The Joy of Copulas: Bivariate Distributions With Uniform Marginals

CHRISTIAN GENEST and JOCK MACKAY*

We describe a class of bivariate distributions whose marginals are uniform on the unit interval. Such distributions are often called "copulas." The particular copulas we present are especially well suited for use in undergraduate mathematical statistics courses, as many of their basic properties can be derived using elementary calculus. In particular, we show how these copulas can be used to illustrate the existence of distributions with singular components and to give a geometric interpretation to Kendall's tau.

KEY WORDS: Archimedean copulas; Fixed marginals; Fréchet bounds; Kendall's tau; Singular distributions.

1. INTRODUCTION

In introductory mathematical statistics courses, a good deal of time is spent studying joint distributions and transformations of random variables. Students are often told that bivariate distributions may include a singular component even though the marginal distributions are absolutely continuous. In general, however, no examples are given. In fact, even graduate textbooks in statistics rarely contain examples of singular distributions or mention as their only example a distribution concentrated on the Cantor set.

In this article, we present a class of bivariate distributions whose members may contain singular parts and illustrate other interesting phenomena. The elements of this class are called "copulas," because their marginal distributions are uniform on the unit interval [this terminology is used by Schweizer and Sklar (1983)]. An attractive feature of these copulas is that it is possible to derive many of their elementary properties by using simple calculus only. In particular, it is very easy to detect the presence of a singular component and to calculate its probability mass. Within this class of distributions, it is also possible to give a geometric interpretation to Kendall's coefficient of concordance.

2. A CLASS OF SYMMETRIC COPULAS

Consider a class Φ of functions $\phi:[0, 1] \rightarrow [0, \infty]$ that have two continuous derivatives on $(0, 1)$ and satisfy

$$\phi(1) = 0, \qquad \phi'(t) < 0, \qquad \phi''(t) > 0 \qquad (2.1)$$

for all $0 < t < 1$. Conditions (2.1) are enough to guarantee that ϕ has an inverse ϕ^{-1} that also has two derivatives. For convenience, we write $\phi(0) = \infty$ if $\lim_{t\to 0^+} \phi(t) = \infty$. Typical members of the class Φ include $\phi(t) = -\log(t)$, $\phi(t) = (1 - t)^\alpha$, and $\phi(t) = t^{-\alpha} - 1$, where $\alpha > 1$.

*Christian Genest is Assistant Professor and Jock MacKay is Associate Professor, both with the Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Every member ϕ of the class Φ generates a bivariate distribution function for the pair (X, Y) as follows:

$$H(x, y) = \phi^{-1}[\phi(x) + \phi(y)] \quad \text{if } \phi(x) + \phi(y) \le \phi(0);$$
$$= 0 \quad \text{otherwise.} \qquad (2.2)$$

If $\phi(0) = \infty$, then $H(x, y)$ is strictly positive except when $x = 0$ or $y = 0$. A typical ϕ and some level curves $H(x, y) = c$ of the corresponding distribution function are depicted in Figure 1.

To find the density $h(x, y)$ associated with (2.2), let $\phi(H) = \phi(x) + \phi(y)$ and differentiate $H(x, y)$ with respect to x and then y. This yields

$$\phi'(H)\frac{\partial H}{\partial x} = \phi'(x)$$

$$\text{and} \quad \phi''(H)\frac{\partial H}{\partial x}\cdot\frac{\partial H}{\partial y} + \phi'(H)\frac{\partial^2 H}{\partial x \partial y} = 0 \qquad (2.3)$$

so that

$$h(x, y) = -\frac{\phi''(H)\phi'(x)\phi'(y)}{[\phi'(H)]^3}.$$

From the properties of ϕ given in (2.1), it is clear that $h(x, y) > 0$ for all (x, y) such that $\phi(x) + \phi(y) < \phi(0)$. In general, the derivatives do not exist on the boundary $\phi(x) + \phi(y) = \phi(0)$.

The following elementary properties of $H(x, y)$ are easily verified and can be assigned to undergraduate students as exercises.

(a) The distribution is symmetric in x and y; that is, $H(x, y) = H(y, x)$.

(b) The marginal distributions of X and Y are uniform on the interval $(0, 1)$. For example, $H(x, 1) = x$ for all $0 \le x \le 1$. In other words, distributions of the form (2.2) are copulas.

(c) The support of the distribution is $\{(x, y): \phi(x) + \phi(y) \le \phi(0)\}$, which is the complete unit square if $\phi(0) = \infty$.

(d) If $c > 0$ is any constant, ϕ and $c\phi$ generate the same copula.

(e) X and Y are independent iff $\phi(t) = -c\log(t)$, where $c > 0$ is arbitrary.

To verify that $\phi(t)$ is of the specified form when X and Y are independent, it is necessary to solve Cauchy's functional equation $\phi(x) + \phi(y) = \phi(xy)$. When ϕ is twice differentiable, this is readily done by differentiating both sides of the equation with respect to x and y. This yields $\phi'(xy) = -xy\phi''(xy)$, or $\phi'(t) = -t\phi''(t)$. This differential equation is now easy to solve.

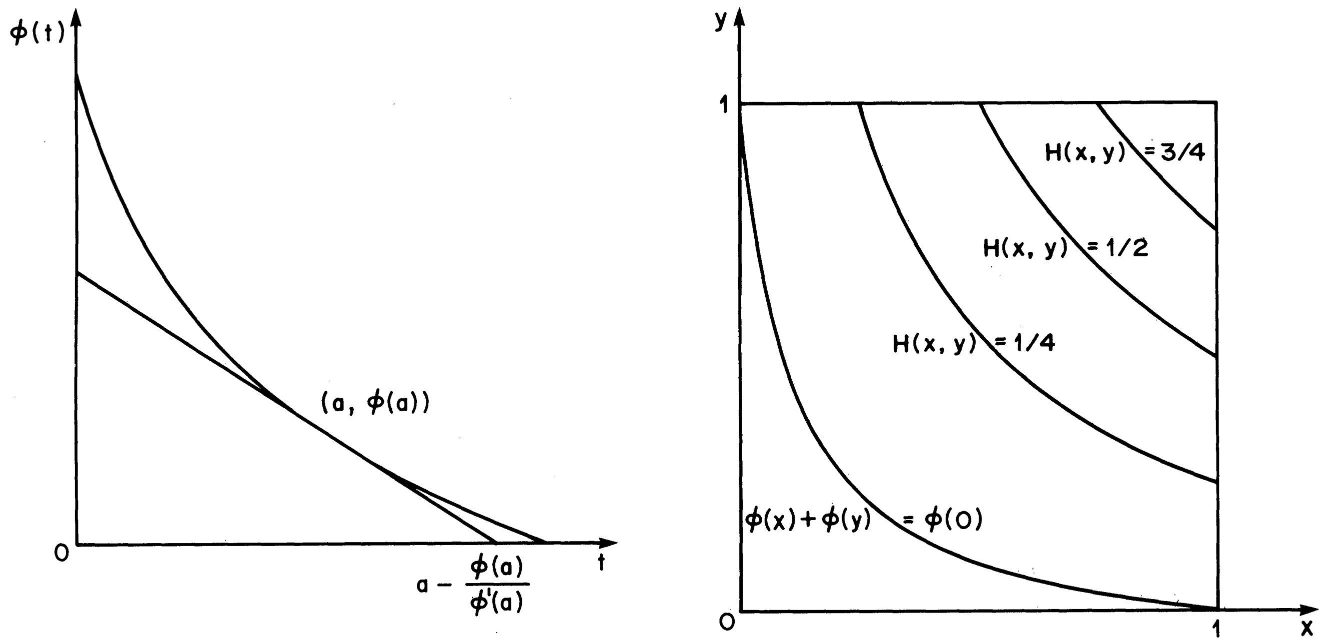

Figure 1. The Graph of a Typical ϕ With $\phi(0) < \infty$ and the Level Curves $H(x, y)$ = ¼, ½, and ¾ of the Corresponding Distribution Function.

(f) $\Pr\{X > x, Y > y\} = H(x, y) - x - y + 1$.
(g) $\max\{0, x + y - 1\} \le H(x, y) \le \min\{x, y\}$.

The lower and upper bounds in the above inequality are usually called the "Fréchet bounds" (Fréchet 1951). The upper bound is the distribution function of the pair (X, Y) when $Y = X$ with probability 1, and the lower bound corresponds to the case $Y = 1 - X$. The first inequality in (g) is a direct consequence of (f), and the other is easily verified using the fact that ϕ (and hence ϕ^{-1}) is decreasing. Actually, both (f) and (g) are true whenever $H(x, y)$ is a bivariate distribution with uniform marginals.

3. COPULAS WITH SINGULAR COMPONENTS

Copulas in the class (2.2) sometimes do and sometimes do not have a singular component. If they do, their singular component is concentrated on the set $\{\phi(x) + \phi(y) = \phi(0)\}$, since the derivatives $\partial H(x, y)/\partial x$ and $\partial H(x, y)/\partial y$ exist everywhere except on that curve. The following theorem tells us exactly when there is a singular component and what probability mass is concentrated there.

Theorem 1. The distribution $H(x, y)$ generated by an element ϕ of Φ has a singular component iff $\phi(0)/\phi'(0) \neq 0$. In that case, $\phi(X) + \phi(Y) = \phi(0)$ with probability $-\phi(0)/\phi'(0)$.

Proof. To derive this fact, we integrate the density $h(x, y)$ over its domain of definition $\{(x, y): \phi(x) + \phi(y) < \phi(0)\}$. To make this easier, let us change variables to

$$u = H(x, y) = \phi^{-1}[\phi(x) + \phi(y)], \qquad v = x$$

so that u and v vary from 0 to 1. For a given value of u, it is easy to see that $u \le v \le 1$ by looking at the level curves as in Figure 1. From (2.3), the Jacobian of the transformation is given by

$$\frac{\partial(u, v)}{\partial(x, y)} = -\frac{\phi'(y)}{\phi'(H)} .$$

Therefore, the probability mass p which is accounted for by the density is

$$\begin{aligned} p &= \iint_{\phi(x) + \phi(y) < \phi(0)} h(x, y)\,dxdy \\ &= -\iint_{0<u<v<1} \frac{\phi''(u)}{[\phi'(u)]^2}\,\phi'(v)\,dvdu \\ &= \int_0^1 \frac{\phi''(u)}{[\phi'(u)]^2}\,\phi(u)\,du. \end{aligned}$$

Integrating by parts, we have

$$p = -\left[\frac{\phi(u)}{\phi'(u)}\right]_0^1 + 1.$$

From Figure 1, we see that the x intercept of the tangent to $\phi(t)$ at $t = a$ is $-\phi(a)/\phi'(a)$, which clearly approaches 0 as a approaches 1 since ϕ is convex. Hence $p = \phi(0)/\phi'(0) + 1$. This probability is less than 1 iff $\phi(0)/\phi'(0) \neq 0$. In this instance, the joint distribution has a singular component on the curve $\phi(x) + \phi(y) = \phi(0)$. That is, with probability $-\phi(0)/\phi'(0)$, the pair (X, Y) will be on the boundary curve.

Example 1. Suppose that $\phi(t) = [t^{-\alpha} - 1]/\alpha$ for some $\alpha > 0$ so that

$$H_\alpha(x, y) = \left[\frac{1}{x^\alpha} + \frac{1}{y^\alpha} - 1\right]^{-1/\alpha}. \tag{3.1}$$

Copulas of this form were suggested by Clayton (1978) and Oakes (1982) for modeling association in bivariate life tables. Cook and Johnson (1981, 1986) also used them to analyze non-elliptically symmetric normal data. In this case, $\phi(0) = \infty$ and $\phi(0)/\phi'(0) = 0$, so these distributions have no singular component and their support is $(0, 1]^2$.

Example 2. The function $\phi(t) = [t^{-\alpha} - 1]/\alpha$ defined in Example 1 is also an element of Φ for $-1 < \alpha \le 0$, where for $\alpha = 0$, we define $\phi(t) = \lim_{\alpha\to 0} [t^{-\alpha} - 1]/\alpha = -\log(t)$. For $\alpha < 0$, $\phi(0) = -1/\alpha$, so the support of the distribution is restricted to the region $\phi(x) + \phi(y) < \phi(0)$. The distribution function on this region is given by (3.1). However, $\phi(0)/\phi'(0) = 0$, so the distribution has no singular component.

Example 3. Consider $\phi(t) = (1 - t)^\alpha$, where $\alpha \geq 1$. These functions give rise to copulas of the form

$$G_\alpha(x, y) = \max\{0, 1 - [(1-x)^\alpha + (1-y)^\alpha]^{1/\alpha}\}. \quad (3.2)$$

In this case, the boundary is $(1 - x)^\alpha + (1 - y)^\alpha = 1$ and the probability that (X, Y) falls on this curve is $-\phi(0)/\phi'(0) = 1/\alpha$. When $\alpha = 1$, (3.2) defines "Fréchet's lower bound," a distribution that is completely singular, as $Y = 1 - X$ with probability 1.

Note that $G_1(x, y)$ is generated by $\phi(t) = 1 - t$, a function that does not quite satisfy the conditions stated in (2.1). In principle, therefore, Theorem 1 does not apply, although it is still true that $-\phi(0)/\phi'(0) = 1$. This example shows that the conditions stated on line (2.1) are sufficient but *not* necessary to imply that (2.2) yields a cumulative distribution function on the unit square. For necessary conditions, see Schweizer and Sklar (1983) or Genest and MacKay (1986).

To complete this section, we also include an example to show that not every symmetric, bivariate distribution function with uniform marginals can be expressed as (2.2).

Example 4. The Fréchet upper bound, defined by $H(x, y) = \min\{x, y\}$, is the distribution function of the pair (X, Y), where $Y = X$ with probability 1. This distribution cannot be written in the form (2.2) for any ϕ in Φ, since $H(x, x) = x$ would imply $2\phi(x) = \phi(x)$ for all $0 < x < 1$.

4. COPULAS AND KENDALL'S TAU

There exist several measures of association for joint distributions. The most familiar is Pearson's correlation, which is ideally suited to the bivariate normal distribution. In this section, however, we wish to focus our attention on Kendall's tau, the population analog of Kendall's coefficient of concordance.

To define this measure, suppose that (X, Y) and (X^*, Y^*) are two independent realizations of a joint distribution. Then τ is the difference between the probability of concordance and the probability of discordance of these two observations, namely,

$$\tau = \Pr\{(X^* - X)(Y^* - Y) \geq 0\} - \Pr\{(X^* - X)(Y^* - Y) < 0\}.$$

Students can derive the following properties of τ in exercises. For simplicity, it is assumed that the marginal distributions are continuous.

(i) $-1 \leq \tau \leq 1$.

(ii) τ is invariant under monotone transformations. That is, if f and g are monotone increasing or decreasing functions, then $\tau(f(X), g(Y)) = \tau(X, Y)$.

Note that if f and g are the marginal distribution functions of X and Y, respectively, then $f(X)$ and $g(Y)$ are uniform. To compute Kendall's tau, it is often convenient to transform the joint distribution of (X, Y) to the unit square.

(iii) $\tau = 1$ $(= -1)$ iff $Y = f(X)$ for some monotone increasing (decreasing) function f.

To prove the necessity of the condition in (iii), first note that we can use the results of (ii) to restrict attention to distributions on the unit square with uniform margins. Having done this, all we need to show is that $Y = X$ with probability 1 when $\tau = 1$. In this case, we know that

$$\Pr\{(X^* - X)(Y^* - Y) \geq 0\} = 1.$$

Conditioning on a possible value of (X, Y), say (x, y), and using the independence of the two observations (X, Y) and (X^*, Y^*), we get

$$\Pr\{(X^* - x)(Y^* - y) \geq 0\} = 1.$$

Using property (f) in Section 2, we can express this equation in terms of the joint distribution function H as $2H(x, y) - x - y + 1 = 1$. So $H(x, y) = (x + y)/2$ for all $0 < x, y < 1$. Since $H(x, y) \leq \min\{x, y\}$, the upper Fréchet bound, it follows that $x = y$ as required.

(iv) $\tau = 0$ if X and Y are independent (but not conversely).

(v) $\tau = 4E[H(X, Y)] - 1$.

The above expression is obtained by writing

$$\tau = 2\Pr\{(X^* - X)(Y^* - Y) \geq 0\} - 1$$
$$= 2E[\Pr\{(X^* - X)(Y^* - Y) \geq 0 | X = x, Y = y\}] - 1.$$

Using the independence of the two vectors (X, Y) and (X^*, Y^*) and identity (f) from Section 2, the result follows easily.

As we will now see, there is a simple formula for computing Kendall's τ when a copula belongs to the family defined by (2.2). The value of τ is related in a linear fashion to the area above the curve $\phi(t)/\phi'(t)$ between 0 and 1. This gives a geometrical interpretation to Kendall's measure of association.

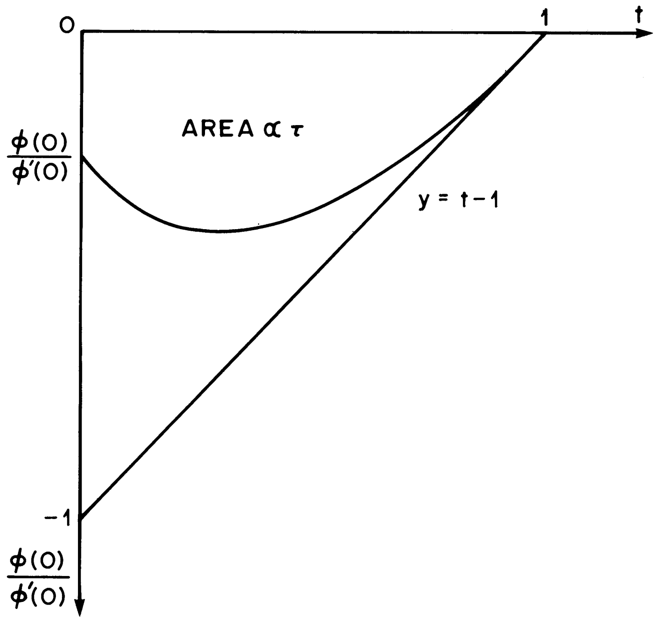

Figure 2. The Graph of $\phi(t)/\phi'(t)$.

Theorem 2. Let (X, Y) be a pair of random variables whose distribution H is of the form (2.2) for some ϕ in Φ. Then

$$\tau(X, Y) = 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt + 1.$$

Proof. First note that $H(x, y) = 0$ for all (x, y) such that $\phi(x) + \phi(y) = 0$. Hence we can compute τ using (v) by integrating H over the region in which there is a density. That is,

$$\tau = \iint_{\phi(x)+\phi(y)<\phi(0)} H(x, y)\frac{-\phi''(H)\phi'(x)\phi'(y)}{[\phi'(H)]^3}dxdy.$$

Making the same transformations as in Theorem 1 and integrating yields the desired conclusion.

Figure 2 summarizes what the graph of $\phi(t)/\phi'(t)$ tells us about $H(x, y)$. The probability associated with the singular component and an estimate of Kendall's τ are readily available. The graph is of most interest when comparing two copulas. Note that Fréchet's lower bound gives $\phi(t)/\phi'(t) = t - 1$. As $\phi(t)/\phi'(t)$ approaches 0, we get a distribution close to Fréchet's upper bound.

Examples 1 and 2 (continued). In this case, it is easy to see that $\tau(X, Y) = \alpha/(\alpha + 2)$ for all $\alpha \geq -1$. In particular, note that the cases $\tau = -1$ and 0 correspond to the Fréchet lower bound and the independence distribution ($\alpha = -1$, 0, respectively), and τ approaches 1 as α tends to infinity. It is easy to verify directly that $\lim_{\alpha\to\infty} H_\alpha(x, y) = \min\{x, y\}$.

Example 3 (continued). Here, we see that $\tau(X, Y) = 1 - 2/\alpha$, which suggests that G_α approaches Fréchet's upper bound as α increases indefinitely. This is easy to check. Note also that $\tau = 0$ when $\alpha = 2$ but that

$$G_2(x, y) = \max\{0, 1 - \sqrt{(1 - x)^2 + (1 - y)^2}\}$$

is not the independence distribution. This illustrates the parenthetical comment made in exercise (iv).

5. COMMENT

Copulas of the form (2.2) serve other purposes besides those outlined in this article. Some of their theoretical uses are described in the book by Schweizer and Sklar (1983) and in a paper by Genest and MacKay (1986). In the latter, for example, it is shown how these copulas can be used to generate one-parameter families of bivariate distributions with prescribed marginals in such a way as to approach the Fréchet bounds "smoothly." Two examples of such families have been presented here (Examples 1 and 2). In general, it turns out that the convergence of a sequence H_n of copulas of the form (2.2) can be determined by simply looking at the graph of ϕ_n/ϕ_n'. See Genest and MacKay (1986) for details.

[*Received October 1985. Revised June 1986.*]

REFERENCES

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Cook, R. D., and Johnson, M. E. (1981), "A Family of Distributions for Modelling Non-elliptically Symmetric Multivariate Data," *Journal of the Royal Statistical Society*, Ser. B, 43, 210–218.

——— (1986), "Generalized Burr-Pareto-Logistic Distributions With Applications to a Uranium Exploration Data Set," *Technometrics*, 28, 123–131.

Fréchet, M. (1951), "Sur les Tableaux de Corrélation Dont les Marges Sont Données," *Annales de l'Université de Lyon*, Série 3, 14, 53–77.

Genest, C., and MacKay, R. J. (1986), "Copules Archimédiennes et Familles de Lois Bidimensionnelles Dont les Marges Sont Données," *The Canadian Journal of Statistics*, 14, 145–159.

Oakes, D. (1982), "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society*, Ser. B, 44, 414–422.

Schweizer, B., and Sklar, A. (1983), *Probabilistic Metric Spaces*, New York: North-Holland.

The Meaning of Kurtosis: Darlington Reexamined

J. J. A. MOORS*

There seems to be no universal agreement about the meaning and interpretation of kurtosis. An easy interpretation is given here: kurtosis is a measure of dispersion around the two values $\mu \pm \sigma$.

The concept of kurtosis seems to be rather difficult to interpret. Most statistical textbooks describe kurtosis in terms of peakedness, and some seek the explanation in heavy tails. Ben-Horim and Levy (1984) is an exception: it presents a more elaborate example featuring a bimodal distribution.

Bimodality as interpretation of kurtosis was introduced in Darlington (1970). Unfortunately, he pushed an otherwise correct argument one step too far. The present note reexamines his reasoning.

The kurtosis k will be defined here as the normalized fourth central moment; compare Kendall and Stuart (1969). So, any random variable X with expectation $E\{X\} = \mu$, variance $V\{X\} = \sigma^2$, and finite fourth moment has

$$k = E\{X - \mu\}^4/\sigma^4.$$

Introduction of the standardized variable $Z := (X - \mu)/\sigma$

*J. J. A. Moors is Senior Lecturer, Department of Econometrics, Tilburg University, 5000LE Tilburg, Netherlands. The author is grateful to two unknown referees, whose comments substantially improved earlier drafts of this article.