A Class of Symmetric Bivariate Uniform Distributions

Thomas S. Ferguson, 07/08/94

A class of symmetric bivariate uniform distributions is proposed for use in statistical modeling. The distributions may be constructed to be absolutely continuous with correlations as close to ± 1 as desired. Expressions for the correlations, regressions and copulas are found. An extension to three dimensions is proposed.

Keywords. Copulas, Spearman's ρ, Kendall's τ, median regression.

1. Introduction

It is useful in statistical modeling to have tractible multivariate distributions with given marginals in order to be able to quantify the effect of dependence of the variables in the model. There are various ways of constructing such distributions. Usually one maps the random variables to the unit cube using the marginal distribution functions, and deals with the problem there. The present paper presents a simple method of constructing bivariate distributions with uniform marginals. The distribution function corresponding to a distribution on the unit cube with uniform marginals is called a copula. See the proceedings of a recent conference on the subject, Dall'Aglio et al. (1991). For various other methods and reviews, see Kimeldorf and Sampson (1975), Cook and Johnson (1986), Marshall and Olkin (1988), Nelsen (1991) and Kotz and Seeger (1991).

The method presented here is a variation of the convex sum of shuffles of min as described in Mikusiński et al. (1991). Let U be a random variable on the unit interval, $[0, 1]$ with distribution function, $G(u)$. The distribution of (X, Y) is determined as follows. First U is chosen from $G(u)$. Then, conditional on $U = u$, (X, Y) is chosen from a uniform distribution on the four line segments that form the boundary of the rectangle with corners $(u, 0)$, $(0, u)$, $(1-u, 1)$ and $(1, 1-u)$. For $u = 0$, this reduces to the single line segment from $(0, 0)$ to $(1, 1)$, and for $u = 1$, this is the line segment from $(0, 1)$ to $(1, 0)$.

This method provides a large class of bivariate uniform distributions, allowing treatment of a variety of types of dependence. For example, it is easy to construct a one-parameter subfamily of distributions indexed by the correlation coefficient ranging from -1 to $+1$ with the distributions stochastically continuous in the parameter. In addition, the method leads to distributions that are symmetric in X and Y, in the sense that the distribution of (X, Y) is the same as the distribution of (Y, X) (exchangeability) and is the same as the distribution of $(1-X, 1-Y)$ (central symmetry). It shares these properties with two important one-parameter families of bivariate uniform distributions, the family of Plackett (1965), see Johnson and Kotz (1972), and the family of Frank (1979), studied further in Nelsen (1986) and Genest (1987).

Because of these properties, the distributions are particularly useful for statistical models where the loss depends on the observations only through their ranks, as for example in the full-information secretary problem and in some poker models, for in such models one may assume without loss of generality that the marginal distributions are uniform. It is also advantageous for arbitrary symmetric distributions on the plane because if X and Y are transformed by the inverses of symmetric distribution functions, the resulting distribution is still centrally symmetric. Thus, one may obtain distributions with normal marginals and symmetric densities that are not elliptically symmetric.

A related method has been proposed by Marshall (1989). A random variable U on $[0, 1)$ is chosen, and given $U = u$, (X, Y) is chosen uniformly on the union of the two line segments from $(u, 0)$ to $(1, 1-u)$ and from $(0, 1-u)$ to $(u, 1)$. This is a convex sum of shuffles of min, but since opposite edges of the unit square are identified, this is best considered as a distribution on the torus. For example, if U gives most of its mass close to 0 (or 1), then (X, Y) gives most of its mass close to the diagonal $x = y$, including the corners $(0, 1)$ and $(1, 0)$.

Other large classes of bivariate uniform distributions have been constructed. However, the weighted linear combination method of Johnson and Tenebeim (1981) generally leads to distributions that are not exchangeable. The important Archimedean class, studied in Genest and MacKay (1986) and Genest and Rivest (1993), are, except for the Frank family, not centrally symmetric.

2. Main Properties

If G is absolutely continuous, then the distribution of (X, Y) as constructed is absolutely continuous as well.

Theorem 1. *Suppose that G is absolutely continuous with density $g(u)$ for $u \in [0,1]$. Then,*

$$(1)\quad f_{X,Y}(x,y) = \frac{1}{2}[g(|x-y|) + g(1-|1-x-y|)] \quad \textit{for} \quad 0 < x < 1 \quad \textit{and} \quad 0 < y < 1.$$

is a joint density on the unit square with uniform marginal distributions.

If g is smooth on $[0,1]$, the joint density $f(x,y)$ will be be smooth except perhaps on the diagonal lines of the square. However, we only need only to require that $g'(0) = 0$ (resp. $g'(1) = 0$,) in order for $f(x,y)$ to be smooth, i.e. have continuous first derivatives, along the diagonal $x = y$ (resp. $x + y = 1$). (See Examples 1 and 2.) Although we can make the joint density as smooth as desired, it cannot be made analytic in the square. Thus, this class of distributions does not contain the families of Plackett or Frank. For the same reasons, the Farlie-Gumbel-Morgenstern family of quadratic densities, and the extensions to more general polynomial densities in Kimmeldorf and Sampson (1976), Johnson and Kotz (1975, 1977) and Cambanis (1977), are also distinct from the above class. However, with bivariate uniform polynomial densities of fixed degree on $[0,1]^2$, correlations arbitrarily close to ± 1 cannot be obtained (see for example Huang and Kotz (1984)).

The correlation and regression structures are easily obtained for these distributions. For an arbitrary joint distribution of X and Y, Spearman's coefficient ρ is defined as $12\mathrm{P}(X < X', Y < Y'') - 3$, where (X, Y), X' and Y'' are independent, and X' has the same distribution as X and Y'' has the same distribution as Y. For bivariate uniform distributions, Spearman's ρ reduces to the Pearson correlation coefficient, $\rho = 12\mathrm{E}XY - 3$. Kendall's τ is defined as $\tau = 4\mathrm{P}(X < X', Y < Y') - 1$, where (X, Y) and (X', Y') are independent and identically distributed.

Theorem 2. (a) $\rho = 1 - 6\mathrm{E}(U^2) + 4\mathrm{E}(U^3)$.
(b) $\tau = 1 - 2\mathrm{E}U^2 - 2\mathrm{E}\max(U, U')(1 - \max(U, U'))$, *where U and U' are i.i.d. from g.*

We can achieve any correlation between -1 (when U puts all its mass on 1), through $+1$ (when U gives all its mass to 0.). When U has the uniform distribution on $[0,1]$, then X and Y are independent, as can be seen from the joint density. If g is symmetric about $1/2$, the resulting distributions provide simple models of dependent uncorrelated variables. (See Examples 3 and 4.)

The description of the conditional distribution of Y given X, found in the next theorem, follows immediately from the definition of the joint distribution of X and Y given in the second paragraph. It also suggests a simple method of Monte Carlo sampling from this joint distribution. Namely, take X from a uniform distribution on (0,1), choose U independently from G, and toss independently a fair coin. On heads, let $Y = |U - X|$, and on tails let $Y = 1 - |1 - U - X|$. From this description of the conditional distribution of Y given X, the regression is easily found. We also give the median regression as more in keeping with the non-parametric description of these distributions.

Theorem 3. *The conditional distribution of Y given $X = x$ is the distribution of*

$$V_x = \begin{cases} |U - x| & \textit{with probability } 1/2 \\ 1 - |1 - U - x| & \textit{with probability } 1/2 \end{cases}.$$

(a) $\mathrm{E}(Y|X = x) = (1/2)(\mathrm{E}|U - x| + 1 - \mathrm{E}|1 - U - x|)$.
(b) $m(x) = \mathrm{med}(Y|X = x) = \mathrm{med}(V_x)$. *If $G(u)$ is continuous and increasing, then for $x \le 1/2$, $m(x)$ is any root of the equation, $G(m+x)+G(m-x) = 1$ for $x \le m \le 1-x$. For $x > 1/2$, $m(x)$ may be found by central symmetry: $m(x) = 1 - m(1 - x)$.*

The corresponding copula is as follows.

Theorem 4. *The distribution function of X and Y is*

$$F_{X,Y}(x, y) = \mathrm{E}\,\mathrm{med}\{\max(0, x + y - 1), \frac{x + y - U}{2}, \min(x, y)\}$$

where $\mathrm{med}\{a, b, c\}$ *represents the median of the three numbers, a, b and c.*

One can determine $g(u)$ from $f(x, y)$, by $g(u) = f(u, 0)$. This should allow us to check if a given copula is in this class. For example, we also have for $0 \le x \le 1/2$ $f(x, x) = (1/2)[g(0) + g(2x)]$ and $f(0, 0) = g(0)$, so that provided $f(0, 0) \ne \infty$,

$$f(x, 0) = g(x) = 2f(x/2, x/2) - f(0, 0) \quad \text{for} \quad 0 \le x \le 1.$$

3. Examples.

Example 1. $g(u) = 2u$. The density of (X, Y) is $f(x, y) = |x - y| + 1 - |1 - x - y|$. See Fig 1. Spearman's ρ is $-2/5$, and Kendall's τ is $-4/15 = \frac{2}{3}\rho$. The regression of Y on X is $\mathrm{E}(Y|X = x) = (2/3) - x^2 + (2/3)x^3$, and median regression is $m(x) = \sqrt{.5 - x^2}$ for $x < 1/2$, and $m(x) = 1 - \sqrt{-.5 + 2x - x^2}$ for $x > 1/2$. The copula is

$$F(x, y) = \begin{cases} x(y^2 + \frac{1}{3}x^2) & \text{if } x \le y \text{ and } x + y \le 1 \\ y(x^2 + \frac{1}{3}y^2) & \text{if } y \le x \text{ and } x + y \le 1 \\ xy - x(1 - x)(1 - y) + \frac{1}{3}(1 - y)^3 & \text{if } x \le y \text{ and } x + y \ge 1 \\ xy - y(1 - x)(1 - y) + \frac{1}{3}(1 - x)^3 & \text{if } y \le x \text{ and } x + y \ge 1 \end{cases}$$

Fig 1.

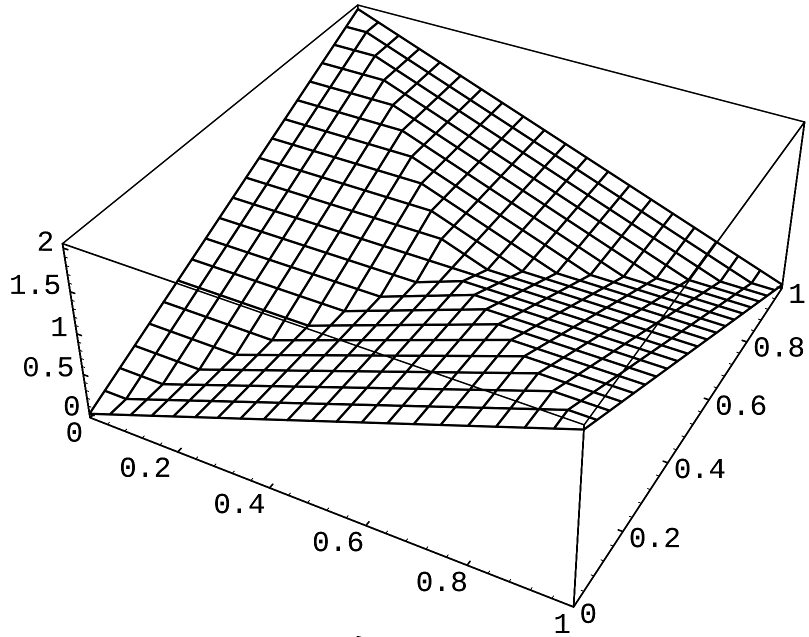
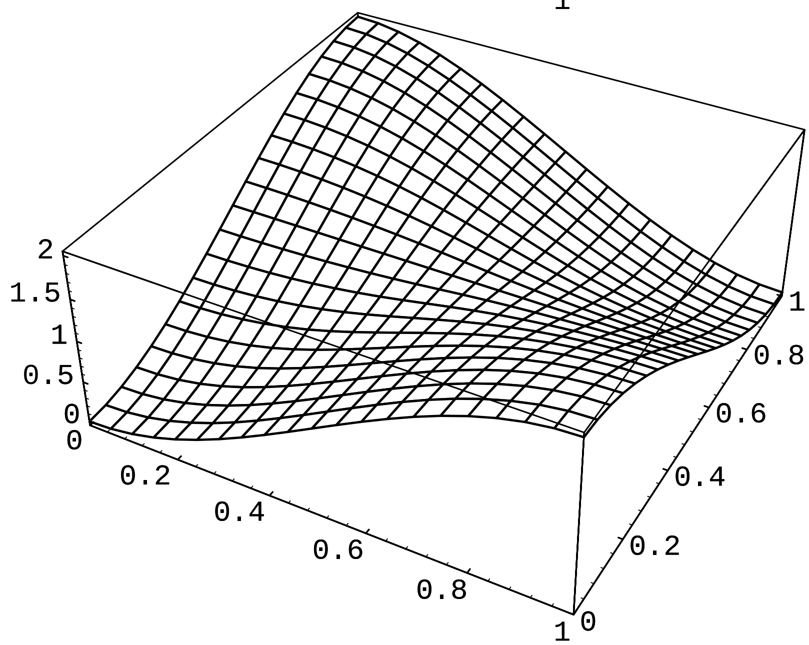


Fig 2.



Example 2. $g(u) = 6u^2 - 4u^3$. $f(x,y) = 6x(1-y) + 6y(1-x) - 2|x-y|^3 - 2(1 - |1-x-y|^3)$. See Fig 2. $\rho = -17/35$, $\tau = -34/105 = \frac{2}{3}\rho$.

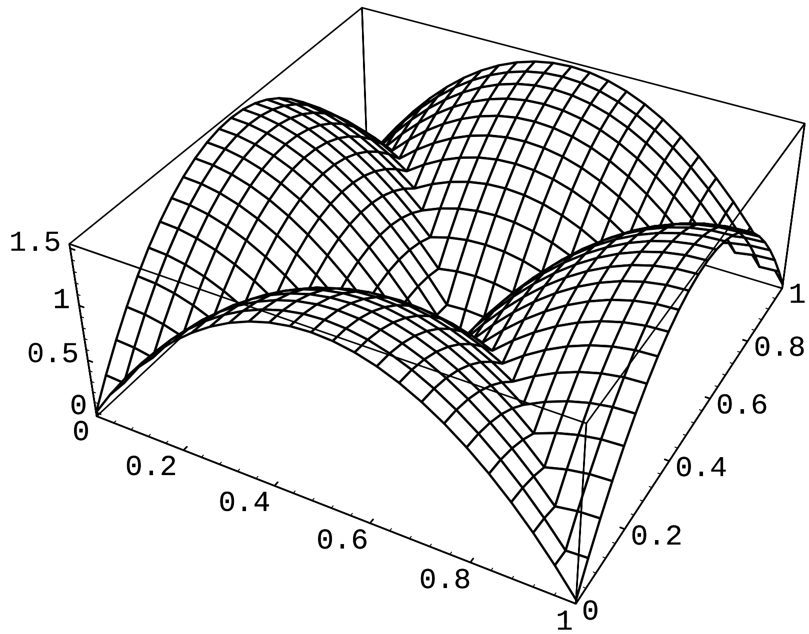
Example 3. $g(u) = 6u(1-u)$. See Fig 3. Correlation zero, regression constant.

Example 4. $g(u) = 2 - 6u(1-u)$. See Fig 4. Correlation zero, regression constant.

4. Proofs

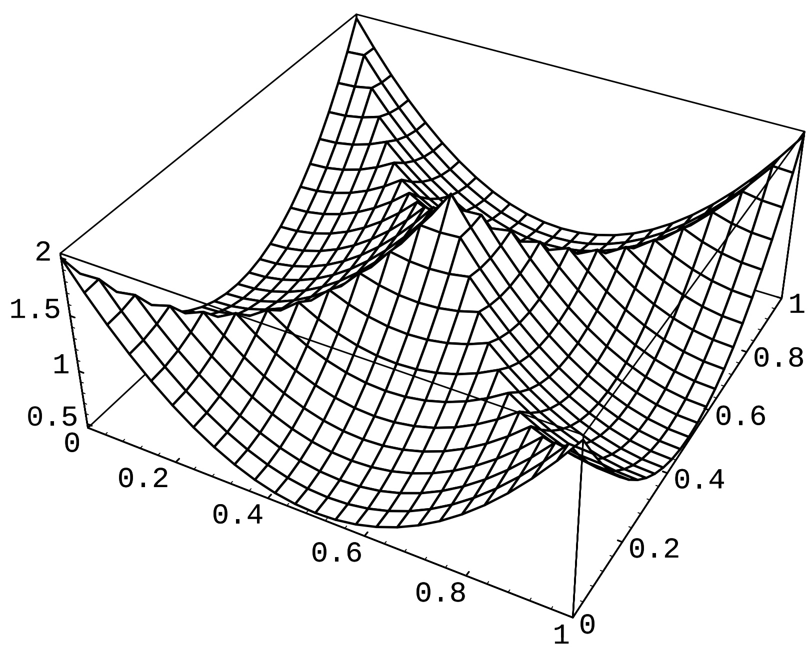
Since conditional on $U = u$, X has a uniform distribution on $[0,1]$, we see that U and X are independent and that X has a uniform distribution on $[0,1]$. Similarly, Y is independent of U, and Y has a uniform distribution on $[0,1]$. If $g(u)$ denotes the density of U, the joint density of (X,Y) at the point (x,y) is the sum of the contributions of the two lines through (x,y) with slopes $+1$ and -1. The line with slope $+1$ is given by

Fig 3.



If the distribution functions are called copulas, might the densities be called cupolas?

Fig 4.



$|x - y| = u$. The line with slope -1 is given by $x + y = u$ or $x + y = 2 - u$ depending on whether $x + y < 1$ or $x + y > 1$. This is the line $1 - |1 - x - y| = u$. From these observations, the joint density (1) follows.

To find the correlation between X and Y, first note that the mean of X and Y is $1/2$, and the variance is $1/12$. To evaluate $\mathrm{E}(XY)$, we first find $\mathrm{E}(XY|U = u)$.

$$\begin{aligned}\mathrm{E}(XY|U=u) &= \frac{1}{2}\Big[\int_0^u x(u-x)\,dx + \int_u^1 x(x-u)\,dx \\ &\qquad + \int_0^{1-u} x(x+u)\,dx + \int_{1-u}^1 x(2-x-u)\,dx\Big] \\ &= \frac{u^3}{3} - \frac{u^2}{2} + \frac{1}{3}\end{aligned}$$

Hence, $\mathrm{E}(XY) = \mathrm{E}(U^3)/3 - \mathrm{E}(U^2)/2 + 1/3$. The covariance of X and Y is $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = \mathrm{E}(U^3)/3 - \mathrm{E}(U^2)/2 + 1/12$, and the correlation is this divided by the common variance, which gives Theorem 2(a).

To compute Kendall's τ, we take U from g and (X,Y) on the unit square as above, and independently take U' from g and (X',Y') as above. Kendall's τ is $\tau = 4\mathrm{P}(X < X', Y < Y') - 1$. It is fairly easy to make the computations

$$\mathrm{P}(X < X', Y < Y'|U,U') = \frac{1}{4}[(1-U)^2 + (1-U')^2 + 2\min(U,U')(1-\min(U,U'))]$$

Then multiplying by 4, subtracting 1 and taking the expectation over U and U' gives the result.

To compute $F(x,y)$, first assume that $x \leq y$ and $x+y \leq 1$. From the representation of the distribution given in Theorem 3, we have

$$\begin{aligned}2F(x,y) &= \int_0^x |P(|U-z| \leq y)\,dz + \int_0^x \mathrm{P}(1-|1-U-z| \leq y)\,dz \\ &= \int_0^x \mathrm{P}(U \leq y+z)\,dz + \int_0^x \mathrm{P}(U \leq y-z)\,dz \qquad \text{since } z \leq y \text{ and } y+z \leq 1 \\ &= \int_0^x \int_0^1 [\mathrm{I}(u \leq y+z) + \mathrm{I}(u \leq y+z)]\,dG(u)\,dz \\ &= \int_0^1 \int_0^x [\mathrm{I}(u \leq y+z) + \mathrm{I}(u \leq y+z)]\,dz\,dG(u) \\ &= \int_0^1 [\mathrm{med}(0, x+y-u, x) + \mathrm{med}(0, y-u, x)]\,dG(u) \\ &= \mathrm{E}\,\mathrm{med}(0, x+y-U, 2x)\end{aligned}$$

Now use exchangeability to find $F(x,y) = \mathrm{E}\,\mathrm{med}(0, (x+y-U)/2, y)$ for $y \leq x$ and $x+y \leq 1$, so that $F(x,y) = \mathrm{E}\,\mathrm{med}(0, (x+y-U)/2, \min(x,y))$ for $x+y \leq 1$. Now use central symmetry to find for $x+y \geq 1$,

$$\begin{aligned}F(x,y) &= F(1-x, 1-y) + x + y - 1 \\ &= \mathrm{E}\,\mathrm{med}(0, \frac{1-x+1-y-U}{2}, \min(1-x, 1-y)) + x + y - 1 \\ &= \mathrm{E}\,\mathrm{med}(x+y-1, \frac{x+y-U}{2}, \min(y,x))\end{aligned}$$

So in all cases, $F(x,y) = \mathrm{E}\,\mathrm{med}(\max(0, x+y-1), (x+y-U)/2, \min(x,y))$.

5. Three dimensions

The extension to three dimensions is probably not unique, but here is one analogous way to construct (X,Y,Z) with uniform marginals.

We choose an arbitrary point $(u,v,0)$ in the xy-plane of the unit cube, $[0,1]^3$. We construct a piecewise linear path moving away from $(u,v,0)$ into the interior of the cube using initial slopes $(+1,+1,+1)$ and then bouncing off the walls of the cube with angle of incidence equal to the angle of reflection until we return to the starting point. This path touches each of the six faces of the cube exactly once. If a point (X,Y,Z), is chosen at random on this path according to a uniform distribution, then the marginal distributions of X, Y and Z are all uniform. Therefore, if we first choose (U,V) at random on the unit square and proceed as above, the marginal distributions of X, Y and Z are still uniform, and each of X, Y and Z is independent of (U,V). If (U,V) has a density, $g_1(u,v)$, then the density of (X,Y,Z) exists and is

$$(2)\qquad f^{(1)}_{X,Y,Z}(x,y,z) = \frac{1}{2}[g_1(|z-x|, |z-y|) + g_1(1-|1-z-x|, 1-|1-z-y|)]$$

From this, we can obtain a density for uniform X, Y and Z that has all three correlations arbitrarily close to $+1$, by choosing g_1 to give most of its mass arbitrarily close to $(0,0)$. Similarly we can obtain

$$\mathrm{Corr}(X,Y) \approx +1 \text{ and } \mathrm{Corr}(X,Z) = \mathrm{Corr}(Y,Z) \approx -1$$

by having g_1 give most of its mass near $(1,1)$,

$$\mathrm{Corr}(Y,Z) \approx +1 \text{ and } \mathrm{Corr}(X,Y) = \mathrm{Corr}(X,Z) \approx -1$$

by having g_1 give most of its mass near $(1,0)$,

$$\mathrm{Corr}(X,Z) \approx +1 \text{ and } \mathrm{Corr}(X,Y) = \mathrm{Corr}(Y,Z) \approx -1$$

by having g_1 give most of its mass near $(0,1)$.

However, there is a completely different path, starting at $(u,v,0)$ but with initial slopes $(+1,-1,+1)$, that hits the other five walls and returns. If we choose (U,V) at random according to a density $g_2(u,v)$ on the unit square, and then choose (X,Y,Z) at random uniformly on this path, we arrive at a second, complementary, joint distribution of (X,Y,Z) with density

$$(3)\qquad f^{(2)}_{X,Y,Z}(x,y,z) = \frac{1}{2}[g_2(1-|1-z-x|, |z-y|) + g_2(|z-x|, 1-|1-z-y|)].$$

This density could be obtained from (2) as the density of $(X,-Y,Z)$.

We may mix these two densities arbitrarily to get a larger class of densities with uniform marginals,

$$f(x, y, z) = \pi f^{(1)}(x, y, z) + (1 - \pi) f^{(2)}(x, y, z) \tag{4}$$

for arbitrary $0 \leq \pi \leq 1$. Since all of the six-sided paths touch each face of the cube exactly once, it does not matter which of the six faces we use as the domain of (U, V); we get the same family of distributions.

Also, since each of the six-sided paths is centrally symmetric, all distributions of this family are centrally symmetric; that is , (X, Y, Z) has the same distribution as $(1 - X, 1 - Y, 1 - Z)$, as is easily seen from the form of the distributions in (2) and (3). However, these distributions are not necessarily exchangeable. In general, we do not obtain exchangeability in the family of distributions given by (2). We can get exchangeability in X and Y in (2) by assuming that $g(u, v)$ is exchangeable. Such a condition gives equal weight to the paths starting at $(u, v, 0)$ and $(v, u, 0)$. To obtain exchangeability in X, Y and Z, we must give equal weight to the six paths starting at $(u, v, 0)$, $(v, u, 0)$, $(u, 0, v)$, $(v, 0, u)$, $(0, u, v)$ and $(0, v, u)$ with initial slopes $(+1, +1, +1)$. Although we can get such distributions using (4), it puts complicated restrictions on g_1 and g_2. It is simpler to choose points $(u, v, 0)$ and $(v, u, 0)$ on the xy-plane, points $(u, 0, v)$ and $(v, 0, u)$ on the xz-plane and points $(0, u, v)$ and $(0, v, u)$ on the yz-plane according to the same exchangeable distribution $g(u, v)$, and use the analogue of (2) for each of them. We arrive at the class of distributions with densities

$$\begin{aligned} f(x, y, z) = \frac{1}{6}[&g(|z - x|, |z - y|) + g(1 - |1 - z - x|, 1 - |1 - z - y|) \\ &+ g(|y - x|, |y - z|) + g(1 - |1 - y - x|, 1 - |1 - y - z|) \\ &+ g(|x - z|, |x - y|) + g(1 - |1 - x - z|, 1 - |1 - x - y|)] \end{aligned} \tag{5}$$

where $g(u, v)$ is a density on $[0, 1]^2$ such that $g(u, v) = g(v, u)$. It is easy to see directly that such distributions are exchangeable.

Acknowledgement. I would like to thank Samuel Kotz for his comments on an earlier version of this paper.

References.

Cambanis, S. (1977) "Some properties and generalizations of multivariate Eyraud-Farlie-Gumbel-Morgenstern distributions" *J. Mult. Anal.* **7**, 551-559.

Cook, R. D. and Johnson, M. E. (1986) "Generalized Burr-Pareto-Logistic distributions with applications to a uranium exploration data set" *Technometrics* **28**, 123-131.

Dall'Aglio, G., Kotz, S. and Salinetti, G., Eds., (1991) *Advances in Probability Distributions with Given Marginals*, Mathematics and Its Applications, vol 67, Klewer Academic Publishers.

Frank, M. J. (1979) "On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$" *Aequationes Math.* **19**, 194-226.

Genest, C. (1987) "Frank's family of bivariate distributions" *Biometrika* **74**, 549-555.

Genest, C. and MacKay, J. (1986) "The joy of copulas: Bivariate distributions with uniform marginals" *Amer. Statist.* **40**, 280-283.

Genest, C. and Rivest, L-P. (1993) "Statistical inference procedures for bivariate Archimedean copulas" *J. Amer. Statist. Assoc.* **88**, 1034-1043.

Gleser, L. J., Perlman, D., Press, S. J. and Sampson, A. R., Eds. (1989) *Contributions to Probability and Statistics. Essays in honor of Ingram Olkin*, Springer-Verlag, New York.

Huang, J. S. and Kotz, S. (1984) "Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions" *Biometrika* **71**, 633-636.

Johnson, M. E. and Tenenbeim, A. (1981) "A bivariate distribution family with specified marginals" *J. Amer. Statist. Assoc.* **76**, 198-201.

Johnson, N. L. and Kotz, S. (1972) *Distributions in Statisics: Continuous Multivariate Distributions*, John Wiley, New York.

Johnson, N. L. and Kotz, S. (1975) "On some generalized Farlie-Gumbel-Morgenstern distributions" *Comm. Stat.* **4**, 415-427.

Johnson, N. L. and Kotz, S. (1977) "On some generalized Farlie-Gumbel-Morgenstern distributions II. Regression, correlation and further generalizations" *Comm. Stat. Th. Meth.* **A6(6)**, 485-496.

Kimeldorf, G. and Sampson, A. (1975) "Uniform representations of bivariate distributions" *Comm. Stat. Th. Meth.* **4**, 617-627.

Kotz, S. and Seeger, J. P. (1991) "A new approach to dependence in multivariate distributions", in Dall'Aglio et al. (1991), 113-127.

Marshall, A. W. and Olkin, I. (1988) "Families of multivariate distributions" *J. Amer. Statist. Assoc.* **83**, 834-841.

Marshall, A. W. (1989) "A bivariate uniform distribution", in Gleser et al. (1989), 99-106.

Mikusiński, P., Sherwood, H. and Taylor, M. D. (1991) "Probabilistic interpretations of copulas and their convex sums", in Dall'Aglio et al. (1991), 95-112.

Nelsen, R. B. (1986) "Properties of a one-parameter family of distributions with specified marginals" *Comm. Stat. Th. Meth.* **15**, 3277-3285.

Nelsen, R. B. (1991) "Copulas and association", in Dall'Aglio et al. (1991), 51-74.

Plackett, R. L. (1965) "A class of bivariate distributions" *J. Amer. Statist. Assoc.* **60**, 516-522.

Thomas S. Ferguson
Mathematics Department
UCLA
405 Hilgard Avenue
Los Angeles, CA 90024