\mathbf{t}; \mathbf{0}, \mathbf{R})}{\Phi^{(m)}(\boldsymbol{\gamma}; \mathbf{0}, \mathbf{R})} \quad (33)$$

There are other ways of arriving at this model. An additive component development is possible involving truncated versions of the W_j 's. Discussion of this route will be subsumed by the later discussion of a quite general additive component approach. A hidden truncation formulation permits perhaps the most simply described version of the general linear constraint skew-normal model as follows. Begin with

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N^{(k+m)} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{v} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \quad (34)$$

and consider the conditional distribution of $\mathbf{X}$ given $\mathbf{Y} \leq \mathbf{c}$ for some $\mathbf{c} \in \mathcal{R}^m$. Without loss of generality one may assume that $\mathbf{v} = \mathbf{0}$ in (34) and that $\boldsymbol{\Sigma}_{22}$ is a correlation matrix.

The single constraint ($m = 1$) version of this distribution, (34), was introduced by Azzalini and Dalla Valle (1996). The multiple constraint version has a density of the form

$$f_{\mathbf{Z}}(\mathbf{z}) \propto \varphi((\mathbf{z} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})) \Phi^{(m)}(\mathbf{a} + \mathbf{Bz}; \mathbf{0}, \mathbf{R}). \quad (35)$$

Extensions of the model (35) are possible. Several authors have considered non-normal versions. Frequently, elliptically contoured distributions replace multivariate normal distributions in the model. Rather than move in that direction, we will focus in the following section on multivariate models with normal components but involving more general forms of truncation. We thus will consider multivariate versions of the models (14) and (16), and we will investigate, in the multivariate setting, whether the two modeling strategies, hidden truncation and additive components, again lead to the same class of distributions.

5 General hidden truncation and additive components in the multivariate case.

If $\mathbf{X} \sim N^{(m)}(\mu, \Sigma)$, we introduced the notations $\Phi^{(m)}(\mathbf{x}; \mu, \Sigma)$ and $\phi^{(m)}(\mathbf{x}; \mu, \Sigma)$ to denote the corresponding distribution and density functions. As in the one dimensional case we will define a related measure on the m -dimensional Borel sets by

$$\Phi^{(m)}(\mathbf{A}; \mu, \Sigma) = \int_{\mathbf{A}} \phi^{(m)}(\mathbf{x}; \mu, \Sigma) d\mathbf{x}, \quad \mathbf{A} \in \mathcal{B}^m. \quad (36)$$

This notation is parallel to that used in the univariate setting and here also there should be no confusion between the distribution function and the measure, even though the notation is almost identical. Again, if the argument is point in $\mathcal{R}^m$ then we are dealing with the distribution function, while if the argument is a Borel set, then we are dealing with the measure. In addition we will define for any $B \in \mathcal{B}^m$ and any $\mathbf{c} \in \mathcal{R}^m$, the set $B + \mathbf{c}$ to consist of all points in $\mathcal{R}^m$ which are equal to $\mathbf{b} + \mathbf{c}$ for some $\mathbf{b} \in B$.

We will consider truncation to a specific but arbitrary set $A \in \mathcal{B}^m$. First we discuss a hidden truncation model. Begin with

$$\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim N^{(k+m)} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} \right). \quad (37)$$

here

$$R = \begin{pmatrix} \mathbf{I} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}$$

is a correlation matrix, with upper left corner submatrix equal to the identity matrix.

Our hidden truncation model corresponds to the conditional distribution of $\mathbf{Z}_1$ given $\mathbf{Z}_2 \in A$. Thus it is the density

$$f_{\mathbf{Z}_1}(\mathbf{z} | \mathbf{Z}_2 \in A) = \left[\prod_{i=1}^k \phi(z_i) \right] \frac{P(\mathbf{Z}_2 \in A | \mathbf{z}_1 = \mathbf{z})}{P(\mathbf{Z}_2 \in A)}.$$

However from (37), we know that

$$\mathbf{Z}_2 \sim N^{(m)}(\mathbf{0}, \mathbf{R}_{22})$$

and that

$$\mathbf{Z}_2 | \mathbf{Z}_1 = \mathbf{z} \sim N^{(m)}(\mathbf{R}_{21}\mathbf{z}, \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{12}).$$

Note that $\mathbf{R}_{21}$ is the transpose of $\mathbf{R}_{12}$.

Consequently

$$f_{\mathbf{Z}_1}(\mathbf{z} | \mathbf{Z}_2 \in A) = \left[\prod_{i=1}^k \phi(z_i) \right] \frac{\Phi^{(m)}(A - \mathbf{R}_{21}\mathbf{z}; \mathbf{0}, \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{12})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R}_{22})}. \quad (38)$$

In order to easily compare this hidden truncation model with a corresponding additive component model, it will be convenient to evaluate the m.g.f. of this distribution. If $\mathbf{Y}$ has the density (38), then its corresponding m.g.f. may be evaluated as follows.

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= \int_{\mathcal{R}^k} e^{\mathbf{t}'\mathbf{y}} \frac{e^{-\mathbf{y}'\mathbf{y}/2}}{(2\pi)^{k/2}} \frac{\Phi^{(m)}(A - \mathbf{R}_{21}\mathbf{y}; \mathbf{0}, \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{12})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R}_{22})} d\mathbf{y} \\
&= e^{\mathbf{t}'\mathbf{t}/2} \int_{\mathcal{R}^k} \frac{e^{-(\mathbf{y}-\mathbf{t})'(\mathbf{y}-\mathbf{t})/2}}{(2\pi)^{k/2}} \frac{\Phi^{(m)}(A - \mathbf{R}_{21}\mathbf{y}; \mathbf{0}, \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{12})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R}_{22})} d\mathbf{y} \\
&= e^{\mathbf{t}'\mathbf{t}/2} \int_{\mathcal{R}^k} \frac{e^{-\mathbf{u}'\mathbf{u}/2}}{(2\pi)^{k/2}} \frac{\Phi^{(m)}(A - \mathbf{R}_{21}(\mathbf{u} + \mathbf{t}); \mathbf{0}, \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{12})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R}_{22})} d\mathbf{u} \\
&= e^{\mathbf{t}'\mathbf{t}/2} \frac{\Phi^{(m)}(A - \mathbf{R}_{21}\mathbf{t}; \mathbf{0}, \mathbf{R}_{22})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R}_{22})}. \tag{39}
\end{aligned}$$

From the discussion in the one-dimensional case, recall that the additive component construction involves a sum of two independent random variables, one of which is truncated. To extend this to k dimensions we begin by taking two independent random vectors

$$\mathbf{Z}_1 \sim N^{(k)}(\mathbf{0}, \mathbf{I})$$

and

$$\mathbf{Z}_2 \sim N^{(m)}(\mathbf{0}, \mathbf{R}),$$

where $\mathbf{R}$ is an $m \times m$ correlation matrix.

For a fixed but arbitrary set $A \in \mathcal{B}^m$, we define $\mathbf{Z}_2(A)$ to be $\mathbf{Z}_2$ truncated to the set A . The m.g.f. of this truncated random vector may be evaluated as follows:

$$\begin{aligned}
M_{\mathbf{Z}_2(A)}(\mathbf{t}) &= \frac{\int_A e^{\mathbf{t}'\mathbf{z}} f_{\mathbf{Z}_2}(\mathbf{z}) d\mathbf{z}}{\int_A f_{\mathbf{Z}_2}(\mathbf{z}) d\mathbf{z}} \\
&= \frac{\int_A e^{\mathbf{t}'\mathbf{z}} |\mathbf{R}|^{-1/2} (2\pi)^{-m/2} e^{-\mathbf{z}'\mathbf{R}^{-1}\mathbf{z}/2} d\mathbf{z}}{\int_A |\mathbf{R}|^{-1/2} (2\pi)^{-m/2} e^{-\mathbf{z}'\mathbf{R}^{-1}\mathbf{z}/2} d\mathbf{z}} \\
&= e^{\mathbf{t}'\mathbf{R}\mathbf{t}/2} \frac{\int_A |\mathbf{R}|^{-1/2} (2\pi)^{-m/2} e^{-(\mathbf{z}-\mathbf{R}\mathbf{t})'\mathbf{R}^{-1}(\mathbf{z}-\mathbf{R}\mathbf{t})/2} d\mathbf{z}}{\int_A |\mathbf{R}|^{-1/2} (2\pi)^{-m/2} e^{-\mathbf{z}'\mathbf{R}^{-1}\mathbf{z}/2} d\mathbf{z}}.
\end{aligned}$$

Consequently

$$M_{\mathbf{Z}_2(A)}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{R}\mathbf{t}/2} \frac{\Phi(A - \mathbf{R}\mathbf{t}; \mathbf{0}, \mathbf{R})}{\Phi(A; \mathbf{0}, \mathbf{R})} \tag{40}$$

Now for some $k \times m$ matrix $\mathbf{D}$, define an “additive components” random variable $\mathbf{X}$ by

$$\mathbf{X} = (\mathbf{I} - \mathbf{D}\mathbf{R}\mathbf{D}')^{1/2} \mathbf{Z}_1 + \mathbf{D}\mathbf{Z}_2(A). \tag{41}$$

Compare this with the univariate case (16). The m.g.f. of $\mathbf{X}$ may then be obtained using (40). Thus

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{t}/2} \frac{\Phi^{(m)}(A - \mathbf{R}\mathbf{D}'\mathbf{t}; \mathbf{0}, \mathbf{R})}{\Phi^{(m)}(A; \mathbf{0}, \mathbf{R})}. \quad (42)$$

A comparison of this m.g.f. with (39), confirms that they will be identical provided we set $\mathbf{R} = \mathbf{R}_{22}$ and $\mathbf{D}' = \mathbf{R}_{22}^{-1}\mathbf{R}_{21}$. This confirms that even in this multivariate setting with arbitrary truncation, the hidden truncation and the additive component models coincide.

6 Envoi.

Even 25 years after the appearance of Azzalini's (1985) seminal discussion of the basic skew-normal density, there still remain a remarkably diverse spectrum of extensions and variations which merit further investigation. Models with normal components and single or multiple linear constraints have been quite thoroughly studied. However, outside of that specialized arena, much work remains to be done. Models are often only briefly described and little related distributional and/or statistical analysis has been provided. I hope that the rapid tour of variations on the Azzalini theme provided in this paper will encourage further study of these models.

References

- Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004) A new class of skew normal distributions. *Comm. Stat: Theory and Methods*, **33**, 1465-1480.
- Arnold, B.C. and Beaver, R.J. (2002) Some multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-35.
- Arnold, B.C. and Gómez, H.W. (2009) Hidden truncation and additive components: Two alternative skewing paradigms. *Calcutta Statistical Association Bulletin*, **61**, 17-31.
- Arnold, B.C., Gómez, H.W. and Salinas, H.S. (2009) On multiple constraint skewed models. *Statistics*, **43**, 279-293.
- Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.
- Azzalini, A. and Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.
- Filus, J.K. and Filus, L.Z. (2007) On new multivariate probability distributions and stochastic processes with systems reliability and maintenance applications. *Methodology and Comp. in Applied Prob.*, **9**, 426-446.
- Henze, N. (1986) A probabilistic representation of the "skew-normal" distribution. *Scand. J. Statist.*, **13**, 271-275.
- Kim, H.J. (2005) On a class of two-piece skew normal distributions. *Statistics*, **39**, 537-553.