

Marginal Replacement in Multivariate Densities, with Application to Skewing Spherically Symmetric Distributions

M. C. Jones

The Open University, Walton Hall, Milton Keynes, United Kingdom
E-mail: m.c.jones@open.ac.uk

Received November 4, 1999; published online November 13, 2001

We consider a simple general construct, that of marginal replacement in a multivariate distribution, that provides, in particular, an interesting way of producing distributions that have a skew marginal from spherically symmetric starting points. Particular examples include a new multivariate beta distribution and a new multivariate t /skew t distribution, along with Azzalini and colleagues' multivariate skew normal distribution. © 2001 Elsevier Science (USA)

AMS 2000 subject classification: 60E05.

Key words and phrases: multivariate beta distribution; multivariate t distribution; skew normal distribution; skew t distribution; transformation.

1. INTRODUCTION

This paper is concerned with a very simple approach to modifying given multivariate densities to produce new multivariate densities with one or more specified marginals. Partition a p -dimensional vector, $p \geq 2$, into (\mathbf{x}, \mathbf{y}) where \mathbf{x} is d -dimensional, $1 \leq d \leq p$, and \mathbf{y} is $(p-d)$ -dimensional. Let the corresponding continuous random variables \mathbf{X} and \mathbf{Y} follow the p -variate distribution with density function $f(\mathbf{x}, \mathbf{y})$ and let this distribution have \mathbf{X} -marginal $f_{\mathbf{X}}(\mathbf{x})$. Suppose it is desired to obtain a new p -variate density $f_1(\mathbf{x}, \mathbf{y})$ with specified \mathbf{X} -marginal $g(\mathbf{x})$, say. An example of this might be producing a distribution with a skew marginal from a symmetric multivariate starting point; this is the motivation for the current work, described below.

Clearly, we can write $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y} | \mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$ using obvious notation for the conditional density. Then, modulo support considerations,

$f_1(\mathbf{x}, \mathbf{y}) = f(\mathbf{y} | \mathbf{x}) g(\mathbf{x})$ is obviously another multivariate density, but with the \mathbf{X} -marginal we desire. Going straight to f_1 without reference to f by specifying marginal and conditional densities in this way is a standard technique. The point here is that even when it is not especially natural to think in terms of conditional distributions, but rather in terms of bivariate or multivariate distributions with variables on an equal footing, this marginal replacement scheme is still available, most usefully written in the multiplicative form

$$f_1(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}) f(\mathbf{x}, \mathbf{y}) / f_X(\mathbf{x}). \quad (1)$$

The only condition on this approach is that the support of g be contained in, or equal to, the support of f_X . Indeed, f_1 then has support contained in, or equal to, respectively, the support of f . This approach seems to be quite often utilised for discrete distributions but not for continuous distributions as here.

As well as giving the desired \mathbf{X} -marginal, this approach has the benefit of retaining the same conditional distributions of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$ as the original distribution, both characteristics by construction. Properties of distributions that depend on conditional distributions only will thus be unaffected by the change. In the bivariate case, an example is that the local dependence function [5, 8] is invariant. The \mathbf{Y} -marginal of f_1 , of course, differs from that of f .

This approach contrasts with, and is largely complementary to, the obvious alternative approach, at least when $d = 1$, of transforming the X variable using $X_1 = G^{-1}(F_X(X))$, where G and F_X are the distribution functions associated with g and f_X , respectively. This, too, results in marginal density $g(x)$, but the conditional distribution of $\mathbf{Y} | X$ is changed; the marginal distribution of \mathbf{Y} , however, is unaffected. Approach (1) is, however, often easier to apply, being involved with density functions only rather than distribution functions. It also extends immediately to $d > 1$, while transformation, although readily available for independent components of \mathbf{X} , is not so easily extendable for general f_X and g . Transformation, on the other hand, does allow an expansion in the support of X , and it also sometimes more readily admits of meaningful interpretation.

In Section 2, formula (1) is applied to the task of introducing a univariate skew marginal to a spherically symmetric distribution. Various properties associated with the spherically symmetric distribution are retained; these include its local dependence function which is discussed there. New multivariate distributions based on the beta, Section 3, and t /skew t [9, 10], Section 4, are the particular proposals given in the following two sections. A further new distribution, with an extreme value marginal distribution introduced to a multivariate t distribution, is described in

the brief Section 5 to emphasise the generality of our approach. We provide formulae for general p , but are usually most interested in the case $p = 2$. This is because, with the symmetric marginals of the spherically symmetric distribution being the same as one another, it is unclear to what extent they are of practical interest for fitting to data when $p > 2$; a bivariate distribution with one skew and one symmetric marginal is, however, of potentially more practical interest.

In Section 6, we look briefly at one existing distribution of this type, the popular multivariate skew-normal distribution of Azzalini and colleagues [2, 3], and make some further remarks. One particular point, of course, is that by rotating f_1 , each co-ordinate marginal can be made skew; another is that general location and scale parameters can readily be introduced. Indeed, the whole of the current paper can be seen as an interpretation of the way in which the multivariate skew-normal distribution arises, and hence its extension to introducing skew marginals into other symmetric distributions, and ultimately much further.

2. A SKEW MARGINAL INTRODUCED TO A SPHERICALLY SYMMETRIC DISTRIBUTION

Let f be a spherically symmetric distribution i.e. one with density of the form $g(\mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y})$ for some function g defined on $[0, \infty)$. See Fang, Kotz and Ng [4] for an excellent review of spherically symmetric distributions which can be consulted for any of the results about such distributions used here.

By construction (1), we can introduce as \mathbf{X} marginal any density with the same support, or smaller, as the \mathbf{X} marginal of f . Moreover, conditionally on $\mathbf{X} = \mathbf{x}$, \mathbf{Y} simply has the same symmetric conditional distribution that it has under f . The symmetric nature of these conditional distributions means that the marginal distribution of \mathbf{Y} is also symmetric.

All spherically symmetric distributions have zero correlations. The diagonal nature of the covariance matrix of the underlying spherically symmetric distribution is retained in f_1 . This is because, for any one-dimensional elements Y of \mathbf{Y} and X of \mathbf{X} , $E(Y|X)$ remains zero and hence $E(Y)$ and $E(XY)$ are zero also. Note that the variances will, in general, change but uncorrelatedness remains.

It is well known that the only spherically symmetric distribution for which random variables are independent as well as uncorrelated is the standard normal. It may not be generally appreciated that the uncorrelatedness of other spherically symmetric distributions disguises what can be a

considerable dependence between random variables. A good way of seeing this is to concentrate on any bivariate marginal, and to look at its local dependence function $\gamma(x, y) \equiv \partial \log f(x, y) / \partial x \partial y$ (Holland and Wang [5], Jones [8]). Since $f(x, y) = g(x^2 + y^2)$, we immediately find that

$$\gamma(x, y) = 4xy(\log g)''(x^2 + y^2).$$

Thus, if g is log-convex or log-concave, as is the case for many standard spherically symmetric distributions, γ exhibits a dependence with one particular sign in the positive and negative quadrants and precisely the negative of that dependence in the other two quadrants. In fact, g log-convex results in negative association between $|X|$ and $|Y|$, see Section 3 for an example, while g log-concave gives positive association between $|X|$ and $|Y|$, Section 4. Note that in the bivariate case g can be interpreted as the density of the squared modulus $R^2 = X^2 + Y^2$.

This local dependence structure is maintained for distributions of the form (1) based on spherically symmetric distributions.

3. A NEW MULTIVARIATE BETA DISTRIBUTION, WITH A SINGLE SKEW MARGINAL

There are surprisingly few multivariate distributions with beta marginals [7, Chapter 40] in the literature. The Dirichlet is one such, but it is restrictive in terms of relationships between beta marginals; it also has the simplex as its sample space. Distributions more fully in the cube include the case of independence, and of the general, but in this case clumsy, method of marginal transformation of copulae e.g. [6]. Spherically symmetric beta distributions have support the interior of the sphere, and it is an extension of this distribution, which has the same support, that we consider here.

The p -variate spherically symmetric beta distribution with parameter $b > (p-1)/2$ has density

$$\frac{\Gamma(b+1/2)}{\Gamma(b-(p-1)/2) \pi^{p/2}} (1 - x_1^2 - \cdots - x_p^2)^{b-(p+1)/2}, \quad 0 < x_1^2 + \cdots + x_p^2 < 1,$$

e.g. [6, p. 128] or [4, Section 3.4] where the distribution is called the symmetric multivariate Pearson Type II distribution.

Each marginal density is the univariate symmetric beta with parameter b which has density

$$\frac{1}{B(b, 1/2)} (1 - x_1^2)^{b-1}, \quad -1 < x_1 < 1, \quad (2)$$

where $B(a, b)$ denotes the beta function. Note that it is convenient to use the beta distribution on support $[-1, 1]$; many readers will be more familiar with the beta distribution on $[0, 1]$, for which simply make the linear transformations $(X_i + 1)/2$.

To obtain an asymmetric univariate beta density with parameters $a > 0$ and $c > 0$, say, simply multiply (2) by $(1 + x_1)^{a-b} (1 - x_1)^{c-b}$ and renormalise accordingly to obtain the resulting density.

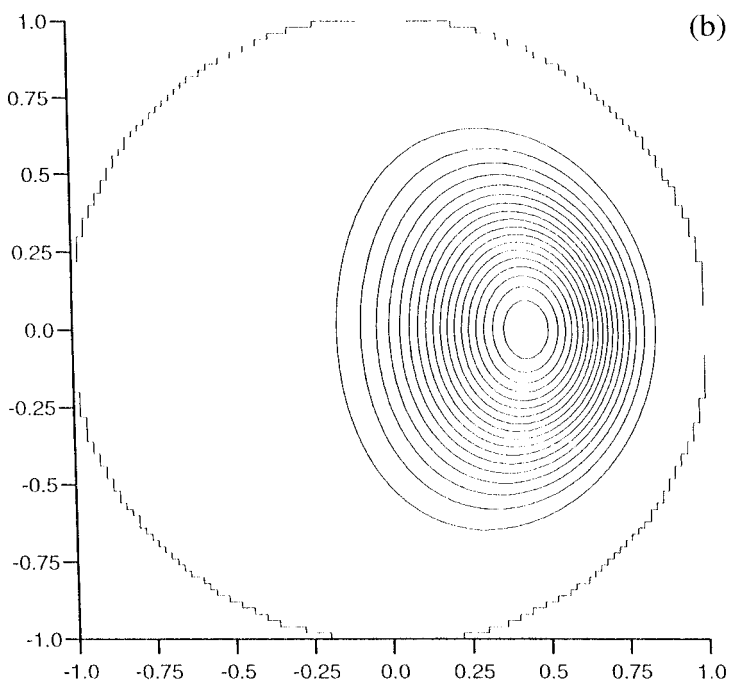
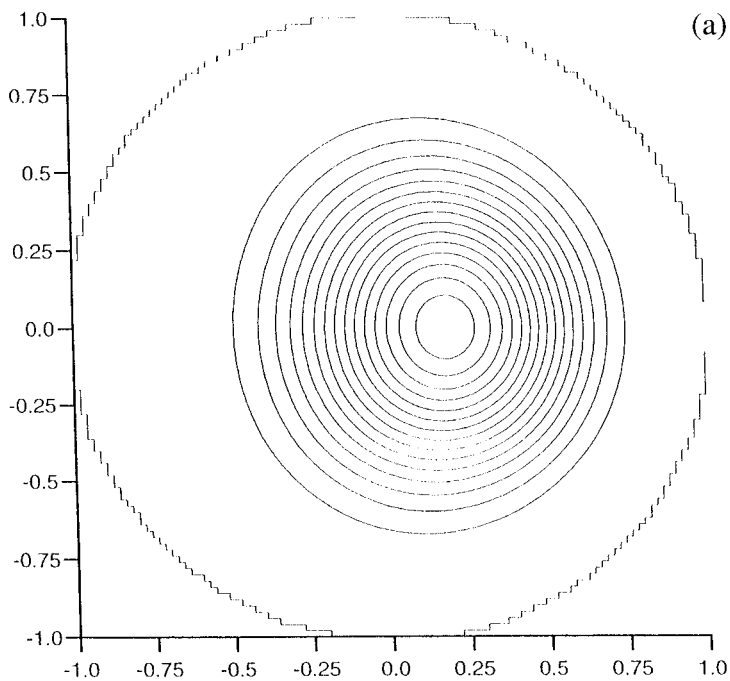
$$\frac{1}{B(a, c) 2^{a+c-1}} (1 + x_1)^{a-1} (1 - x_1)^{c-1}, \quad -1 < x_1 < 1.$$

We therefore propose the multivariate beta distribution with density

$$\frac{\Gamma(a+c) \Gamma(b)}{\Gamma(a) \Gamma(c) \Gamma(b-(p-1)/2) 2^{a+c-1} \pi^{(p-1)/2}} \times (1 + x_1)^{a-b} (1 - x_1)^{c-b} (1 - x_1^2 - \dots - x_p^2)^{b-(p+1)/2}, \quad (3)$$

$0 < x_1^2 + \dots + x_p^2 < 1$, $a > 0$, $b > (p-1)/2$, $c > 0$. By construction, this has as X_1 marginal the Beta(a, c) distribution on $(-1, 1)$. Moreover, conditionally on $X_1 = x_1$, each X_i simply has a symmetric beta distribution with parameter $b-1/2$, rescaled to the interval $(-\sqrt{1-x_1^2}, \sqrt{1-x_1^2})$, as has the original spherically symmetric distribution. The marginal distribution of X_i , $i = 2, \dots, p$, is also symmetric.

The resulting density shapes in the bivariate case are illustrated in Figs. 1 and 2. In Fig. 1, contour plots of three such densities, each with $b = c = 5$ and with $a-b = 2, 7$ and 20 , respectively, are shown. The X_1 marginals are, therefore, Beta(7,5), Beta(12,5) and Beta(25,5), respectively. The conditional distributions of $X_2 | X_1$ are rescaled Beta(5,5)s. Figure 2 is similarly based on the spherically symmetric beta distribution with parameter 5, but shows other X_1 marginal densities: Beta(3,3), Beta(3,7) and Beta(7,7), respectively.



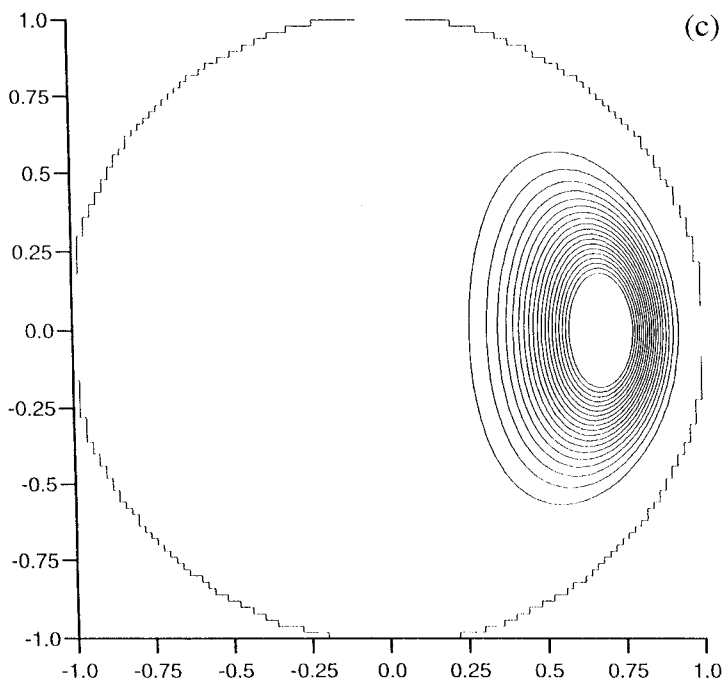


FIG. 1. Bivariate beta densities (3) for $b = c = 5$ and (a) $a = 7$, (b) $a = 12$ and (c) $a = 25$, respectively.

The X_2 marginal distribution is a (symmetric) scale mixture of $\text{Beta}(b, b)$ distributions which can be well approximated, at least provided neither a nor c is too small, by a single symmetric beta distribution with the same variance. Details are not given to save space.

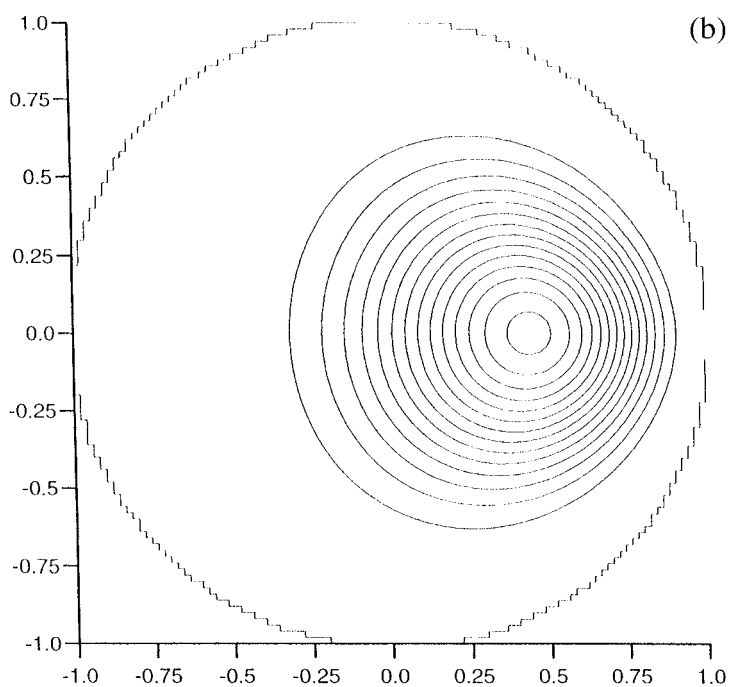
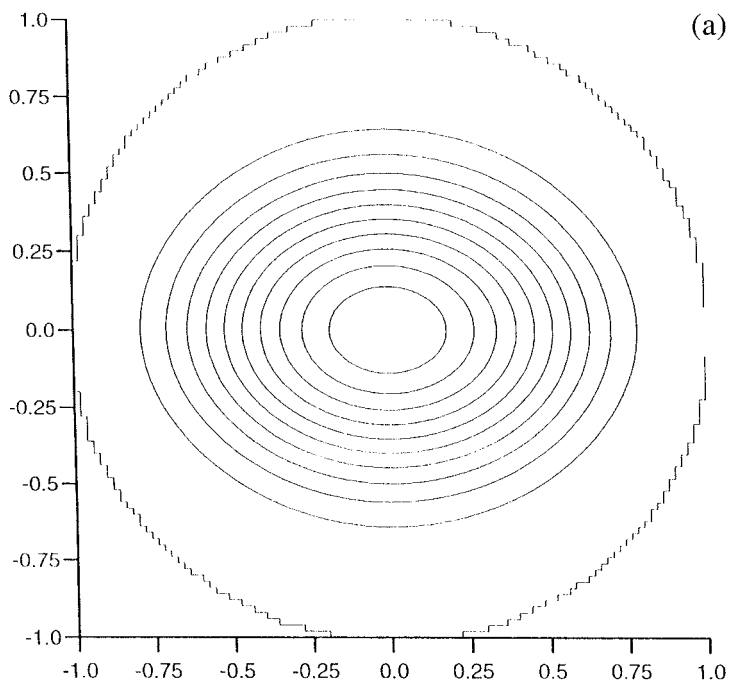
After a little manipulation, the conditional mean of X_1 given $X_2 = x_2, \dots, X_p = x_p$ can be written as $P(1, a-b, c-b, b-(p+1)/2)/P(0, a-b, c-b, b-(p+1)/2)$ where

$$P(Q, R, S, T) = \int_{-1}^1 (ku)^Q (1+ku)^R (1-ku)^S (1-u^2)^T du.$$

This depends on x_2, \dots, x_p only through $k \equiv \sqrt{1-x_2^2-\dots-x_p^2}$. It is not difficult to see that $E(X_1 | X_2, \dots, X_p)$ is positive/zero/negative (for all $0 < k < 1$) whenever $a > / = / < c$.

As previously mentioned, the diagonal nature of the covariance matrix of the underlying spherically symmetric beta distribution is retained in (3).

Also, the local dependence function associated with every density in Figs. 1 and 2 is that associated with the $\text{Beta}(5,5)$ distribution, shown in



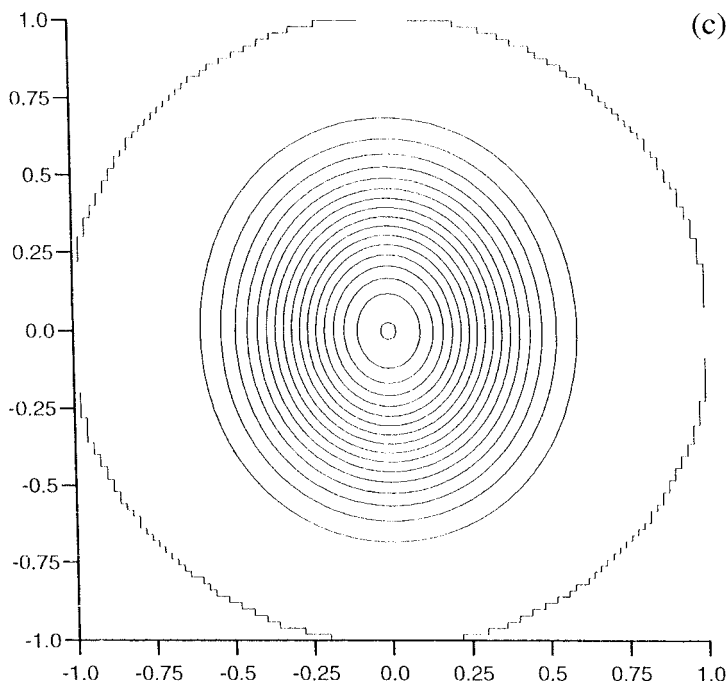


FIG. 2. Bivariate beta densities (3) for $b=5$ and (a) $a=c=3$, (b) $a=7$, $c=3$ and (c) $a=c=7$, respectively.

Fig. 3. The beta distribution is a case for which g is log-convex for $b > 3/2$ —and log-concave otherwise—and hence when $b=5$ there is negative dependence in the positive and negative quadrants and a matching positive dependence elsewhere, corresponding to a negative association between $|X_1|$ and $|X_2|$.

4. A NEW MULTIVARIATE t DISTRIBUTION, WITH A SINGLE SKEW t MARGINAL

The standard multivariate t distribution with parameter $\nu > 0$ is a spherically symmetric distribution on the whole of \Re^p with density

$$\frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2) (\nu\pi)^{p/2}} \frac{1}{(1 + \nu^{-1}(x_1^2 + \cdots + x_p^2))^{(\nu+p)/2}} \quad (4)$$

e.g. [4, Section 3.3, 7, Chapter 37]. Its univariate marginals are ordinary symmetric Student's t distributions with density

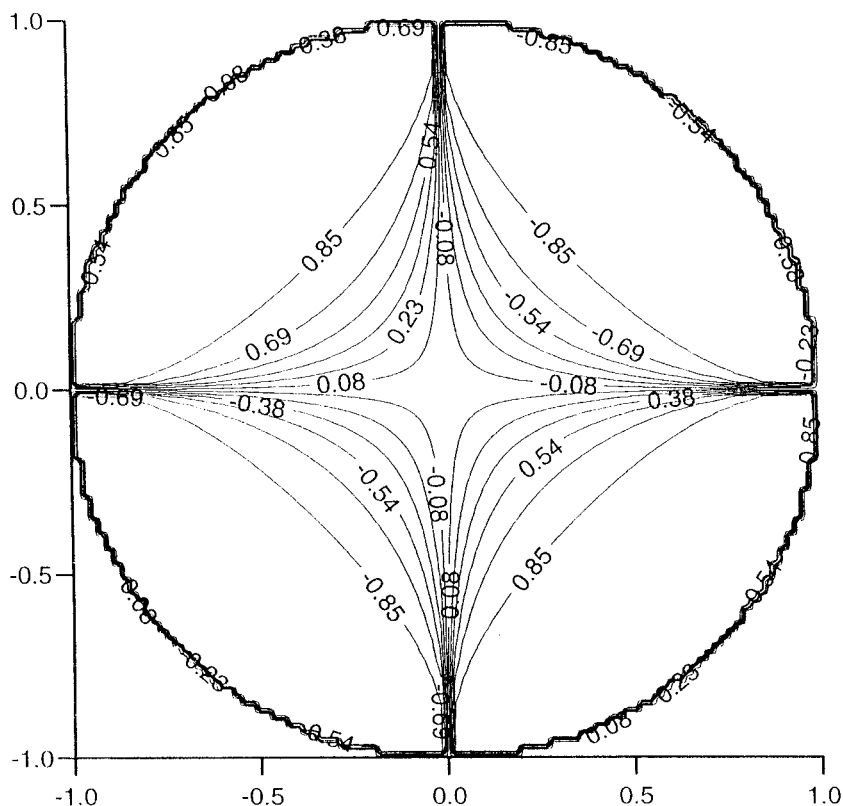


FIG. 3. The local dependence function associated with the Beta(5,5) density and all the densities in Figs. 1 and 2.

$$\frac{1}{B(v/2, 1/2) v^{1/2}} \frac{1}{(1 + v^{-1} x_1^2)^{(v+1)/2}}. \quad (5)$$

Jones [9] introduced a univariate skew t distribution with parameters $a > 0$ and $c > 0$ which has density

$$\frac{1}{B(a, c) (a+c)^{1/2} 2^{a+c-1}} \left(1 + \frac{x_1}{(a+c+x_1^2)^{1/2}}\right)^{a+1/2} \left(1 - \frac{x_1}{(a+c+x_1^2)^{1/2}}\right)^{c+1/2}; \quad (6)$$

this reduces to the t distribution on $2a$ degrees of freedom when $a = c$. For much more about this distribution see [10].

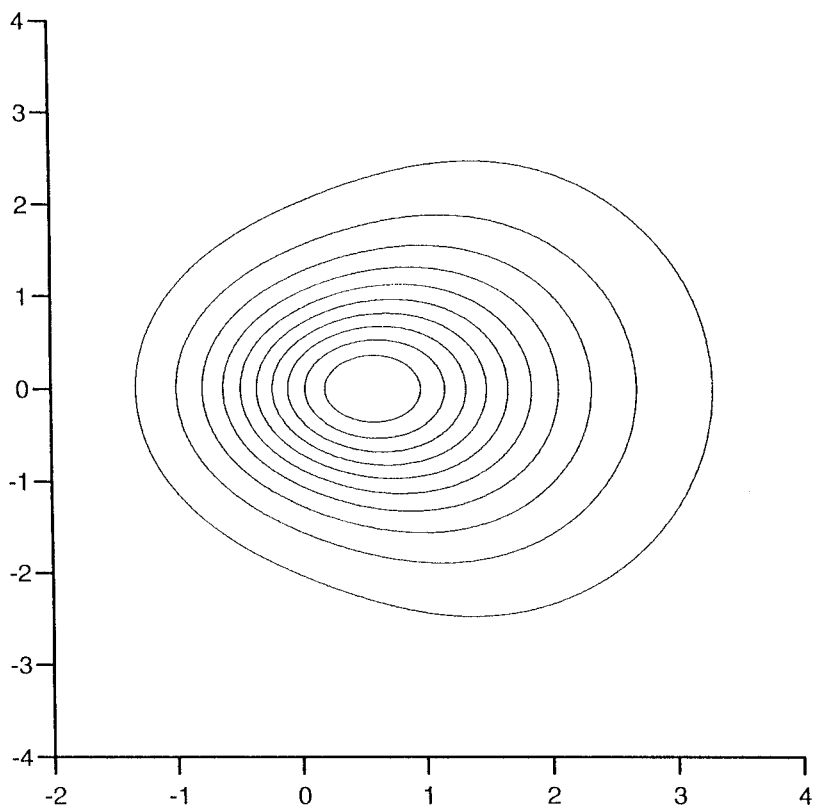


FIG. 4. The bivariate t /skew t density (7) with $a = 4$, $v = 3$ and $c = 2$.

Multiplying (4) by (6) and dividing by (5), as per the recipe in (1), yields the multivariate distribution at the centre of this section. Let us call it a t /skew t distribution. This has density

$$\frac{\Gamma((v+p)/2)}{\Gamma((v+1)/2) B(a, c) (a+c)^{1/2} 2^{a+c-1} (v\pi)^{(p-1)/2}} \times \frac{(1+v^{-1}x_1^2)^{(v+1)/2} \left(1 + \frac{x_1}{(a+c+x_1^2)^{1/2}}\right)^{a+1/2} \left(1 - \frac{x_1}{(a+c+x_1^2)^{1/2}}\right)^{c+1/2}}{(1+v^{-1}(x_1^2 + \dots + x_p^2))^{(v+p)/2}}. \quad (7)$$

Here, a , c and v are all positive; this is (4) if $a = c = v/2$.

In the bivariate case, then, (7) is a distribution with: (i) a skew t marginal with parameters a and c in the X_1 direction; (ii) conditional distributions for $X_2|X_1$ which match those of the bivariate t distribution being t distributions on $\nu+1$ degrees of freedom scaled by a factor of $\{(\nu+1)^{-1}(x_1^2+\nu)\}^{1/2}$; (iii) diagonal covariance matrix; and (iv) the local dependence function of the bivariate t distribution, a special case of which is given as Fig. 2 of [8]. The t distribution is a case for which g is log-concave and hence there is positive association between absolute values of random variables, for any $\nu > 0$. The (symmetric) X_2 marginal of this distribution can be well approximated by a t distribution with the same variance. The conditional distribution of X_1 given X_2, \dots, X_p has similar qualitative properties as that for the beta distribution of Section 3 depending on the relative values of a and c .

To save space, just one example of distribution (7), when $p = 2$, is given in Fig. 4. This has parameters $a = 4$, $\nu = 3$, $c = 2$.

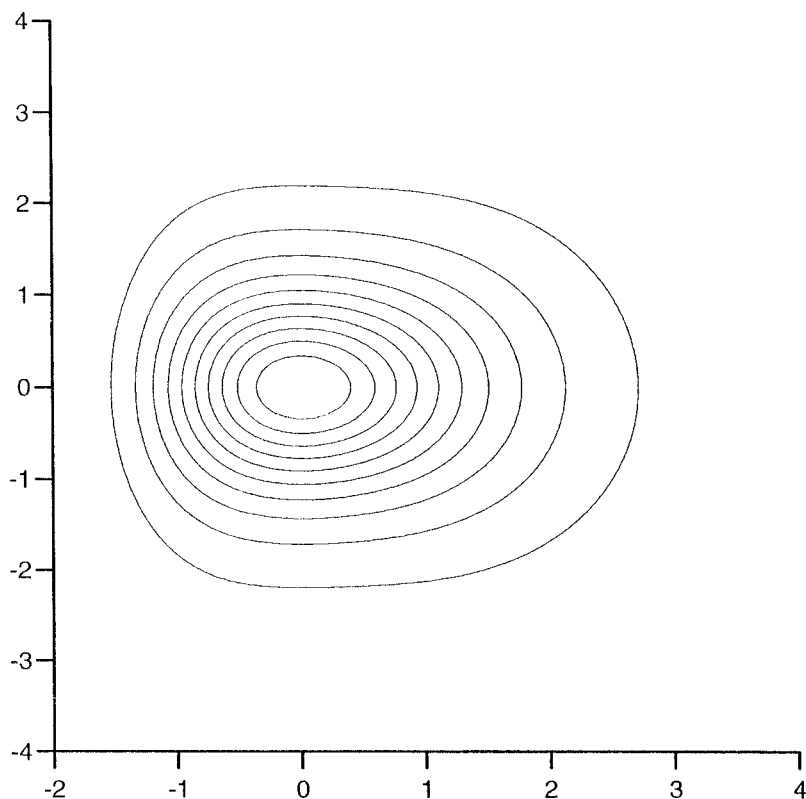
Jones [9] notes that if $B \sim \text{Beta}(a, c)$ then $T = \sqrt{a+c} B / \sqrt{1-B^2} \sim t(a, c)$, the skew t distribution with parameters a and c given at (6). There is also a transformation relationship between the bivariate beta distribution given at (3) and the bivariate t /skew t distribution given at (7) when we take $b = \nu/2 + 1$. Let (B_1, B_2) be random variables following the former distribution. Then

$$(T_1, T_2) = \left(\frac{\sqrt{a+c} B_1}{\sqrt{1-B_1^2}}, \frac{B_2 \sqrt{\nu+B_1^2(a+c-\nu)}}{\sqrt{1-B_1^2-B_2^2} \sqrt{1-B_1^2}} \right) \quad (8)$$

has distribution (7). The relationship between T_1 and B_1 is immediate. To obtain the formula for T_2 , let B_3 have the $\text{Beta}(\frac{1}{2}(\nu+1), \frac{1}{2}(\nu+1))$ distribution independent of B_1 . Then, using the conditional distributions already mentioned, $B_2/\sqrt{1-B_1^2} \sim B_3$ and $\sqrt{\nu+1} T_2 / \sqrt{\nu+T_1^2} \sim \sqrt{\nu+1} B_3 / \sqrt{1-B_3^2}$, and a little manipulation gives the result.

5. ANOTHER NEW MULTIVARIATE DISTRIBUTION, WITH A SPECIALLY CHOSEN MARGINAL

To emphasise that the construction proposed in this paper is very general, let us get away from t /skew t and beta marginal distributions. Suppose again that f is the multivariate t distribution given at (5) and that we construct f_1 by introducing a quite different (skew) marginal for X_1 . This might be, for example, the standard extreme value distribution with $g(x_1) = \exp(-x_1 - e^{-x_1})$, which results in

FIG. 5. The bivariate density (9) with $\nu = 3$.

$$f_1(\mathbf{x}) = \frac{\Gamma((\nu+p)/2)}{(v\pi)^{(p-1)/2} \Gamma((\nu+1)/2)} \frac{\exp(-x_1 - e^{-x_1}) (1 + v^{-1}x_1^2)^{(\nu+1)/2}}{(1 + v^{-1}(x_1^2 + \dots + x_p^2))^{(\nu+p)/2}}. \quad (9)$$

This distribution is given in the bivariate case in Fig. 5 for the case $\nu = 3$.

We stress that this distribution retains much in common with the spherically symmetric t distribution and the distribution of Section 4: each has the same conditional distributions given X_1 , the same local dependence function and zero correlations. What differs between the distributions are the marginals (though those other than X_1 remain symmetric) and the conditional distribution of X_1 given X_2, \dots, X_p .

6. ROTATION, NON-IDENTITY COVARIANCE, THE MULTIVARIATE SKEW-NORMAL DISTRIBUTION

There is, of course, nothing special about choosing X_1 to have its marginal changed above. Any \mathbf{X} would do, including d -dimensional ones, as would changing the marginal distribution along any linear combination $\mathbf{a}'\mathbf{X}$ of the original variables. The latter is equivalent to rotating the kinds of distribution developed in this paper. Notice that any such rotation would in general result in skew marginal distributions along all the current coordinate directions.

Non-identity covariance Σ , say, is easily introduced in the usual manner by multiplying a random vector \mathbf{X} with a distribution of the form (1) for spherically symmetric f by a matrix \mathbf{V} such that $\mathbf{V}'\mathbf{V} = \Sigma$.

An example of an existing distribution which falls within the class of a skew marginal introduced to a spherically symmetric distribution and which has been presented with the involvement of general \mathbf{a} and Σ is the multivariate skew normal distribution of Azzalini and colleagues [2, 3]. Let ϕ and Φ denote the normal density and distribution functions, respectively. Then the univariate skew normal distribution, for which [1] is a key reference, has density

$$2\Phi(\lambda x) \phi(x), \quad (10)$$

the parameter λ controlling the degree of skewness introduced.

Azzalini and Capitanio [2] write the multivariate generalisation of (10) in the form

$$2\Phi(\mathbf{a}'\mathbf{x}) \phi_p(\mathbf{x}; \Sigma), \quad (11)$$

where $\phi_p(\cdot; \Sigma)$ denotes the p -dimensional normal distribution with mean zero and variance Σ . Note that if $\mathbf{a} = (1, 0, \dots, 0)$ and $\Sigma = \mathbf{I}_p$, the identity, this is of the form (1) with $g(x)/f_X(x) = 2\Phi(x)$; call this the standardised multivariate skew normal distribution S say. If $\Sigma = \mathbf{I}_p$, then (11) gives S except with rotation such that $\mathbf{a}'\mathbf{x}$ is identified with x_1 . If $\mathbf{a} = (1, 0, \dots, 0)$, (11) is like applying construction (1) to the elliptically symmetric general multivariate normal distribution, and, by rotation to a different version of the elliptically symmetric multivariate normal distribution, likewise for general \mathbf{a} . Equivalently, starting with $2\Phi(\mathbf{a}'\mathbf{x}) \phi_p(\mathbf{x}; \mathbf{I}_p)$ and transforming \mathbf{X} to $\mathbf{V}\mathbf{X}$ also yields (11) only with \mathbf{a} replaced by $(\mathbf{V}^{-1})'\mathbf{a}$. The latter might in general be the more attractive way of thinking about how general Σ is introduced rather than the former. If (1) is based on a distribution centred at zero, non-zero location μ can, of course, be introduced by replacing \mathbf{x} by $\mathbf{x} - \mu$.

Because of the independence of the marginals of the standard multivariate normal distribution, the marginal replacement scheme underlying the standard multivariate skew normal distribution is essentially trivial. And yet the technique and its extension in this paper are quite powerful. The multivariate skew normal distribution itself is very attractive, with some special properties described in [2] that are specific to the multivariate normal base distribution and the particular construction for skewing it.

Azzalini and Capitanio [2] give an example of the fitting of their distribution to data; data modelling and inference can, of course, be pursued for the novel models in this paper also.

ACKNOWLEDGMENTS

I am grateful to Professor Samuel Kotz for his encouragement on reading a draft of this paper and to the referee and editor for useful comments.

REFERENCES

1. A. Azzalini, A class of distributions which includes the normal ones, *Scand. J. Statist.* **12** (1985), 171–178.
2. A. Azzalini and A. Capitanio, Statistical applications of the multivariate skew-normal distribution, *J. Roy. Statist. Soc. Ser. B* **61** (1999), 579–602.
3. A. Azzalini and A. Dalla Valle, The multivariate skew-normal distribution, *Biometrika* **83** (1996), 715–726.
4. K. T. Fang, S. Kotz, and K. W. Ng, “Symmetric Multivariate and Related Distributions,” Chapman and Hall, London, 1990.
5. P. W. Holland and Y. J. Wang, Dependence function for continuous bivariate densities, *Commun. Statist. Theory Methods* **16** (1987), 863–876.
6. H. Joe, “Multivariate Models and Dependence Concepts,” Chapman and Hall, London, 1997.
7. N. L. Johnson and S. Kotz, “Distributions in Statistics: Continuous Multivariate Distributions,” Wiley, New York, 1972.
8. M. C. Jones, The local dependence function, *Biometrika* **83** (1996), 899–904.
9. M. C. Jones, A skew t distribution, in “Probability and Statistical Models with Applications” (C. A. Charalambides, M. V. Koutras, and N. Balakrishnan, Eds.), pp. 269–278, Chapman and Hall, London, 2001.
10. M. C. Jones and M. J. Faddy, A skew extension of the t distribution with applications, manuscript submitted for publication.