Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 206-226

CONDITIONALLY SPECIFIED MULTIVARIATE SKEWED DISTRIBUTIONS

By BARRY C. ARNOLD,
University of California, Riverside, USA
ENRIQUE CASTILLO
and
JOSÉ MARÍA SARABIA
University of Cantabria, Santander, Spain

SUMMARY. The basic skew normal model $f(x; \lambda) = 2\phi(x)\Phi(\lambda x)$, where ϕ (respectively Φ) is the standard normal density (respectively distribution), may be used as a component in the construction of flexible families of multivariate densities using a conditional specification paradigm. Parallel developments are outlined in the case in which the basic distribution is something other than normal (for example, Cauchy). The development of these and related models involves identification of broad classes of particular solutions to certain functional equations.

1. **Conditional Specification**

Since conditional densities (i.e. cross sections of multivariate densities) are easier to visualize and inherently more informative than are marginal densities, a case can be made for specifying joint densities entirely in terms of conditional densities. A survey of such conditionally specified models may be found in Arnold, Castillo and Sarabia (1999). The basic strategy in the bivariate case involves positing that all conditionals of X given $Y = y$ should be members of some specified family of densities while, at the same time, all conditional densities of Y given $X = x$ should be members of a second specified family of densities. The end result is a flexible family of bivariate densities which includes the independent marginal case and

Paper received October 2000; revised October 2001.
AMS (2000) subject classification. Primary 62H05.
Keywords and phrases. Skewed normal density, hidden truncation, functional equations, Cauchy, Laplace, logistic, normal conditionals.

provides a convenient alternative to more classical bivariate densities, which are often constructed by marginal specification. One advantage of the conditional specification route is that the resulting densities are easy to simulate using Gibbs sampler techniques, indeed they are tailor made for such simulation. The prototypical such conditionally specified density is the normal conditionals density introduced by Bhattacharyya (1944), whose bivariate version can be written in compact form as

$$f_{X,Y}(x,y) = \exp((1,x,x^2)A(1,y,y^2)') \tag{1.1}$$

where A is a 3×3 matrix of parameters which must satisfy certain constraints in order for (1.1), to represent a valid (integrable) density. Actually there are only 8 parameters in (1.1) since a_{11} will be a function of the other a_{ij}'s chosen so that (1.1) integrates to 1. It is evident that (1.1) indeed does have all of its conditional densities (of X given Y and of Y given X) of the normal form. It is also evident that (1.1) includes the classical bivariate normal density as a special case. We will return to consider (1.1) further in section 5. We will now turn to consider arguments in favour of the use of skewed rather than symmetric normal densities in modeling scenarios.

2. Skewing by Hidden Truncation

To help visualize matters, consider the distribution of weights of basketball players who are at least 6 feet tall. It is eminently plausible that heights and weights (untruncated) might have a classical bivariate normal distribution. But since the variables are correlated, after truncation with respect to height, the resulting distribution of the untruncated variable (weight) will be skewed and not normal. Such models have been discussed in Azzalini (1985) and also Arnold, Beaver, Groeneveld and Meeker (1993). The basic skewed normal (or Azzalini) density takes the form

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

where $\phi(x)$ and $\Phi(x)$ denote, here and henceforth, the standard normal density and distribution functions and where $\lambda \in \mathbb{R}$ is a parameter which governs the skewness of the density. The parameter λ reflects the correlation between the observed variable and the hidden variable with respect to which truncation has occurred. If X has density (2.1), we will write $X \sim SN(\lambda)$.

Actually the Azzalini model corresponds to the case where the hidden variable was truncated at its mean value. If we wish to entertain the possibility of other truncation thresholds we are led to consider a slightly more general model which we call the linearly skewed normal density. It is of the form:

$$f(x;\lambda_0,\lambda_1) = 2\phi(x)\Phi(\lambda_0+\lambda_1 x)/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right), \quad x \in \mathbb{R}. \tag{2.2}$$

If X has density (2.2) we will write $X \sim LSN(\lambda_0,\lambda_1)$. We shall actually consider certain higher order skewed distributions. The quadratically skewed normal distribution will have density of the form

$$f(x;\lambda_0,\lambda_1,\lambda_2) \propto \phi(x)\Phi(\lambda_0+\lambda_1 x+\lambda_2 x^2) \tag{2.3}$$

which can be written as $X \sim QSN(\lambda_0,\lambda_1,\lambda_2)$. More generally we may speak of polynomially skewed normal densities of the form

$$f(x) \propto \phi(x)\Phi(P_k(x)), \tag{2.4}$$

where P_k is a polynomial in x of degree k. It is natural to introduce location and scale parameters μ and σ in the families (2.1), (2.2), (2.3) leading to notation such as $X \sim LSN(\mu,\sigma;\lambda_0,\lambda_1)$. To avoid confusion we might, on occasion, write the slightly more cumbersome $X \sim LSN(0,1;\lambda_0,\lambda_1)$ instead of $LSN(\lambda_0,\lambda_1)$ to denote the centered and unscaled version of (2.2). Similarly we can use the notation $SN(0,1,\lambda)$ and $QSN(0,1,\lambda_0,\lambda_1,\lambda_2)$ for densities (2.1) and (2.3).

Properties of the basic skewed normal and the linearly skewed normal are extensively discussed in Azzalini (1985) and Arnold et al (1993). Expressions for the moments can be obtained from available expressions for the moment generating functions of these densities.

3. Skewed Densities in Higher Dimensions

A k-dimensional version of the basic skew normal density was introduced by Azzalini and Dalla Valle (1996).

The basic Azzalini-Dalla Valle k-dimensional distribution takes the form

$$f(\boldsymbol{x};\boldsymbol{\lambda}) = 2\left[\prod_{i=1}^{k}\phi(x_i)\right]\Phi\left(\sum_{i=1}^{k}\lambda_i x_i\right), \quad \boldsymbol{x} \in \mathbb{R}^k, \tag{3.1}$$

where $\boldsymbol{\lambda}$ is a k-dimensional skewness parameter. A possible genesis for (3.1) involves beginning with a $k+1$ dimensional normal random vector $(\boldsymbol{X},Y)$ and retaining $\boldsymbol{X}$ observations only if Y is above average. By allowing the hidden truncation with respect to Y to be at an arbitrary but fixed level (not necessarily equal to $E(Y)$), a more general model was introduced and discussed in detail in Arnold and Beaver (2000). It has one additional skewness parameter and is of the form:

$$f(\boldsymbol{x}; \lambda_0, \boldsymbol{\lambda}) = \left[\prod_{i=1}^{k} \phi(x_i)\right] \Phi(\lambda_0 + \sum_{i=1}^{k} \lambda_i x_i) / \Phi\left(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\lambda}}}\right), \quad x \in \mathbb{R}. \quad (3.2)$$

The general k-dimension skew normal densities of the forms (3.1) and (3.2) are given by

$$\boldsymbol{W} = \boldsymbol{\mu} + \Sigma^{1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \quad (3.3)$$

where $\boldsymbol{\mu} \in \mathbb{R}$, $\Sigma^{1/2}$ is positive definite and $\boldsymbol{X}$ has density (3.1) or (3.2). Of course (3.1) is a special case of (3.2) corresponding to the choice $\lambda_0 = 0$, so we really don't need two notational conventions for the density. Nevertheless we will continue to distinguish between the models in a manner parallel to that used in Section 2, speaking of k-dimensional skewed normal densities and k-dimensional linearly skewed normal densities. Thus if $\boldsymbol{W} = \mu + \Sigma^{1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ where $\boldsymbol{X}$ has density (3.1), we will write $\boldsymbol{W} \sim SN^{(k)}(\mu, \sum, \boldsymbol{\lambda})$. In parallel fashion, if $\boldsymbol{X}$ has density (3.2) we will write $\boldsymbol{W} \sim LSN^{(k)}(\mu, \sum, \lambda_0, \boldsymbol{\lambda})$. Azzalini and Dalla Valle (1996) observe that any marginal density of a $SN^{(k)}$ density is again of the $SN^{(k_1)}$ form (here we partition $\boldsymbol{X} = (\dot{\boldsymbol{X}}, \ddot{\boldsymbol{X}})$ where $\dot{\boldsymbol{X}}$ is of dimension k_1 and $\ddot{\boldsymbol{X}}$ is of dimension $k - k_1$). However the conditional densities of $SN^{(k)}$ densities (i.e. $f_{\dot{\boldsymbol{X}}|\ddot{\boldsymbol{X}}}(\dot{\boldsymbol{x}}|\ddot{\boldsymbol{x}})$) are not of the $SN^{(k_1)}$ form. In contrast Arnold and Beaver (2000) show that the multivariate linearly skewed normal family of densities has all marginals and all conditional densities of the multivariate linearly skewed normal form.

Are there other multivariate densities which have all their conditional densities of the linearly skewed normal form? This question will be addressed in the following section. However, to simplify notation we initially consider the bivariate case.

4. Bivariate Distributions with Skewed Normal Conditionals

Rather than following Bhattacharyya's assumption of normal conditional densities for our bivariate density, we will allow the conditional densities first to be skewed normal, and then consider linearly skewed normal conditionals.

4.1. *Skewed normal conditionals.* We are interested in the form of the density for a two dimensional random variable (X, Y) such that:

$$\text{for each } y \in \mathbb{R}, \quad X|Y = y \sim SN(\lambda^{(1)}(y))$$

and

$$\text{for each } x \in \mathbb{R}, \quad Y|X = x \sim SN(\lambda^{(2)}(x)) \tag{4.1}$$

for some functions $\lambda^{(1)}(y)$ and $\lambda^{(2)}(x)$. Location and scale parameters can be added later. If (4.1) is to hold, there must exist densities $f_X(x)$ and $f_Y(y)$ such that

$$f_{X,Y}(x,y) = 2\phi(x)\Phi(\lambda^{(1)}(y)x)f_Y(y) = 2\phi(y)\Phi(\lambda^{(2)}(x)y)f_X(x). \tag{4.2}$$

In this functional equation, $f_X(x), f_Y(y), \lambda^{(1)}(y)$ and $\lambda^{(2)}(x)$ are unknown functions to be determined.

We are able to identify two types of solutions to the functional equation (4.2). We conjecture that these are the only solutions. Further discussion of this functional equation is deferred to Section 8. The two solutions that have been identified are as follows:
Type I. (Independence). If $\lambda^{(1)}(y) \equiv \lambda^{(1)}$ and $\lambda^{(2)}(x) \equiv \lambda^{(2)}$ then

$$\begin{aligned} f_X(x) &= 2\phi(x)\Phi(\lambda^{(2)}x), \\ f_Y(y) &= 2\phi(y)\Phi(\lambda^{(1)}y) \end{aligned} \tag{4.3}$$

and

$$f_{X,Y}(x,y) = 4\phi(x)\phi(y)\Phi(\lambda^{(2)}x)\Phi(\lambda^{(1)}y). \tag{4.4}$$

Type II. (Dependent case). Here $\lambda^{(1)}(y) = \lambda y$ and $\lambda^{(2)}(x) = \lambda x$ and consequently

$$f_X(x) = \phi(x), \quad f_Y(y) = \phi(y) \tag{4.5}$$

and

$$f_{X,Y}(x,y) = 2\phi(x)\phi(y)\Phi(\lambda xy). \tag{4.6}$$

These claims can be justified as follows. First, if $\lambda^{(1)}(y) \equiv \lambda^{(1)}$ and $\lambda^{(2)}(x) \equiv \lambda^{(2)}$ then from (4.2) we can conclude that X and Y are independent and then that the left hand and right hand sides of the equations in (4.3) are proportional. Since left and right hand sides are densities they must actually be equal. Thus (4.3) and then (4.4) are verified.

In fact if either of the functions $\lambda^{(1)}(y), \lambda^{(2)}(x)$ is a constant function, the random variables X and Y will be independent and eventually the joint density will be given by (4.4).

Suppose now that $\lambda^{(1)}(y)$ and $\lambda^{(2)}(x)$ are both non-constant functions. Consider the case in which $\lambda^{(1)}(0) = 0$ and $\lambda^{(2)}(0) = 0$. With this condition, by setting $y = 0$ in (4.2) we conclude that $f_X(x) = \phi(x)$. Analogously, setting $x = 0$ in (4.2) yields $f_Y(y) = \phi(y)$.

From this, substituting back in (4.2), we find that

$$\Phi(\lambda^{(1)}(y)x) = \Phi(\lambda^{(2)}(x)y), \quad \forall (x,y) \in \mathbb{R}^2.$$

But, since Φ is a strictly monotone function, this can only hold if $\lambda^{(1)}(y)x = \lambda^{(2)}(x)y, \ \forall(x,y)$. However, by differencing, this occurs only if $\lambda^{(1)}(y) = \lambda y$ and $\lambda^{(2)}(x) = \lambda x$ for some $\lambda \in \mathbb{R}$. Thus we arrive at (4.6).

It remains to consider (in Section 8) the case in which $\lambda^{(1)}(y)$ and $\lambda^{(2)}(x)$ are non-constant functions and at least one of $\lambda^{(1)}(0), \lambda^{(2)}(0)$ is not zero.

It should be remarked that for any $\lambda^{(1)} \in \mathbb{R}$ and $\lambda^{(2)} \in \mathbb{R}$, (4.4) is a proper (integrable) model. Similarly, for any $\lambda \in \mathbb{R}$, (4.6) clearly integrates to 1.

The density (4.6) has, as remarked, standard normal marginals together with skewed normal conditionals. The corresponding regression functions are non-linear and take the form:

$$E(X|Y=y) = \sqrt{\frac{2}{\pi}} \times \frac{\lambda y}{\sqrt{1+\lambda^2 y^2}} \tag{4.7}$$

$$E(Y|X=x) = \sqrt{\frac{2}{\pi}} \times \frac{\lambda x}{\sqrt{1+\lambda^2 x^2}}. \tag{4.8}$$

To evaluate the correlation between X and Y we argue as follows

$$\rho(X,Y) = E(XY) = E(E(XY|Y)) = \sqrt{\frac{2}{\pi}} E\left(\frac{\lambda Y^2}{\sqrt{1+\lambda^2 Y^2}}\right).$$

Since $Y \sim N(0,1)$, making the change of variable $\lambda^2 x^2 = u$, we get

$$\rho(X,Y) = \operatorname{sign}(\lambda) \times \frac{U(3/2, 2, 1/2\lambda^2)}{2\lambda^2\sqrt{\pi}} \tag{4.9}$$

where $U(a,b,z)$ represents the "Confluent Hypergeometric" function, defined as

$$U(a,b,z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} dt \ , \tag{4.10}$$

in which $b > a > 0$ and $z > 0$. It can be verified numerically that $|\rho(X,Y)| \leq 0.63662$.

Table 1 shows some representative values of the correlation coefficient $\rho(X,Y)$ in (4.9) for the case $\lambda \geq 0$. The corresponding negative values can be obtained by symmetry. This table is useful for estimating λ using the method of moments.

Table 1. Some representative values of $\rho(X,Y)$ in (4.9) for $\lambda \geq 0$.

λ	$\rho(X,Y)$	λ	$\rho(X,Y)$	λ	$\rho(X,Y)$	λ	$\rho(X,Y)$
0.0	0.0000	1.1	0.4674	3.0	0.5862	10	0.6284
0.1	0.0786	1.2	0.4819	3.5	0.5960	11	0.6296
0.2	0.1512	1.3	0.4945	4.0	0.6031	12	0.6306
0.3	0.2145	1.4	0.5056	4.5	0.6084	13	0.6313
0.4	0.2681	1.5	0.5154	5.0	0.6125	14	0.6319
0.5	0.3130	1.6	0.5241	5.5	0.6157	15	0.6324
0.6	0.3506	1.7	0.5318	6.0	0.6183	20	0.6340
0.7	0.3821	1.8	0.5388	6.5	0.6204	40	0.6358
0.8	0.4087	1.9	0.5450	7.0	0.6222	60	0.6362
0.9	0.4313	2.0	0.5507	8.0	0.6249	80	0.6364
1.0	0.4507	2.5	0.5721	9.0	0.6269	∞	0.6366

What do densities of the form (4.6) look like? They can be unimodal or bimodal depending on the value of λ.

Lemma. *If* $|\lambda| \leq \sqrt{\pi/2} \approx 1.25$, *the density (4.6) has a unique mode at the origin,* $(0,0)$. *If* $|\lambda| > \sqrt{\pi/2}$, *the density (4.6) is bimodal. Its modes can be obtained as the solutions (for* (x,y)*) of the system of equations*

$$\lambda\phi(\lambda x^2) - \Phi(\lambda x^2) = 0 \tag{4.11}$$

$$x^2 - y^2 = 0. \tag{4.12}$$

Proof. By differentiating the conditional densities corresponding to (4.6) we find that necessary conditions for a local maximum are

$$\lambda y\phi(\lambda xy) - x\Phi(\lambda xy) = 0 \tag{4.13}$$

$$\lambda x\phi(\lambda xy) - y\Phi(\lambda xy) = 0. \tag{4.14}$$

This leads to the condition $x^2 = y^2$. Consider the case where $x = y$ and make that substitution in (4.13). The resulting function on the left hand side of (4.13) is

$$g(x;\lambda) = \lambda x\phi(\lambda x^2) - x\Phi(\lambda x^2).$$

It can be verified that $g(x;\lambda)$ has 3 real roots if $|\lambda| > \sqrt{\pi/2}$ and only one if $|\lambda| \leq \sqrt{\pi/2}$. Checking the corresponding Hessian matrices it can be shown that the three roots correspond to two maxima and one saddle point.

As an illustration, Figures 1 and 2 show the intersection of the regression curves for the cases $\lambda = 0.5$, and 5 respectively. The following pairs of

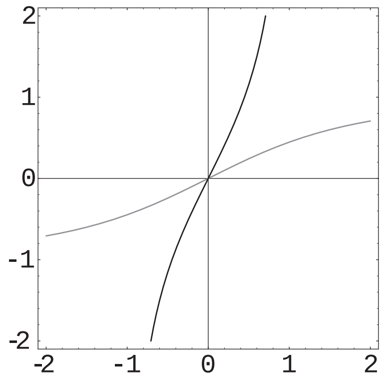

Figure 1: Regression lines of model (4.6) with $\lambda = 0.5$

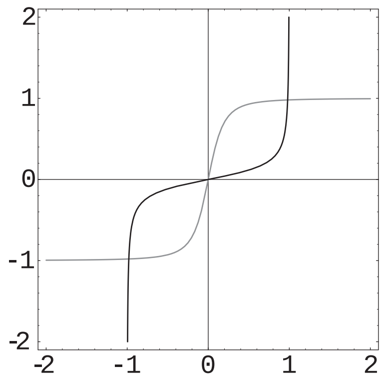

Figure 2: Regression lines of model (4.6) with $\lambda = 5$

Figures 3-4 and 5-6 show the contours and the density of (4.6) for these two values of the parameter λ.

Practical application of the density (4.6) would undoubtedly involve introduction of location and scale parameters leading to a density of the form:

$$f(x, y; \lambda, \mu, \sigma) = \frac{2}{\sigma_1 \sigma_2} \phi \left(\frac{x - \mu_1}{\sigma_1} \right) \phi \left(\frac{y - \mu_2}{\sigma_2} \right) \Phi \left(\lambda \frac{x - \mu_1}{\sigma_1} \times \frac{y - \mu_2}{\sigma_2} \right). \tag{4.15}$$

If a sample $(x_1, y_1) \cdots (x_n, y_n)$ is available from the density (4.15), the corresponding sample moments will provide strongly consistent estimates of means, variances and the correlation between X and Y.

Observe that the marginal densities of X and Y are respectively Normal (μ_1, σ_1^2) and Normal (μ_2, σ_2^2). Since $\lambda = u^{-1}(\rho)$, where $u(\lambda)$ is given by (4.9),

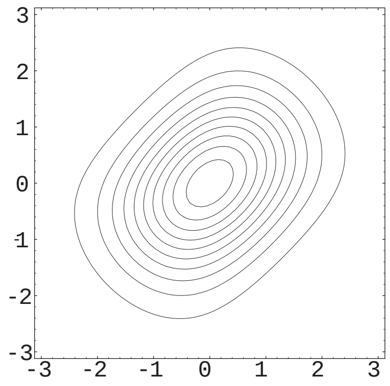

Figure 3: Contours of (4.6) with $\lambda = 0.5$

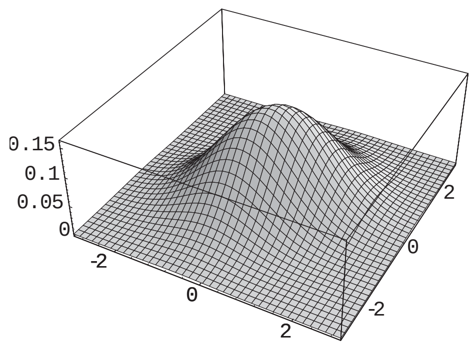

Figure 4: Density function of (4.6) with $\lambda = 0.5$

interpolation in Table 1 will yield an acceptable point estimate of λ based on the sample correlation coefficient $\hat{\rho}$. When λ is large, however, this estimate can be expected to have a large variance.

As a summary of this subsection, we mention that density (4.6) can be used to model two dimensional data with the following characteristics:

- Weakly asymmetric conditional distributions.
- Possible bimodality.
- Moderate values of the correlation coefficient.

4.2 *Linearly skewed normal conditionals.* We wish to identify, if possible, all

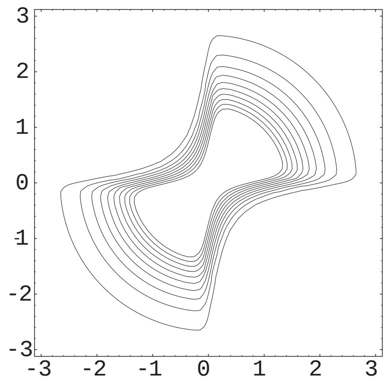

Figure 5: Contours of (4.6) with $\lambda = 5$

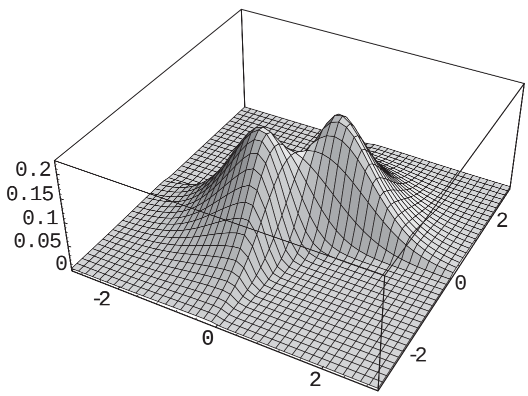

Figure 6: Density function of (4.6) with $\lambda = 5$

joint densities for a two dimensional random variable (X, Y) such that:

$$\text{for each } y \in \mathbb{R},\ X|Y = y \sim LSN(\lambda_0^{(1)}(y), \lambda_1^{(1)}(y)) \tag{4.16}$$

and

$$\text{for each } x \in \mathbb{R},\ Y|X = x \sim LSN(\lambda_0^{(2)}(x), \lambda_1^{(2)}(x))$$

for some functions $\lambda_0^{(1)}(y), \lambda_1^{(1)}(y), \lambda_0^2(x), \lambda_1^{(2)}(x)$. Location and scale parameters will be added later.

If we write the joint density of (X, Y) as a product of marginal and conditional densities in the two possible ways using (4.16) we obtain the following functional equation:

$$\frac{\phi(x)\Phi(\lambda_0^{(1)}(y)+\lambda_1^{(1)}(y)x)f_Y(y)}{\Phi(\lambda_0^{(1)}(y)[1+(\lambda_1^{(1)}(y))^2]^{-1/2})}=\frac{\phi(y)\Phi(\lambda_0^{(2)}(x)+\lambda_1^{(2)}(x)y)f_X(x)}{\Phi(\lambda_0^{(2)}(x)[1+(\lambda_1^{(2)}(x))^2]^{-1/2})} \tag{4.17}$$

where $f_X(x)$ and $f_Y(y)$ denote the marginal densities. The functional equation (4.17) seems difficult to solve in general. It will be discussed further in Section 8. It is possible to find a set of particular solutions to (4.17) that may turn out to be the general solution. In any case, the particular solutions obtained will provide us with a flexible family of joint densities with linearly skewed normal conditionals, i.e., satisfying (4.16). We will use a 3 stage procedure to identify solutions to (4.17) as follows:

- Step 1. We solve the functional equation

  $$\Phi(\lambda_0^{(1)}(y)+\lambda_1^{(1)}(y)x)=\Phi(\lambda_0^{(2)}(x)+\lambda_1^{(2)}(x)y) \tag{4.18}$$

  and obtain the unknown functions $\lambda_i^{(j)}(\cdot),\ i=0,1,\ j=1,2.$

- Step 2. We replace the solutions obtained in Step 1 into (4.17). Then we solve for the remaining two unknowns, $f_X(x)$ and $f_Y(y)$.

- Step 3. Finally, we verify that the resulting functions are density functions.

Since $\Phi(\cdot)$ is a strictly monotone function, (4.18) is equivalent to

$$\lambda_0^{(1)}(y)+\lambda_1^{(1)}(y)x=\lambda_0^{(2)}(x)+\lambda_1^{(2)}(x)y. \tag{4.19}$$

The general solution to (4.19) is (see, for example, Castillo and Ruiz-Cobo (1992, p. 52)):

$$\begin{aligned}\lambda_0^{(1)}(y) &= \lambda_{00}+\lambda_{01}y\\ \lambda_1^{(1)}(y) &= \lambda_{10}+\lambda_{11}y\\ \lambda_0^{(2)}(x) &= \lambda_{00}+\lambda_{10}x\\ \lambda_1^{(2)}(x) &= \lambda_{01}+\lambda_{11}x\end{aligned} \tag{4.20}$$

where $\lambda_{ij},\ i,j=0,1$ are arbitrary real constants. Now for step 2, if we substitute (4.20) back into (4.17) we readily obtain marginal densities of the form:

$$f_X(x)\propto\phi(x)\Phi((\lambda_{00}+\lambda_{10}x)[1+(\lambda_{01}+\lambda_{11}x)^2]^{-1/2}),\ x\in\mathbb{R} \tag{4.21}$$

and

$$f_Y(y) \propto \phi(y)\Phi((\lambda_{00} + \lambda_{01}y)[1 + (\lambda_{10} + \lambda_{11}y)^2]^{-1/2}), \ y \in \mathbb{R}. \quad (4.22)$$

Combining (4.17), (4.20), (4.21) and (4.22) we obtain our class of bivariate densities with linearly skewed normal conditionals in the form:

$$f(x, y; \boldsymbol{\lambda}) \propto \phi(x)\phi(y)\Phi(\lambda_{00} + \lambda_{10}x + \lambda_{01}y + \lambda_{11}xy), \ (x, y) \in \mathbb{R}^2. \quad (4.23)$$

Finally we need to verify that (4.23) is integrable for every choice of $\boldsymbol{\lambda}$. But this is evident since $f(x, y; \boldsymbol{\lambda}) \leq \phi(x)\phi(y), \ \forall x, y \in \mathbb{R}^2$. Note that (4.6) is a special case of (4.23). Unfortunately it is not possible to provide a closed form expression for the normalizing constant in (4.23) (in contrast to the special case (4.6) where the normalizing constant was particularly simple, viz. 2).

The density (4.23) covers a broad spectrum of possibilities. It contains, as examples, the following submodels.

- Product of independent normal densities ($\lambda_{ij} = 0, \ \forall i, j$).
- Model (4.6) with skewed normal conditionals and normal marginals ($\lambda_{00} = \lambda_{01} = \lambda_{10} = 0$).
- A model with skewed normal marginals and linearly skewed normal conditionals ($\lambda_{00} = \lambda_{11} = 0$).
- A model with linearly skewed normal marginals and conditionals ($\lambda_{11} = 0$) (the Arnold-Beaver model (3.2)).

The density (4.23) has at most two modes. The modes lie on the curve

$$\lambda_{01}x - \lambda_{10}y + \lambda_{11}(x^2 - y^2) = 0. \quad (4.24)$$

A representative example of the density and contours corresponding to the density (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$ is provided in Figures 7 and 8. The density (4.23) admits a larger range of values for the correlation than does the model (4.6). However, since the correlation must be evaluated numerically for many values of $\boldsymbol{\lambda}$, we have not yet determined the full range of correlations that are possible.

At this stage, it is natural to complete the model (4.23) by introducing location and scale parameters μ_1, μ_2, σ_1 and σ_2, to arrive at a family of models that is closed under location and scale changes. However, our experience with classical bivariate normal theory might lead us to ask for more. We

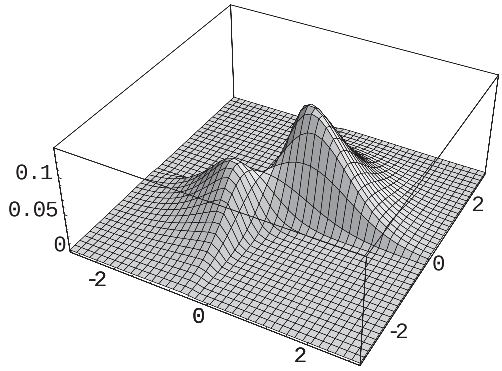

Figure 7: Density function of (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$.

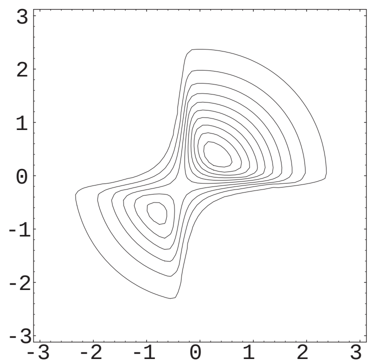

Figure 8: Contours of (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$.

might well ask for a model closed under affine transformations, i.e., under changes of location and scale and rotations.

Rotation of the density (4.23) takes us out of the class of densities with linearly skewed normal conditionals and is the subject of the next section.

5. Rotation and Quadratically Skewed Normal Conditionals

Suppose that we begin with the model (4.23) and apply a rotation and a location and scale transformation. The resulting density will be of the form

$$f(x, y) \propto f_C(x, y)\Phi(a + b_1x + b_2y + c_1x^2 + c_2y^2 + c_3xy) \qquad (5.1)$$

where f_C is a classical bivariate normal density. This joint density will have quadratically (rather than linearly) skewed normal conditionals. The family of densities (5.1) is an 11 parameter family $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, a, b_1, b_1, c_1, c_2, c_3)$ which is closed under changes of location and scale as well as rotations.

The dimension of the parameter space for (5.1) is daunting but in fact the family falls far short of being the most general family with quadratically skewed normal (QSN) conditionals. A classical bivariate normal density weighted by $\Phi(h(x, y))$ will have QSN conditionals provided that h is a quadratic in x for each fixed y and quadratic in y for each fixed x. Thus the following 14 parameter family of densities has QSN conditionals

$$f(x, y) \propto f_C(x, y)\Phi((1, x, x^2)A(1, y, y^2)'), \tag{5.2}$$

where $f_C(x, y)$ is a classical bivariate normal density with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ and where A is 3×3 matrix of real parameters.

But we can't stop here. The Bhattacharyya normal conditionals density (1.1) can be used instead of the classical normal density in (5.2). This leads us to the following very general family of densities with QSN conditionals.

$$f(x, y) \propto \exp[(1, x, x^2)A_1(1, y, y^2)']\Phi((1, x, x^2)A_2(1, y, y^2)'). \tag{5.3}$$

The parameter space for (5.3) is of dimension 17.

We remark in passing, that the Bhattacharyya normal density can be used to extend (4.28) to a quite broad class of densities with linearly skewed normal conditionals. This family will be of the form

$$f(x, y) \propto \exp((1, x, x^2)A(1, y, y^2)')\Phi(\lambda_{00} + \lambda_{10}x + \lambda_{01}y + \lambda_{11}xy). \tag{5.4}$$

The family (5.4) (just as was the case for (4.28)) is closed under location and scale changes, but *not* under rotations.

6. Multivariate Extensions

We have restricted discussion to the bivariate case, but extensions to higher dimension can be readily accomplished using suitable notation. For a random variable $\boldsymbol{X}$ of dimension k, we define $(k-1)$ dimensional subvectors $\boldsymbol{X}_{(1)}, \ldots, \boldsymbol{X}_{(k)}$ such that, for each i, $\boldsymbol{X}_{(i)}$ is the vector $\boldsymbol{X}$ with the i'th coordinate X_i deleted. Analogously for a real vector $\boldsymbol{x}$ we define $\boldsymbol{x}_{(i)}$. We will say that a random vector $\boldsymbol{X}$ has skewed normal conditionals (SN conditionals) if for each $\boldsymbol{x}_{(i)} \in \mathbb{R}^{k-1}$ and for each i,

$$X_i | \boldsymbol{X}_{(i)} = \boldsymbol{x}_{(i)} \sim SN(\lambda_i(\boldsymbol{x}_{(i)})) \tag{6.1}$$

Analogously we define k-dimensional densities with linearly skewed normal (LSN) and quadratically skewed normal (QSN) conditionals. The k-dimensional analog of (4.6) with SN conditionals is of the form

$$f(x_1,\ldots,x_k;\lambda) = 2\left[\prod_{i=1}^{k}\phi(x_i)\right]\Phi\left(\lambda\prod_{i=1}^{k}x_i\right);\ \boldsymbol{x}\in\mathbb{R}^k. \tag{6.2}$$

In this case

$$\lambda_i(\boldsymbol{x}_{(i)}) = \lambda\prod_{j\neq i}x_j,\quad i=1,\ldots,k. \tag{6.3}$$

The modes of the density are the solutions of the system:

$$\lambda\left[\prod_{j\neq i}x_j\right]\phi\left(\lambda\prod_{i=1}^{k}x_i\right) - x_i\Phi\left(\lambda\prod_{i=1}^{k}x_i\right) = 0;\quad i=1,\ldots,k. \tag{6.4}$$

Introducing location and scale parameters we get the general model

$$f(x_1,\ldots,x_k;\lambda,\boldsymbol{\mu},\boldsymbol{\sigma}) = 2\left[\prod_{i=1}^{k}\frac{1}{\sigma_i}\phi\left(\frac{x_i-\mu_i}{\sigma_i}\right)\right]\Phi\left(\lambda\prod_{i=1}^{k}\frac{x_i-\mu_i}{\sigma_i}\right);\ \boldsymbol{x}\in\mathbb{R}^k. \tag{6.5}$$

The k-dimensional analog of (4.28) with LSN conditionals is given by:

$$f(x_1,x_2,\ldots,x_k)\propto\left[\prod_{i=1}^{k}\phi(x_i)\right]\Phi\left(\sum_{\boldsymbol{s}\in S_k}\lambda_{\boldsymbol{s}}\prod_{i=1}^{k}x_i^{s_i}\right) \tag{6.6}$$

where S_k is the set of all vectors of 0's and 1's of dimension k. Location and scale parameters can then be used to extend model (6.6), retaining LSN conditionals.

It is a straightforward matter to write down k-dimensional analogs for (5.1), (5.2) and (5.3). The most general of these, extending (5.3) is given by

$$f(x_1,x_2,\ldots,x_k)\propto\exp\left[\sum_{\boldsymbol{t}\in T_k}a_{\boldsymbol{t}}^{(1)}\prod_{i=1}^{k}x_i^{t_i}\right]\Phi\left[\sum_{\boldsymbol{t}\in T_k}a_{\boldsymbol{t}}^{(2)}\prod_{i=1}^{k}x_i^{t_i}\right], \tag{6.7}$$

where T_k is the set of all vectors of 0's,1's and 2's of dimension k. The model (6.7) has QSN conditionals and is closed under changes of location and scale and under general rotations in k-space. It is daunting to note the dimension of the parameter space for (6.7). It is $2(3^k)-1$.

7. Non-normal Variants

In much of the discussion in the earlier sections, the roles played by ϕ and Φ (the standard normal density and distribution function) could just as well be played by other densities and distributions. See, for example, Arnold and Beaver (2000) who discuss a variety of non-normal distributions skewed by hidden truncation. For notational simplicity let us return to the bivariate setting, realizing that higher dimensional extensions can be readily accomplished.

In equation (4.23), the functions ϕ and Φ can be replaced by ψ and Ψ, the standard Cauchy density and distribution function, respectively (a similar substitution can be made in (4.6)). In this manner a linearly skewed Cauchy conditionals density can be obtained. Parallel models involving the standard Laplace and logistic distribution can also be developed.

In Figures 9, 10 and 11, we present representative contours for the linearly skewed (LS) Cauchy conditionals, LS Laplace conditionals and LS logistic conditionals distributions. In all cases we have taken $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$, to facilitate comparison with the LS normal conditionals contours displayed in Figure 7.

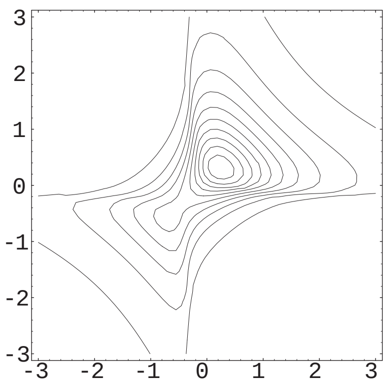

Figure 9: Contours of the linearly skewed Cauchy conditionals (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$.

Of course, one can be even more flexible in constructing linearly skewed distributions. We could let $f_1(x), f_2(y)$ be two quite arbitrary densities and let H be a quite arbitrary distribution function. We then construct linearly skewed joint densities of the form

$$f(x, y) \propto f_1(x)f_2(y)H(\lambda_{00} + \lambda_{10}x + \lambda_{01}y + \lambda_{11}xy) \tag{7.1}$$

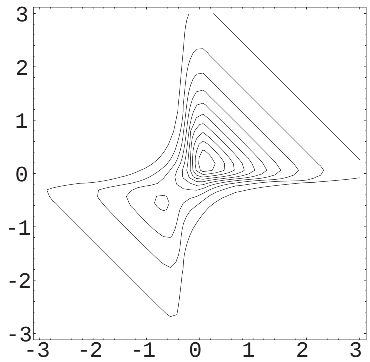

Figure 10: Contours of the linearly skewed Laplace conditionals (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$.

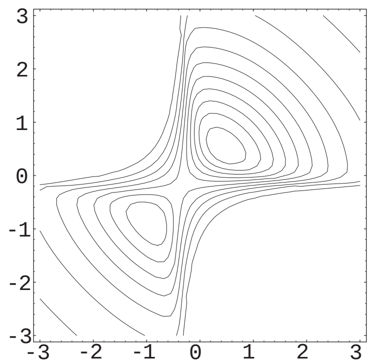

Figure 11: Contours of the linearly skewed Logistic conditionals (4.23) with $\lambda_{00} = 0, \lambda_{01} = 2, \lambda_{10} = 1$ and $\lambda_{11} = 6$.

8. **Functional Equations**

During the discussion presented in Section 4, two non-trivial functional equations were considered. Particular but not general solutions were provided. In this section, we formulate the open questions relating to these equations.

First consider equation (4.2). It will be convenient to use slightly different notation so that our equation is of the form

$$2\phi(x)\Phi(a(y)x)f_Y(y) = 2\phi(y)\Phi(b(x)y)f_X(x) \tag{8.1}$$

and we wish to solve for $a(y)$ and $b(x)$ (as well as $f_X(x)$ and $f_Y(y)$). First set $x = 0$ in (8.1). This yields

$$2\phi(0)\Phi(0)f_Y(y) = 2\phi(y)\Phi(b_0 y)f_X(0) \tag{8.2}$$

where $b_0 = b(0)$. From this we obtain

$$f_Y(y) = 2\phi(y)\Phi(b_0 y). \tag{8.3}$$

Analogously setting $y = 0$ in (8.1) we find

$$f_X(x) = 2\phi(x)\Phi(a_0 x) \tag{8.4}$$

where $a_0 = a(0)$. Substituting (8.3) and (8.4) back in (8.1) and simplifying yields the following functional equation to be solved

$$\Phi(a(y)x)\Phi(b_0 y) = \Phi(b(x)y)\Phi(a_0 x). \tag{8.5}$$

Of course if a_0 and b_0 are both zero, (8.5) can be readily solved as in Section 4, to get $a(y) = \lambda y$ and $b(x) = \lambda x$ for some $\lambda \in \mathbb{R}$. In an effort to solve (8.5) when a_0 and b_0 are not both zero, we can set $x = 1$ in (8.5). This yields

$$\Phi(a(y))\Phi(b_0 y) = \Phi(b_1 y)\Phi(a_0).$$

From this we obtain

$$a(y) = \Phi^{-1}\left(\frac{\Phi(b_1 y)}{\Phi(b_0 y)}\Phi(a_0)\right), \tag{8.6}$$

where $b_1 = b(1)$. Analogously, setting $y = 1$ in (8.5), we get

$$b(x) = \Phi^{-1}\left(\frac{\Phi(a_1 x)}{\Phi(a_0 x)}\Phi(b_0)\right). \tag{8.7}$$

where $a_1 = a(1)$.

If $a_0 = a_1$ then $b(x) \equiv b_0$ which implies eventually that the random variables X and Y are independent skewed normal variables. A similar conclusion can be reached if $b_0 = b_1$. If $a_0 \neq a_1$ and $b_0 \neq b_1$ then from (8.6) or (8.7) we must have

$$\Phi(a_1)/\Phi(a_0) = \Phi(b_1)/\Phi(b_0). \tag{8.8}$$

If we make choices for a_0, a_1, b_0 and b_1 subject to the constraint (8.8) and $a_0 \neq a_1, b_0 \neq b_1$ we can substitute the resulting expressions (8.6) and (8.7) back into (8.5) to obtain:

$$\Phi\left(x\Phi^{-1}\left(\frac{\Phi(b_1 y)}{\Phi(b_0 y)}\Phi(a_0)\right)\right)\Phi(b_0 y)$$

$$= \Phi\left(y\Phi^{-1}\left(\frac{\Phi(a_1x)}{\Phi(a_0x)}\right)\right)\Phi(a_0x). \tag{8.9}$$

We conjecture but are unable to prove that no solution exists to (8.9) unless $a_0 = b_0 = 0$ (the case already resolved in Section 4).

The second recalcitrant functional equation is (4.17). Again we will introduce slightly different notation to reduce the number of subscripts and superscripts in the discussion. The equation to be solved is equivalent to

$$\Phi(a(y) + b(y)x)g_1(y) = \Phi(c(x) + d(x)y)g_2(x)\ , \tag{8.10}$$

to be solved for $a(y), b(y), c(x)$ and $d(x)$ and subsequently for $g_1(y)$ and $g_2(x)$. We can rephrase (8.10) as follows. We need to find functions a, b, c and d such that the ratio

$$\frac{\Phi(a(y) + b(y)x)}{\Phi(c(x) + d(x)y)} \tag{8.11}$$

factors into a function of x times a function of y. The easy solution occurs when the ratio (8.10) is identically equal to one (this is the solution described in Section 4). Here again, we conjecture but are unable to prove that no other solutions exist.

Remark: Throughout this section it may be observed that the only properties of Φ and $\phi = \Phi'$ that were used, were that Φ is strictly monotone on $\mathbb{R}$ and $\Phi(0) = 1/2$. Consequently similar observations and conjectures could be made for Cauchy, Laplace and Logistic models, etc.

9. Polynomially Skewed Normal Models

If we are willing to countenance polynomial weight functions of degree higher than two we can construct skewed models with any desired number of modes. Thus in general we can begin with a basic density f_0 and a possibly unrelated distribution function G_0 and define a polynomially skewed density of the form:

$$f(x) \propto f_0(x)G_0\left(\sum_{i=0}^{k}\lambda_i x^i\right). \tag{9.1}$$

A bivariate version of such a density would be given by

$$f(x,y) \propto f_0(x)f_0(y)G_0(\mathcal{P}_{k_1,k_2}(x,y)) \tag{9.2}$$

where $\mathcal{P}_{k_1,k_2}(x,y)$ is, for each fixed y, a polynomial of degree k_1 in x, and for each fixed x, a polynomial of degree k_2 in y. In particular we might choose

f_0 and G_0 to be ϕ and Φ, the standard normal density and distribution respectively. For example, Figure 12 shows the polynomially skewed normal density

$$f(x) \propto \phi(x)\Phi(x(x-1)(x+1)). \tag{9.3}$$

Figure 13 and 14 show the densities and contours of the densities

$$f(x,y) \propto \phi(x)\phi(y)\Phi\left[(x-1)^2(x+1)^2(y-1)(y+1)\right] \tag{9.4}$$

and

$$f(x,y) \propto \phi(x)\phi(y)\Phi\left[(x-1)^2(x+1)^2(y-1)^2(y+1)^2\right] \tag{9.5}$$

respectively.

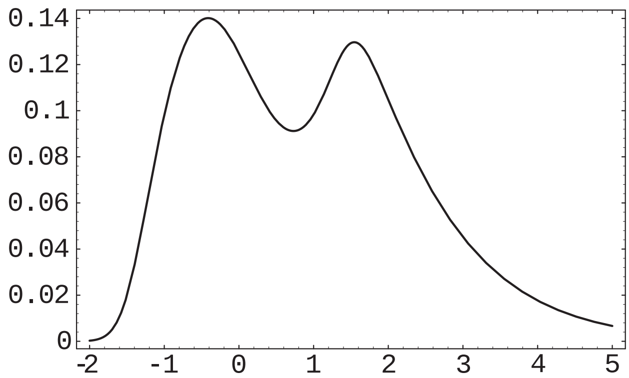

Figure 12: Polynomially skewed normal density (9.3).

References

Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models. *Sankhyā Ser. A*, **62**, 22-35.

Arnold, B.C., Beaver, R.J., Groeneveld, R.A. and Meeker, W.Q. (1993). The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-478.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999). *Conditional Specification of Statistical Models.* Springer Series in Statistics, Springer-Verlag, New York.

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.

Bhattacharyya, A. (1944). On some sets of sufficient conditions leading to the normal bivariate distribution. *Sankhyā*, **6** 399-406.

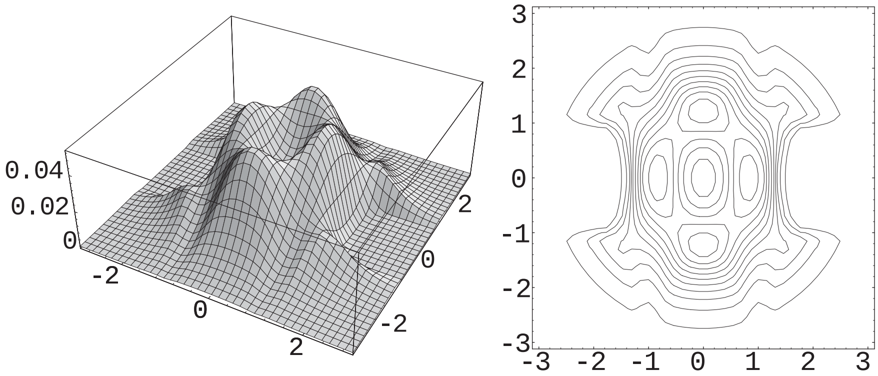

Figure 13: Density and contours of (9.4).

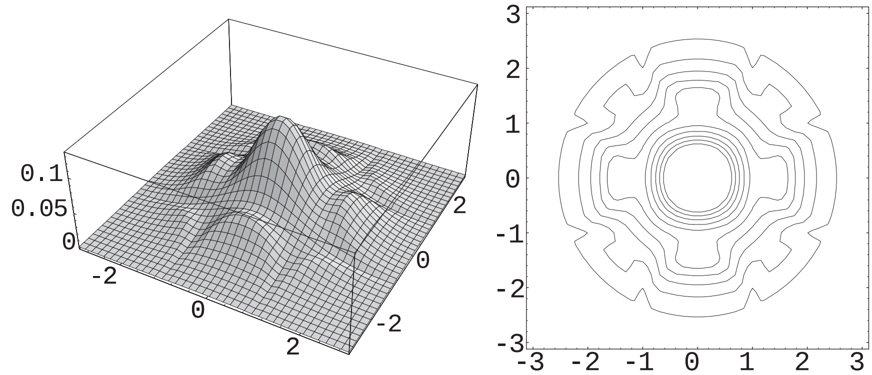

Figure 14: Density and contours of (9.5).

Castillo, E. and Ruiz-Cobo, R. (1992). *Functional Equations in Science and Engineering.* Marcel Dekker, New York, 1992.

Barry C. Arnold
Department of Statistics
University of California
Riverside, USA
E-mail: barry_arnold@ucr.edu

José María Sarabia
Department of Economics
University of Cantabria
Santander, Spain.
E-mail: sarabiaj@unican.es

Enrique Castillo
Dept. of Applied Mathematics
and Computing Sciences
University of Cantabria
Santander, Spain
and
University of Castilla, La Mancha
Ciudad Real, Spain
E-mail: castie@unican.es
castie@ccp-cr.uclm.es