Journal of Statistical Planning and Inference 139 (2009) 3741 – 3749

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Flexible univariate and multivariate models based on hidden truncation

Barry C. Arnold*

University of California, Riverside, USA

ARTICLE INFO

Available online 22 May 2009

MSC:
60E05
62E10

Keywords:
Skew-normal distribution
Conditional specification
Weighted distribution
Multivariate normal
Normal conditionals
Exponential conditionals
Pareto distribution

ABSTRACT

A broad spectrum of flexible univariate and multivariate models can be constructed by using a hidden truncation paradigm. Such models can be viewed as being characterized by a basic marginal density, a family of conditional densities and a specified hidden truncation point, or points. The resulting class of distributions includes the basic marginal density as a special case (or as a limiting case), but also includes an array of models that may unexpectedly include many well known densities. Most of the well known skew-normal models (developed from the seed distribution popularized by Azzalini [(1985). A class of distributions which includes the normal ones. Scand. J. Statist. 12(2), 171–178]) can be viewed as being products of such a hidden truncation construction. However, the many hidden truncation models with non-normal component densities undoubtedly deserve further attention.

1. Introduction

The skew-normal model, popularized and studied by Azzalini (1985) and his coworkers, is a one parameter family of densities of the form

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x), \quad -\infty < x < \infty, \tag{1}$$

where ϕ (respectively, Φ) denotes the standard normal density (respectively, distribution) function and $\lambda \in \mathbb{R}$ is a parameter which controls skewness. The addition of a location parameter (μ) and a scale parameter (σ) yields the following flexible family, useful for fitting asymmetric data sets:

$$f(x;\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{x-\mu}{\sigma}\right)\right). \tag{2}$$

A good survey of results related to this skew-normal model may be found in Genton (2004).

A scenario which leads to data satisfying the model (2) is one involving hidden truncation of the following form (see Arnold and Beaver, 2000 for more details on hidden truncation models which are also known as frontier models in the Economics literature, for example, in Kumbhakar and Lovell, 2000). In this setting, we begin with a two dimensional random variable (X, Y) which has a classical bivariate normal distribution with mean vector (μ_X, μ_Y) and variance–covariance matrix Σ. Only the X coordinate random variable is observed and it is only observed if the concomitant Y exceeds μ_Y (i.e. if Y is above average). Data collected in this fashion are readily confirmed to have density (2).

A slight variant of this model involves observing X's only if their concomitant Y variable exceeds a given level y_0 (not necessarily equal to μ_Y). This extension was alluded to by Azzalini (1985), Henze (1986) and Arnold et al. (1993). It is discussed in more detail

* Tel.: +1 951 827 5939; fax: +1 951 827 3286.
E-mail address: barry.arnold@ucr.edu.

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.05.013

in Arnold and Beaver (2000). The resulting extensions of (1) and (2) now involve two skewing parameters λ_0 and λ_1 and are of the forms

$$f(x;\lambda_0,\lambda_1)=\frac{\phi(x)\Phi(\lambda_0+\lambda_1 x)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)} \tag{3}$$

and

$$f(x;\mu,\sigma,\lambda_0,\lambda_1)=\frac{\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda_0+\lambda_1\left(\frac{x-\mu}{\sigma}\right)\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)}. \tag{4}$$

Many extensions of these models have been proposed. For example, we could select an arbitrary density $\psi_1(x)$ to play the role of $\phi(x)$ in (3) and an equally arbitrary distribution function $\Psi_2(x)$ to play the role of $\Phi(x)$. The resulting family of models, before introducing location and scale parameters, is of the form

$$f(x;\lambda_0,\lambda_1)\propto\psi_1(x)\Psi_2(\lambda_0+\lambda_1 x). \tag{5}$$

Computation of the required normalizing constant in (5) may be troublesome and indeed it will frequently be necessary to determine the constant by numerical integration.

There are multivariate extensions of the model (3) which may be viewed as having begun with a $k+m$ dimensional random vector $(\underline{X},\underline{Y})$ (here $\underline{X}$ is of dimension k and $\underline{Y}$ of dimension m) and only observing $\underline{X}$ if $\underline{Y}>\underline{y}_0$, where $\underline{y}_0$ is a pre specified vector in $\mathbb{R}^m$. Most skew models of this genre begin by assuming a classical $k+m$ dimensional normal distribution for $(\underline{X},\underline{Y})$. An extensive survey of such models may be found in Azzalini (2006). In the present paper we will focus on general hidden truncation models (which of course include skew-normal models) beginning with a completely general distribution for (X,Y) (or for $(\underline{X},\underline{Y})$ in higher dimensional settings). This general hidden truncation paradigm will be shown to yield a remarkably rich vein of models which may profitably be used to fit univariate and multivariate data sets. Naturally, it would be desirable to identify a stochastic mechanism involving hidden truncation which can plausibly be argued to have played a role in generating the data set that is fitted by such a model. However, absent such identification, the hidden truncation model, provided that it fits well, may still be useful for prediction purposes.

Returning to the simple hidden truncation model in which we observe X only if $Y>y_0$, it is evident that the density of the observed X's will have a (conditional) distribution of the form

$$F_{X|Y>y_0}(x_0)=P(X\le x_0|Y>y_0)=\frac{\int_{-\infty}^{x_0}\int_{y_0}^{\infty}f_{X,Y}(x,y)\,dx\,dy}{\int_{y_0}^{\infty}f_Y(y)\,dy}. \tag{6}$$

Assuming the existence of densities (as will be done throughout most of this paper) we can write the corresponding conditional density as

$$f_{X|Y>y_0}(x)=\frac{\int_{y_0}^{\infty}f_{X|Y}(x|y)\,dy}{\bar{F}_Y(y_0)}. \tag{7}$$

In this formulation, the marginal density of Y and the conditional density of X given Y will determine the resulting hidden truncation model. In a sense, the model is parameterized by $y_0\in\mathbb{R}$ and $f_Y(y)$ (or, more generally, a parametric family of densities $f_Y(y;\underline{\theta})$) and by $f_{X|Y}(x|y)$ (or, more generally, by a parametric family of densities $f_{X|Y}(x|y;\underline{\tau})$). Clearly this represents an enormously flexible family of models. For example, we could take $f_Y(y)$ to be a normal density and take $f_{X|Y}(x|y)$ to be normal with linear regression and constant conditional variance. Inexorably we are led to the skew-normal model (4). But we could get a richer family by allowing $f_{X|Y}(x|y)$ to have a more general regression function and perhaps a non-constant conditional variance function. This approach merits further investigation. However it is not the approach that will be followed in the rest of the present paper.

A joint density for (X,Y) can be written as the product $f_Y(y)f_{X|Y}(x|y)$, but equally well it can be written as $f_X(x)f_{Y|X}(y|x)$. Using this expression for the joint density, it is readily verified that

$$f_{X|Y>y_0}(x)=f_X(x)\frac{P(Y>y_0|X=x)}{P(Y>y_0)}. \tag{8}$$

In this formulation, the skewed distribution obtained by hidden truncation is clearly shown to be a weighted version of the original density for X. The weight function, $P(Y>y_0|X=x)$, depends on y_0 and on the conditional density of Y given X. The representation of the hidden truncation density in the form (8) may be found in Arellano-Valle et al. (2002) (their equation (5.1)) in the case in which $y_0=0$ but, as they remark, it is likely that it had appeared elsewhere at some time previous to 2002.

In subsequent sections, we will investigate hidden truncation models of the form (8) (truncation from below) as well as other truncation paradigms. In all cases, the basic components of the models will be a given density for X (or $\underline{X}$ in higher dimensions) and a given conditional density for Y given X (or for $\underline{Y}$ given $\underline{X}$).

2. Basic hidden truncation models

Begin with a two dimensional absolutely continuous random vector (X, Y). We might focus on the conditional distribution of X given $Y \in C$ where C is a Borel set in $\mathbb{R}$. Indeed we could write

$$f_{X|Y \in C}(x) = f_X(x) \frac{P(Y \in C|X = x)}{P(Y \in C)} \tag{9}$$

(see Arellano-Valle et al., 2006, where such general models are introduced). However, we will concentrate on hidden truncation of one of three forms only:

1. Lower truncation, where $C = (y_0, \infty)$.
2. Upper truncation, where $C = (-\infty, y_0)$.
3. Two sided truncation where $C = (a, b)$.

For upper truncation at y_0, in which observations are only available for X's whose concomitant variable Y is less than y_0, Eq. (9) becomes

$$f_{y_0-}(x) = f_X(x) \frac{P(Y \leq y_0|X = x)}{P(Y \leq y_0)}. \tag{10}$$

Models of this type are thus characterized by

1. $f_X(x)$, the density assumed for X.
2. The conditional density of Y given X, $f_{X|Y}(x|y)$.
3. The specific truncation point, y_0.

Note that the distribution function corresponding to (10) is of the form

$$F_{y_0-}(x) = P(X \leq x|Y \leq y_0). \tag{11}$$

Consequently, a convenient way to generate models of this type is to begin with a joint distribution for (X, Y) for which $P(X \leq x|Y \leq y)$ is available in a simple form (discussion of such bivariate distributions may be found in Arnold et al., 1999 and Arnold, 1995).

Models involving lower truncation will be of the form

$$f_{y_0+}(x) = f_X(x) \frac{P(Y > y_0|X = x)}{P(Y > y_0)}, \tag{12}$$

with corresponding survival function

$$\overline{F}_{y_0+}(x) = P(X > x|Y > y_0). \tag{13}$$

Technically, models of the form (12) could be viewed as equivalent to those given by (10). One merely needs to replace the concomitant variable Y by $-Y$ (or for non-negative variables, by $1/Y$) to go from one to the other. In practice, such a transformation may not seem to be natural and the concepts of upper and lower truncation are best dealt with separately.

Two sided truncation models are of the form

$$f_{a,b}(x) = f_X(x) \frac{P(a < Y \leq b|X = x)}{P(a < Y \leq b)}. \tag{14}$$

Such models are determined by the choice of the basic marginal density $f_X(x)$, the choice of conditional density $f_{Y|X}(y|x)$ and the truncation points, a and b.

It will be observed that the upper and lower truncation models can be obtained as limiting cases of two sided truncation models, so in a sense we need only to deal with two sided truncation models. Typically the one sided models are simpler in structure and they sometimes can be obtained directly more easily, without first considering a two sided model. Note that, in order for any of these models to assume a tractable form, it is necessary that the conditional distribution of Y given X should have an analytic expression for its distribution function, or at least that the conditional distribution can be evaluated by reference to available tables.

When using the formulations (10), (12) and (14) to construct flexible families of densities it will, as remarked earlier, typically be the case that the density of X is assumed to be a member of some parametric family of densities $f_X(x, \underline{\theta})$ and that the conditional density of Y given X is a member of another parametric family of densities $f_{X|Y}(x|y; \underline{\eta})$. We will consider some examples in which the family of marginal densities for X and the family of conditional densities for Y given X are of the same form (e.g. they might both be normal), but we can increase flexibility by mixing and matching (e.g. one family might be Weibull and the other gamma).

Before embarking on an investigation of some of the many parametric families of models that can be generated by such hidden truncation constructions, it is appropriate to remark that, beginning with a given choice of density function for X, say $f_0(x)$, it is possible to generate, via hidden truncation, an extremely broad class of densities by judicious choice of the conditional density of Y given X. Just about any density with the same support as $f_0(x)$ and lighter tails than $f_0(x)$ can be generated in this fashion. For example, suppose that we wish to generate the density $f_1(x)$ by applying hidden truncation to $f_0(x)$. In order to achieve this in a simple fashion, we need to assume that there exists $c>0$ such that $f_1(x)/f_0(x) \le c, \forall x$. If such a c exists, we can choose a family of conditional densities of Y given X in such a fashion that

$$P(Y \le 0|X=x) = \frac{1}{c}\frac{f_1(x)}{f_0(x)}. \tag{15}$$

With this choice of conditional distributions of Y given X and by setting $y_0=0$, we may verify that hidden truncation above at 0, applied to $f_0(x)$ will yield, via Eq. (10), the desired density $f_1(x)$.

3. Hidden truncation using normal component densities

We begin by considering hidden truncation applied to classical bivariate normal data. In this case two sided hidden truncation will be considered (from which, results for upper and lower truncation can be readily derived). Thus we begin with $X \sim N(\mu, \sigma^2)$ and we will assume the linear regression and constant conditional variance that is associated with the classical bivariate normal distribution for (X,Y). Thus we assume that $Y|X=x \sim N(\alpha+\beta x, \tau^2)$. Referring to (14), for hidden truncation points a and b, we have

$$\begin{aligned} f_{a,b}(x) &= f_X(x)\frac{P(a<Y\le b|X=x)}{P(a<Y\le b)} \\ &\propto \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{b-\alpha-\beta x}{\tau}\right)-\Phi\left(\frac{a-\alpha-\beta x}{\tau}\right)\right]. \end{aligned} \tag{16}$$

In this expression, $\mu, \alpha, \beta \in \mathbb{R}$ and $\sigma, \tau \in \mathbb{R}^+$ and $a<b$. It is convenient to introduce new parameters δ_1, δ_2 and λ_1 where $-\infty < \delta_1 < \delta_2 < \infty$ and $\lambda_1 \in \mathbb{R}$, allowing us to rewrite the model (16) as

$$f_{a,b}(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\frac{\left[\Phi\left(\delta_2+\lambda_1\left(\frac{x-\mu}{\sigma}\right)\right)-\Phi\left(\delta_1+\lambda_1\left(\frac{x-\mu}{\sigma}\right)\right)\right]}{\left[\Phi\left(\frac{\delta_2}{\sqrt{1+\lambda_1^2}}\right)-\Phi\left(\frac{\delta_1}{\sqrt{1+\lambda_1^2}}\right)\right]}. \tag{17}$$

This is the model involving two sided hidden truncation that was discussed, for example, in Arnold et al. (1993). If we consider upper truncation (letting $\delta_1 \to -\infty$ in (18)) we obtain the Henze–Arnold–Beaver skew-normal model (4) (where δ_2 is replaced by λ_0).

Instead of using a conditional distribution for Y given X that is normal with a linear regression function and a constant conditional variance function, we could consider a normal distribution with more general regression and conditional variance functions. Thus, if we assume that $Y|X=x \sim N(\mu(x), \tau^2(x))$, our two sided hidden truncation model becomes

$$f_{a,b}(x) \propto \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{b-\mu(x)}{\tau(x)}\right)-\Phi\left(\frac{a-\mu(x)}{\tau(x)}\right)\right]. \tag{18}$$

The model (18) includes (as limiting cases) densities of the form

$$f(x;\underline{\lambda}) \propto \phi(x)\Phi\left(\frac{\lambda_{00}+\lambda_{10}x}{\sqrt{1+(\lambda_{01}+\lambda_{11}x^2)}}\right). \tag{19}$$

Such densities have been studied earlier in the literature. They are identifiable as marginal densities of the following class of bivariate distributions with conditionals in the skew-normal family (4)

$$f(x,y;\underline{\lambda}) \propto \phi(x)\phi(y)\Phi(\lambda_{00}+\lambda_{10}x+\lambda_{01}y+\lambda_{11}xy) \tag{20}$$

(See, for example, Arnold et al., 2002). Such models can also be obtained as mixtures of univariate skew-normal densities (see, for example, Arellano-Valle et al., 2004).

In fact, model (18) is, in a sense, completely general. Any weighted version of the normal $N(\mu, \sigma^2)$ density can be represented in the form (18). Suppose that we wish to have

$$f_{a,b}(x) \propto w(x)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \tag{21}$$

for some specified weight function $w(x)$. We can choose $a=-\infty$, $b=0$ and $\tau(x)=1$. The choice of $\mu(x)$ which will then enable us to identify (21) as a special case of (18) will be such that $\Phi(-\mu(x)) = w(x)$, i.e. we should choose $\mu(x) = -\Phi^{-1}(w(x))$.

4. Hidden truncation applied to normal conditionals distributions

Following early work by Bhattacharyya (1944), Arnold et al. (1999) provided detailed discussion of the class of bivariate densities, $f_{X,Y}(x,y)$, which have all of their conditional densities (of X given Y and of Y given X) of the normal form. Such bivariate densities are necessarily of the form

$$f_{X,Y}(x,y) = \exp - \left\{ (1,x,x^2) \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix} \right\}, \tag{22}$$

where the m_{ij}'s satisfy certain constraints to ensure integrability. For our hidden truncation constructions, we need expressions for the corresponding marginal $f_X(x)$ and for the conditional densities $f_{Y|X}(y|x)$. It is not difficult to verify that, if (X,Y) has density (22), then

$$f_X(x) = \frac{\exp - \left(\frac{1}{2} \left(2(m_{20}x^2 + m_{10}x + m_{00}) - \frac{(m_{21}x^2 + m_{11}x + m_{01})^2}{2(m_{22}x^2 + m_{12}x + m_{02})} \right) \right)}{\sqrt{2(m_{22}x^2 + m_{12}x + m_{02})}}, \tag{23}$$

while $Y|X = x \sim N(\mu(x), \sigma^2(x))$ in which

$$\mu(x) = -\frac{(m_{21}x^2 + m_{11}x + m_{01})}{2(m_{22}x^2 + m_{12}x + m_{02})} \tag{24}$$

and

$$\sigma^2(x) = \frac{1}{2(m_{22}x^2 + m_{12}x + m_{02})}. \tag{25}$$

The corresponding two sided hidden truncation model will be

$$f_{a,b}(x) \propto f_X(x) \left[\Phi(\frac{b - \mu(x)}{\sigma(x)}) - \Phi\left(\frac{a - \mu(x)}{\sigma(x)} \right) \right], \tag{26}$$

where $f_X(x)$, $\mu(x)$ and $\sigma(x)$ are as defined in (23)–(25), respectively.

The centered normal conditionals model is considerably simpler. For it, we set $m_{01} = m_{10} = m_{11} = m_{12} = m_{21} = 0$ in (22). This leaves us with a three parameter bivariate density for which

$$f_X(x) = \frac{e^{-m_{20}x^2}}{\sqrt{2(m_{22}x^2 + m_{02})}} \tag{27}$$

and

$$Y|X = x \sim N\left(0, \frac{1}{2(m_{22}x^2 + m_{02})}\right), \tag{28}$$

so that

$$f_{a,b}(x) \propto \frac{e^{-m_{20}x^2}}{\sqrt{2(m_{22}x^2 + m_{02})}} [\Phi(b\sqrt{2(m_{22}x^2 + m_{02})}) - \Phi(a\sqrt{2(m_{22}x^2 + m_{02})})]. \tag{29}$$

5. Hidden truncation with exponential component densities

Suppose now that X has an exponential distribution, i.e.

$$P(X > x) = e^{-\alpha x}, \quad x > 0. \tag{30}$$

Now assume that, for each $x > 0$, the conditional density of Y given $X = x$ is also an exponential density with a constant failure rate which depends linearly on x. Thus

$$P(Y > y|X = x) = e^{-(\beta + \gamma x)y}, \quad y > 0. \tag{31}$$

The resulting joint density is of the form

$$f(x,y) = (\alpha\beta + \alpha\beta x) \exp(-[\alpha x + \beta y + \gamma xy]), \quad x > 0, \ y > 0. \tag{32}$$

The corresponding two sided hidden truncation model will then be

$$f_{a,b}(x) \propto \alpha e^{-\alpha x}[e^{-(\beta+\gamma x)a} - e^{-(\beta+\gamma x)b}], \quad x > 0, \tag{33}$$

a linear combination of two exponential densities. The lower hidden truncation model is obtained from (33) by setting $b = \infty$ and $a = y_0$, in this manner we find

$$f_{y_0+}(x) = (\alpha + \gamma y_0)e^{-(\alpha+\gamma y_0)x}, \quad x > 0, \tag{34}$$

i.e. again an exponential density. Thus, in this situation, lower hidden truncation does not lead to an enrichment of the class of distributions for X. A similar phenomenon is observable if we begin with (X, Y) having an exponential conditionals distribution (see Arnold and Strauss, 1988). The corresponding joint density is of the form

$$f(x, y) \propto \exp -(\alpha x + \beta y + \gamma xy), \quad x > 0, \ \ y > 0. \tag{35}$$

In this case the marginal density for X is

$$f_X(x) \propto (\beta + \gamma x)^{-1}e^{-\alpha x}, \quad x > 0, \tag{36}$$

and the conditional survival function of Y given $X = x$ is of the form

$$P(Y > y|X = x) = e^{-(\beta+\gamma x)y}, \quad x > 0 \tag{37}$$

(the same as (31)). The corresponding lower hidden truncation model will be

$$f_{y_0+} \propto (\beta + \gamma x)^{-1}e^{-(\alpha+\gamma y_0)x}, \quad x > 0. \tag{38}$$

Observe that (38) is obtainable from (36) by a simple change of one of the parameters and the family of lower hidden truncation models coincides with the original family of densities for X.

It becomes evident that the use of lower hidden truncation with a conditional distribution given by (31) will be ineffective in enriching the class of densities assumed for X whenever $f_X(x)$ includes a factor of the form $e^{-g(\underline{\theta})x}$, where $g(\underline{\theta}) > 0$. Thus, for example, if we begin with an assumption that $X \sim N(\mu, \sigma^2)$ where $\mu < 0$ and assume that $Y|X = x$ has a distribution satisfying (31) (i.e. an exponential conditional distribution with a constant failure rate that is a linear function of x), then the resulting lower hidden truncation models will again be normal with negative means.

If we allow the failure rate for the conditional distribution of Y given $X = x$ to depend on x in a non-linear fashion, we can, of course, get new densities by using the lower hidden truncation paradigm as the following examples show.

Let us begin with $X \sim \exp(\alpha)$, i.e. $P(X > x) = e^{-\alpha x}$. Now assume that $P(Y > y|X = x) = e^{-\gamma(x)y}$ for some positive function $\gamma(x)$ defined on $\mathbb{R}^+$. It follows that

$$f_{y_0+} \propto e^{-(\alpha x+\gamma(x)y_0)}, \quad x > 0. \tag{39}$$

As a special case, consider $\gamma(x) = \gamma x^2$ for some $\gamma > 0$. In this case, we find that

$$f_{y_0+} \propto e^{-(\alpha x+\gamma y_0 x^2)}, \quad x > 0. \tag{40}$$

It is perhaps surprising that a truncated normal density such as (40) can arise via hidden truncation applied to a model with an exponential marginal distribution for X and exponential conditionals for Y given X.

However, it is true that a very broad class of densities can be obtained by hidden truncation, even when we restrict attention to pre-truncated models involving an exponential marginal and exponential conditionals. Suppose that we wish to obtain a specific target density $g(x)I(x > 0)$ in this manner. To do this, we must select y_0, k, α and $\gamma(x)$ such that, for $x > 0$,

$$g(x) = \exp(m - \alpha x - y_0\gamma(x)). \tag{41}$$

So $\gamma(x)$ must satisfy

$$\gamma(x) = \frac{m - \alpha x - \log g(x)}{y_0}. \tag{42}$$

For certain choices of $g(x)$, the corresponding function $\gamma(x)$ given by (42) will not be always positive, so that some mild conditions must be imposed on the form of $g(x)$ in order for it to be obtainable via hidden truncation using exponential model components. Nevertheless, an extremely broad class of densities $g(x)$ on $\mathbb{R}^+$ can be so constructed.

6. Hidden truncation with Pareto component densities

A bivariate Pareto conditionals density is of the form (Arnold et al., 1999)

$$f(x,y)=(\alpha+\beta x+\gamma y+\delta xy)^{-(\tau+1)}, \tag{43}$$

where $\alpha,\beta,\gamma,\delta,\tau>0$. The corresponding marginal and conditional densities are

$$f_X(x)\propto[(\alpha+\beta x)^{\tau}(\gamma+\delta x)]^{-1}, \quad x>0 \tag{44}$$

and for $x>0$,

$$f_{Y|X}(y|x)\propto\left[1+\frac{\gamma+\delta x}{\alpha+\beta x}y\right]^{-(\tau+1)}, \quad y>0 \tag{45}$$

(i.e. $Y|X=x\sim$ Pareto$((\gamma+\delta x)/(\alpha+\beta x),\tau)$). The two sided hidden truncation distribution derived from this joint density will then be given by

$$\begin{aligned}f_{a,b}(x)&\propto f_X(x)P(a<Y\le b|X=x)\\&\propto[(\alpha+\beta x)^{\tau}(\gamma+\delta x)]^{-1}\left\{\left[1+\frac{\gamma+\delta x}{\alpha+\beta x}a\right]^{-\tau}-\left[1+\frac{\gamma+\delta x}{\alpha+\beta x}b\right]^{-\tau}\right\}\\&\propto\frac{1}{\gamma+\delta x}\left\{\frac{1}{(\alpha+\beta x+\gamma a+\delta ax)^{\tau}}-\frac{1}{(\alpha+\beta x+\gamma b+\delta bx)^{\tau}}\right\},\end{aligned} \tag{46}$$

for $x>0$. Recalling that $\alpha,\beta,\gamma,\delta>0$ and $0\le a<b$, we may write this as

$$f_{a,b}(x)\propto\frac{1}{\gamma+\delta x}\left[\frac{1}{(\alpha_1+\beta_1x)^{\tau}}-\frac{1}{(\alpha_2+\beta_2x)^{\tau}}\right], \quad x>0, \tag{47}$$

a linear combination of two densities of the same form as the original marginal density of X (as in (44)).

To obtain $f_{y_0-}(x)$ we just set $a=0$ and $b=y_0$ in (46) and we obtain a density, also of the form (47), with $\alpha_1=\alpha$ and $\beta_1=\beta$.

The lower hidden truncation model is simpler. To get $f_{y_0+}(x)$ we must set $a=y_0$ and $b=\infty$ in (46) to obtain

$$f_{y_0+}(x)\propto\frac{1}{\gamma+\delta x}\frac{1}{(\alpha+\beta x+\gamma y_0+\delta y_0x)^{\tau}}, \quad x>0, \tag{48}$$

or equivalently

$$f_{y_0+}(x)\propto\frac{1}{\gamma+\delta x}\frac{1}{(\alpha_2+\beta_2x)^{\tau}}, \quad x>0. \tag{49}$$

So in this case we again observe the phenomenon in which lower hidden truncation fails to augment the class of models already assumed for X.

If we begin with a Pareto distribution for X and Pareto conditional distributions for Y given X, as in (45), then lower hidden truncation will lead to an enriched family of densities. We will have

$$f_X(x)\propto(\alpha+\beta x)^{-(\eta+\tau+1)}, \quad x>0 \tag{50}$$

and

$$P(Y>y_0|X=x)=\left[1+\frac{\gamma+\delta x}{\alpha+\beta x}y_0\right]^{-\tau}, \tag{51}$$

so that, after reparameterization, we have

$$f_{y_0+}(x)\propto(\alpha+\beta x)^{-(\eta+1)}(\alpha'+\beta'x)^{-\tau}, \tag{52}$$

where $\alpha\le\alpha'$ and $\beta\le\beta'$.

7. Multivariate cases

In the development thus far, both X and Y have been scalar variables. Of course, analogous arguments can be advanced when the variables are of higher dimensions. Thus one may consider a $k+m$ dimensional random vector $(\underline{X},\underline{Y})$ where $\underline{X}$ is of dimension

k and $\underline{Y}$ is of dimension m. We will consider the distribution of $\underline{X}$ subject to hidden truncation on $\underline{Y}$ of the form $\underline{Y} \leq \underline{y}_0$. We can write this conditional density (in a form analogous to (10)) as

$$f_{\underline{y}_0} - (\underline{x}) = f_{\underline{X}}(\underline{x}) \frac{P(\underline{Y} \leq \underline{y}_0 | \underline{X} = \underline{x})}{P(\underline{Y} \leq \underline{y}_0)}. \tag{53}$$

Hidden truncation models of this type will be determined by the marginal density $f_{\underline{X}}(\underline{x})$, the conditional density of $\underline{Y}$ given $\underline{X}$, $f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$, and the truncation point $\underline{y}_0$.

In this section, we will restrict attention to an illustrative case in which the component densities ($f_{\underline{X}}$ and $f_{\underline{Y}|\underline{X}}$) are multivariate normal although, of course, the ideas discussed can be extended readily to deal with other examples, perhaps with component densities of different (non-normal) types.

We consider such hidden truncation in a setting in which $(\underline{X}.\underline{Y})$ has a classical $k+m$ dimensional normal distribution. Thus we begin with

$$\begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix} \sim N^{(k+m)} \left(\begin{pmatrix} \underline{\mu} \\ \underline{\nu} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right). \tag{54}$$

In this case $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma_{11})$ and the conditional distribution of $\underline{Y}$ given $\underline{X} = \underline{x}$ is of the form

$$\underline{Y}|\underline{X} = \underline{x} \sim N^{(m)}(\underline{\nu} + \Sigma_{21}\Sigma_{11}^{-1}(\underline{x} - \underline{\mu}), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}). \tag{55}$$

We will introduce notation as follows:

$$\phi^{(k)}(\underline{x}) = \prod_{i=0}^{k} \phi(x_i) \tag{56}$$

and

$$\Phi^{(m)}(\underline{y}; \underline{\delta}, \Lambda) = P(\underline{Y} \leq \underline{y}), \tag{57}$$

where $\underline{Y} \sim N(\underline{\delta}, \Lambda)$. With this notation, referring to (53), we will have

$$f_{\underline{y}_0-}(\underline{x}) = |\Sigma_{11}|^{-1/2} \phi^{(k)}(\Sigma_{11}^{-1/2}(\underline{x} - \underline{\mu})) \frac{\Phi^{(m)}(\underline{y}_0 - \underline{\nu} - \Sigma_{21}\Sigma_{11}^{-1}(\underline{x} - \underline{\mu}); \underline{0}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})}{\Phi^{(m)}(\underline{y}_0 - \underline{\nu}; \underline{0}, \Sigma_{22})}. \tag{58}$$

At this point it is convenient to make a change of variables, defining $\underline{Z} = \Sigma_{11}^{-1/2}(\underline{X} - \underline{\mu})$ so that $\underline{Z} \sim N^{(k)}(\underline{0}, I)$. If $\underline{X}$ has density (58), then $\underline{Z}$ will have a density of the following form:

$$f_{\underline{y}_0-}(\underline{z}) = \phi^{(k)}(\underline{z}) \frac{\Phi^{(m)}(\underline{\lambda}_0 + \Lambda\underline{z}; \underline{0}, \Delta)}{\Phi^{(m)}(\underline{\lambda}_0; \underline{0}, \Delta + \Lambda^T\Lambda)}, \tag{59}$$

for suitably defined $\underline{\lambda}_0$, Δ and Λ (which will depend on the choice of $\underline{y}_0$). The model (59) is known in the literature under a variety of names with variations in the labeling of the parameters. For example González-Farías et al. (2004) call it the closed skew-normal family, while Arellano-Valle and Genton (2005) refer to it as the fundamental skew-normal distribution. See Azzalini (2005) and Arnold and Beaver (2007) for further discussion of these and other aliases.

It is not difficult to deal with analogous lower and two sided hidden truncation models. It is convenient to use the notation $\overline{\Phi}^{(m)}(\underline{y}; \underline{\delta}, \Lambda)$ to denote $P(\underline{Y} > \underline{y})$ where $\underline{Y} \sim N^{(m)}(\underline{\delta}, \Lambda)$. The resulting models obtained by hidden truncation on $\underline{Y}$ applied to the density of $\underline{Z}$ are

$$f_{\underline{y}_0+}(\underline{z}) = \phi^{(k)}(\underline{z}) \frac{\overline{\Phi}^{(m)}(\underline{\lambda}_0 + \Lambda\underline{z}; \underline{0}, \Delta)}{\overline{\Phi}^{(m)}(\underline{\lambda}_0; \underline{0}, \Delta + \Lambda^T\Lambda)} \tag{60}$$

and

$$f_{\underline{a},\underline{b}}(\underline{z}) = \phi^{(k)}(\underline{z}) \frac{\Phi^{(m)}(\underline{\delta}_2 + \Lambda\underline{z}; \underline{0}, \Delta) - \Phi^{(m)}(\underline{\delta}_1 + \Lambda\underline{z}; \underline{0}, \Delta)}{\Phi^{(m)}(\underline{\delta}_2; \underline{0}, \Delta + \Lambda^T\Lambda) - \Phi^{(m)}(\underline{\delta}_1; \underline{0}, \Delta + \Lambda^T\Lambda)}. \tag{61}$$

There is a considerable literature devoted to the discussion of the distribution of $\underline{X} = \underline{\mu} + \Sigma_{11}^{1/2}\underline{Z}$ where $\underline{Z}$ has a hidden truncation density of the form (59). Of course, (60) can be viewed as a special case of (59) in which $\underline{Y}$ has been replaced by $-\underline{Y}$ and $\underline{y}$ by $-\underline{y}$. Densities of the form (61) have received less attention, even though such two sided hidden truncation can be expected to be encountered in many real world data configurations.

8. Envoi

Generally speaking, hidden truncation models will be difficult to deal with analytically unless the joint density of (X, Y) (or of $(\underline{X}, \underline{Y})$ in higher dimensions) is a member of some tractable family of multivariate distributions. Even in such cases, an awkward normalizing constant may be associated with the hidden truncation distribution. Techniques for dealing with inference problems, even for hidden truncation models as simple as the basic Azzalini model (1), still require refinement. Much work remains to be done before the more complicated hidden truncation models can be expected to enter into the applied statistician's toolkit.

References

Arellano-Valle, R.B., Branco, M., Genton, M., 2006. A unified view on skewed distributions arising from selections. Canad. J. Statist. 34, 581–601.
Arellano-Valle, R.B., Gonez, H., Quintana, F., 2004. A new class of skew-normal distributions. Commun. Statist. Theory Methods 33, 1465–1480.
Arellano-Valle, R.B., Genton, M., 2005. On fundamental skew distributions. J. Multivariate Anal. 96, 93–116.
Arellano-Valle, R.B., del Pino, G., San Martín, E., 2002. Definition and probabilistic properties of skew-distributions. Statist. Probab. Lett. 58 (2), 111–121.
Arnold, B.C., 1995. Conditional survival models. In: Balakrishnan, N. (Ed.), Recent Advances in Life-Testing and Reliability. CRC Press, Boca Raton, FL, pp. 589–601.
Arnold, B.C., Beaver, R.J., 2000. Hidden truncation models. Sankhyā Ser. A 62 (1), 22–35.
Arnold, B.C., Beaver, R.J., 2007. Skewing around: relationships among classes of skewed distributions. Methodol. Comput. Appl. Probab. 9 (2), 153–162.
Arnold, B.C., Beaver, R.J., Groeneveld, R.A., Meeker, W.Q., 1993. The nontruncated marginal of a truncated bivariate normal distribution. Psychometrika 58 (3), 471–488.
Arnold, B.C., Castillo, E., Sarabia, J.M., 1999. Conditional specification of statistical models. In: Springer Series in Statistics. Springer, New York.
Arnold, B.C., Castillo, E., Sarabia, J.M., 2002. Conditionally specified multivariate skewed distributions. Sankhyā Ser. A 64 (2), 206–226 (selected articles from San Antonio Conference in honour of C.R. Rao, San Antonio, TX, 2000).
Arnold, B.C., Strauss, D., 1988. Bivariate distributions with exponential conditionals. J. Amer. Statist. Assoc. 83 (402), 522–527.
Azzalini, A., 1985. A class of distributions which includes the normal ones. Scand. J. Statist. 12 (2), 171–178.
Azzalini, A., 2005. The skew-normal distribution and related multivariate families. Scand. J. Statist. 32 (2), 159–200 (with discussion by Marc G. Genton and a rejoinder by the author).
Bhattacharyya, A., 1944. On some sets of sufficient conditions leading to the normal bivariate distribution. Sankhyā 6, 399–406.
Genton, M.G. (Ed.), 2004. Skew-elliptical distributions and their applications. A Journey Beyond Normality. Chapman & Hall, CRC Press, Boca Raton, FL.
González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K., 2004. The closed skew-normal distribution. In: Genton, M.G. (Ed.), Skew-Elliptical Distributions and their Applications. Chapman & Hall, CRC Press, Boca Raton, FL, pp. 25–42.
Henze, N., 1986. A probabilistic representation of the "skew-normal" distribution. Scand. J. Statist. 13 (4), 271–275.
Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press, Cambridge.