THE JOY OF COPULAS

Why Copulas?

- Understanding relationships among multivariate outcomes is the goal of researchers.
- Regression analysis limited by the fact that one variable must be selected as dependent and the others as explanatory.
- Correlation coefficient widely used and perhaps misused measure of dependence.
- Correlation coefficient not invariant under monotonic transformation of variables.
- Marginal distributions and correlation coefficient do not determine the joint distribution, except for the elliptical distributions.
- Given marginals F_1 and F_2, not all linear correlations between -1 and 1 are attainable.

Definition

A two dimensional copula is a real function, C(x, y), defined on $I^2 = [0, 1]x[0, 1]$, with the range I = [0, 1], such that:

C(0, v) = C(u, 0) = 0, for all u and v in I.

C(1,v) = v, C(u, 1) = u, for all u and v in I.

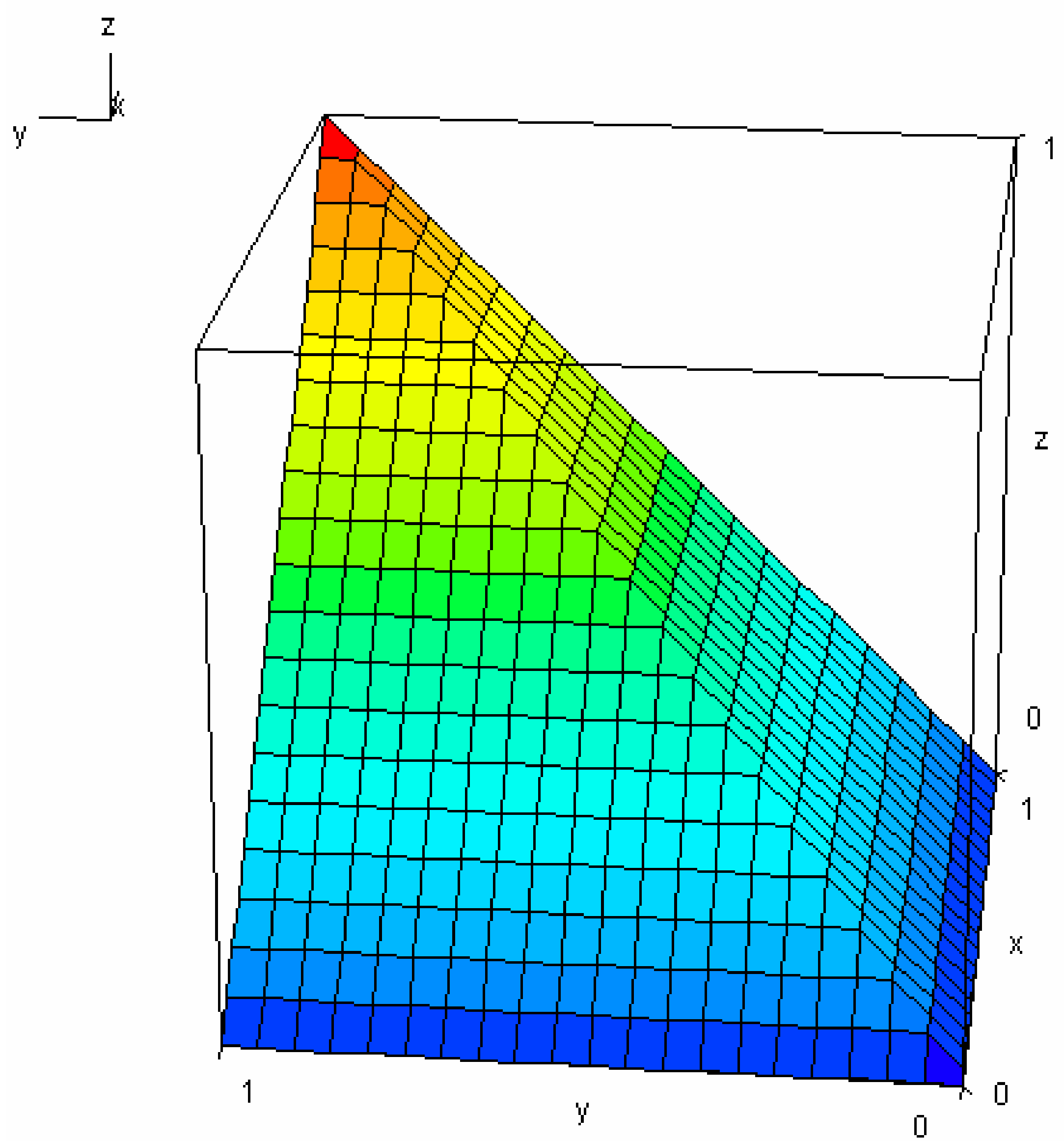
$C(u_2, v_2) – C(u_2, v_1) – C(u_1, v_2) + C(u_1, v_1) \geq 0$, for all rectangles $[u_1, v_1]x[u_2, v_2]$ in I^2. That is C is 2-increasing.

Remark1. We can see, from the above definition, that a copula represents the joint distribution of a two standard uniform random variables U_1 and U_2:

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$$

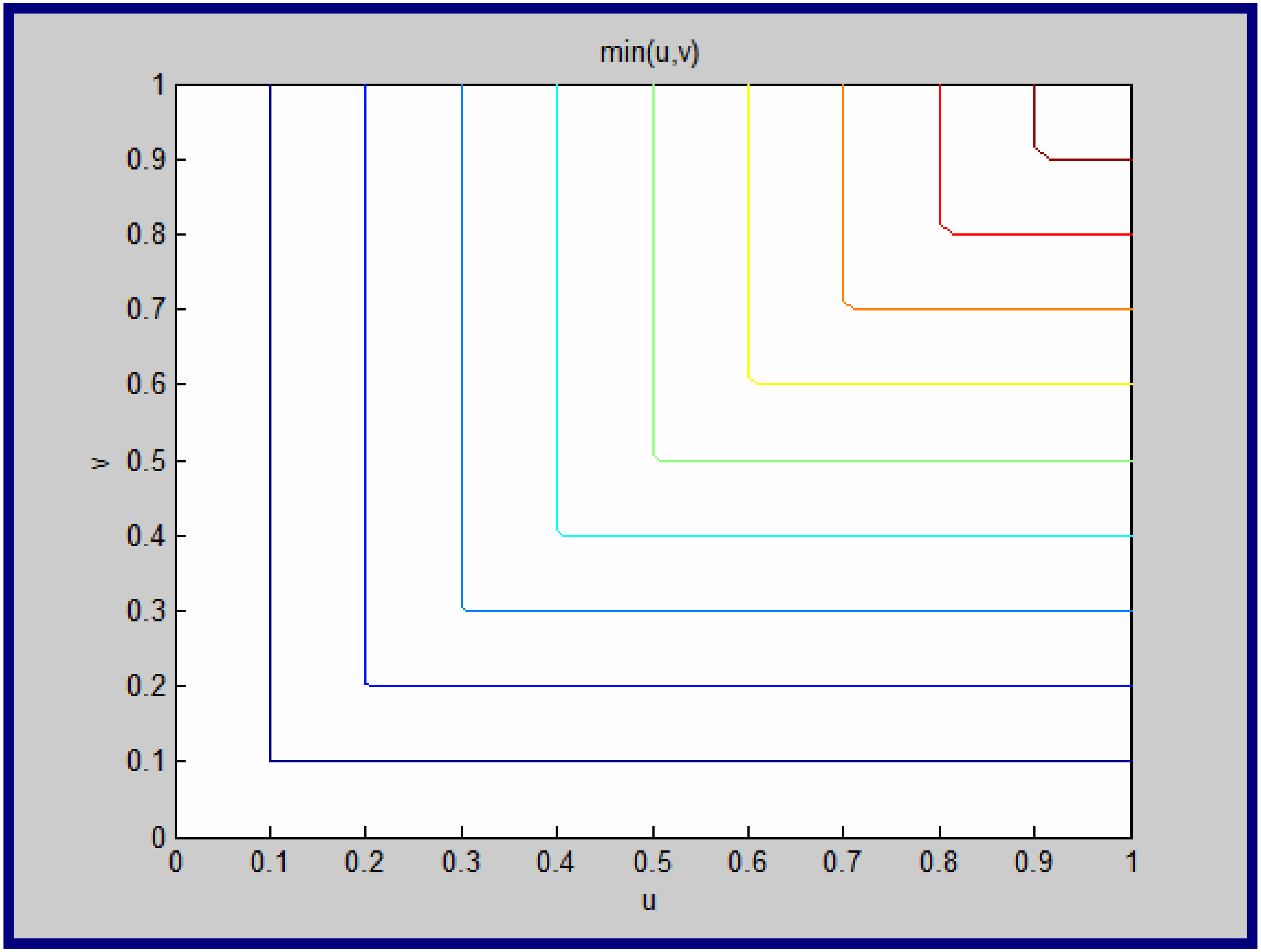
Remark2. Property *a)* says that C is grounded and b*)* says C has uniform margin.

Some important examples of copulas:

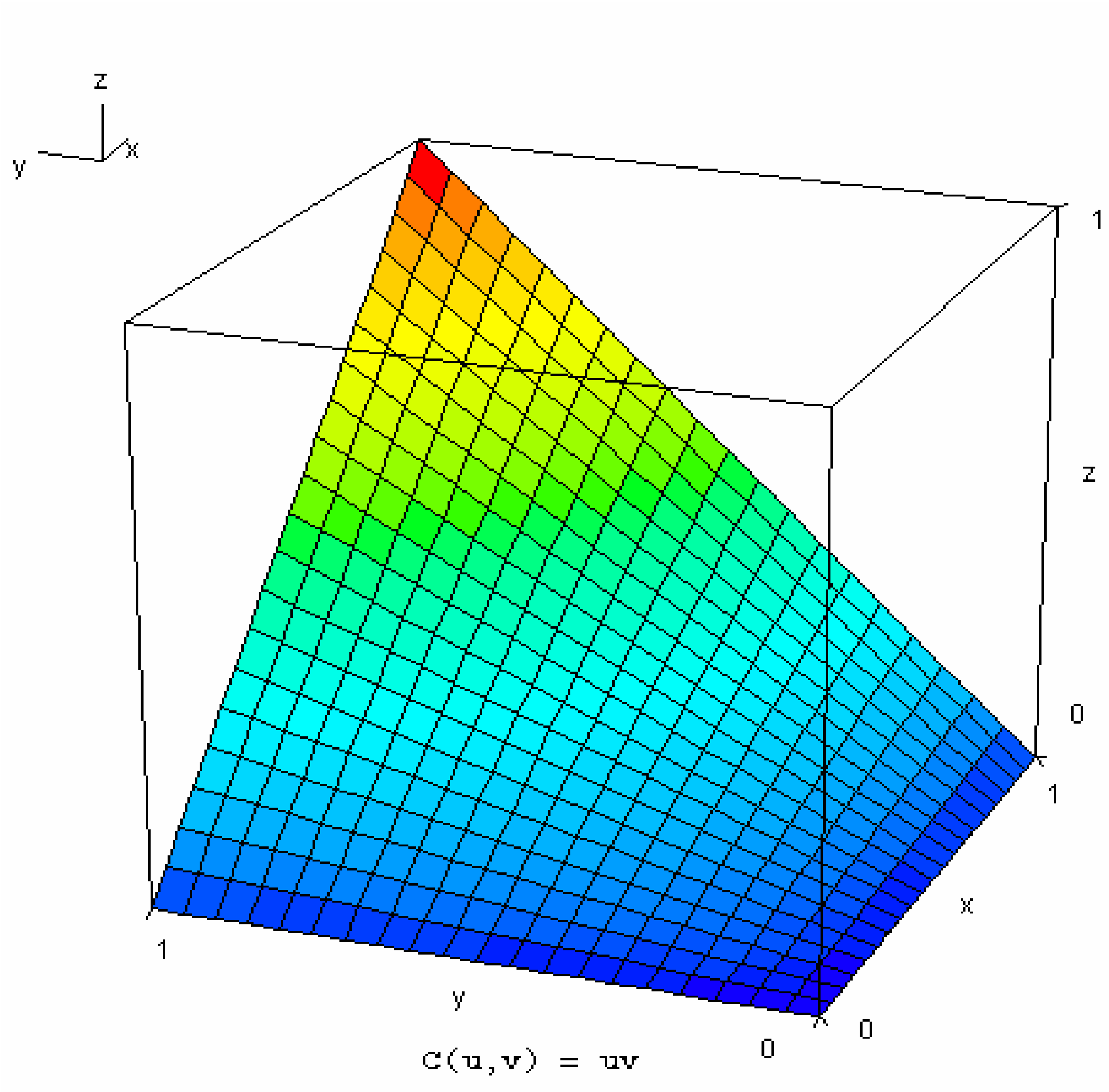


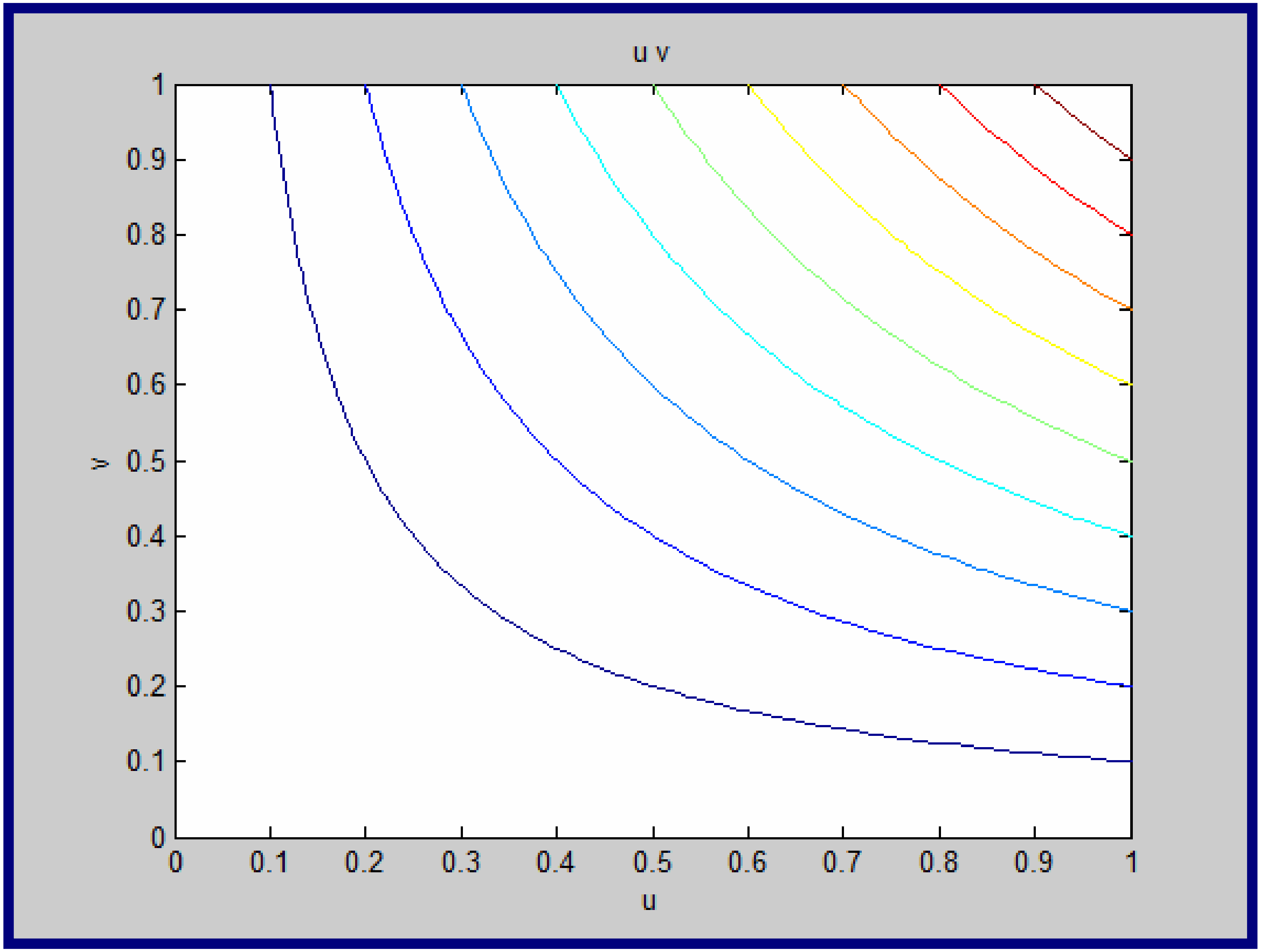
$C(u,v) = Min(u,v)$

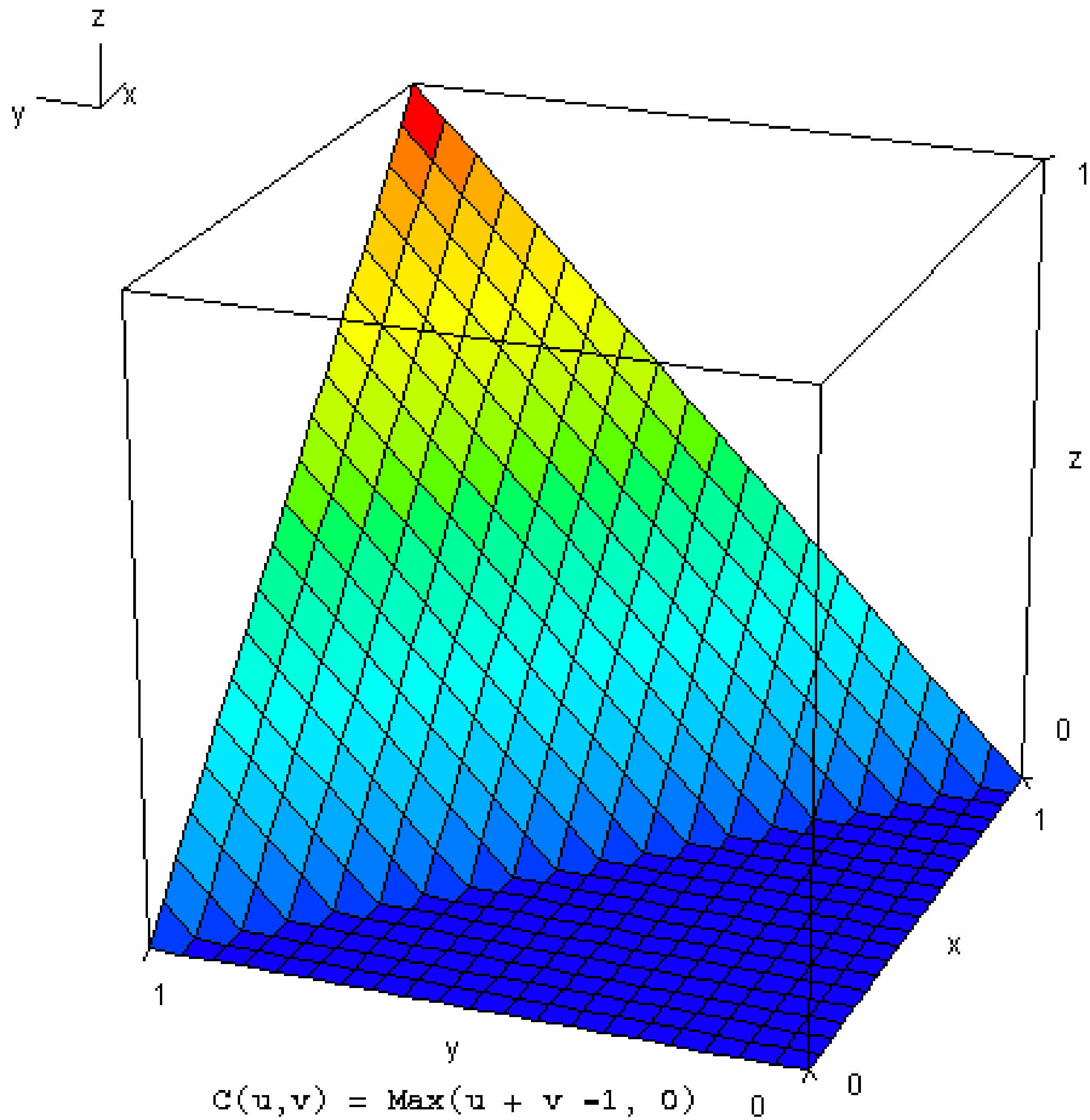
The level curves of Min(u, v)

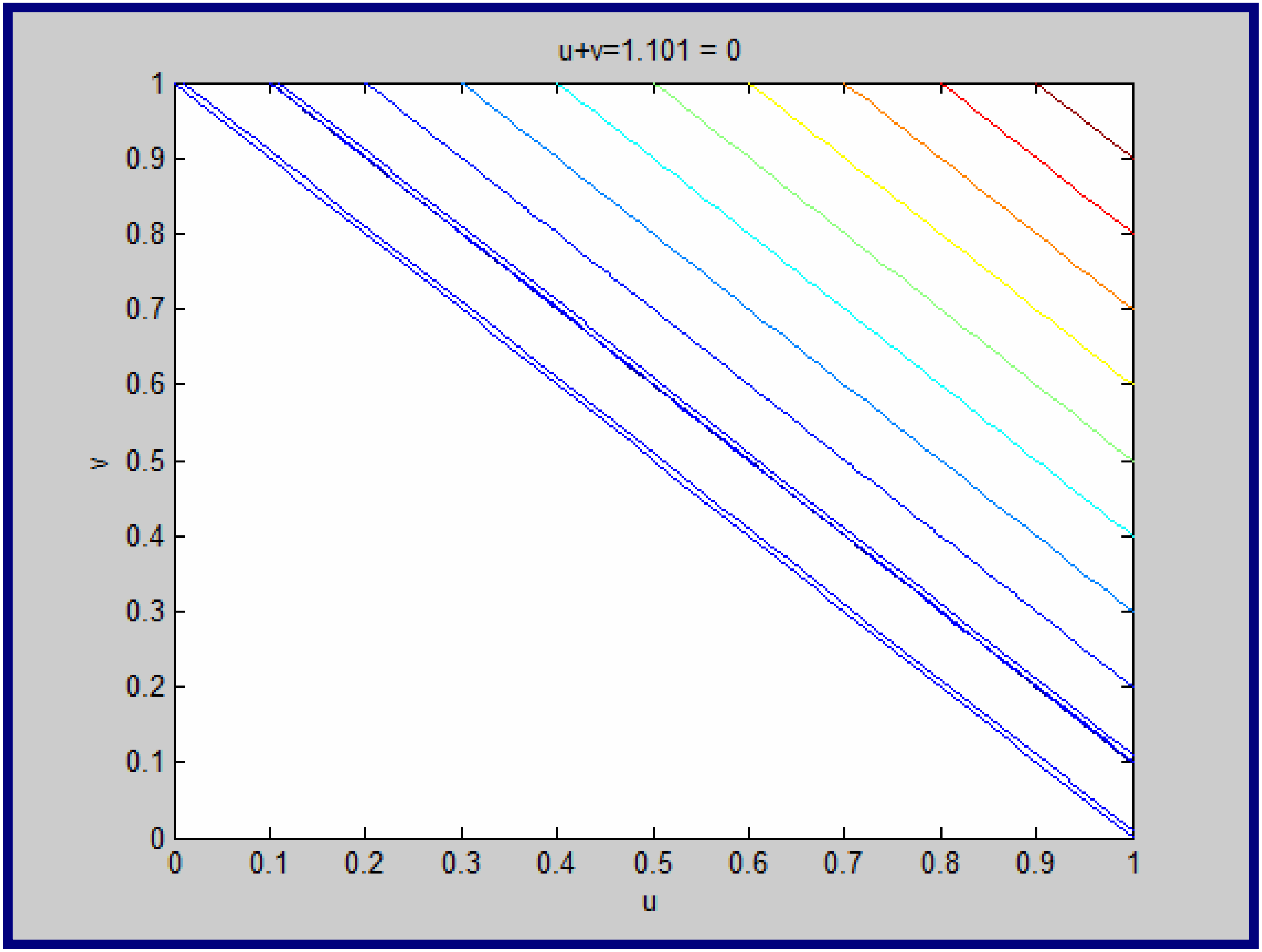


The Product Copula: $\Pi(u,v) = uv$




u v
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
v
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
u


z
y
x
1
z
0
1
x
1
y
0
0
C(u,v) = Max(u + v -1, 0)



u+v=1.101 = 0
v
u
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Fréchet bounds

For every multivariate distribution function $F(x_1,\ldots,x_n)$, we have:

$$\max[\sum_{i=1}^{n} F_i(x_i)+1-n,0] \le F(x_1,\ldots,x_n) \le \min(F_1(x_1),\ldots,F_n(x_n))$$

The bounds are known as Fréchet bounds.

If we reduce n to 2, the bounds themselves are copulas and are denoted by:

*$M(x_1, x_2) = \min(x_1, x_2)$, and $W(x_1, x_2) = \max(x_1 + x_2 - 1, 0)$. Thus for every copula C and every (x_1, x_2) in **I**, we have:*

$W(x_1, x_2) \le C(x_1, x_2) \le M(x_1, x_2)$.

*Remark. For $n > 2$, $M(x_1,\ldots,x_n)$ and the product Π^n for x_i in **I** are still copulas, but the lower bound is no longer a copula.*

What can one say about random variables with copulas *M* or *W*?

Assume $X_1 \sim F_1$, $X_2 \sim F_2$. Also assume $X_2 = T(X_2)$, *T strictly increasing. Then:*

$$F_2(x) = P(X_2 \le x) = P(T(X_1) \le x) = P(X_1 \le T^{-1}(x))$$

$$F_2(x) = (F_1 \circ T^{-1})(x) \Rightarrow F_2 = F_1 \circ T^{-1}$$

$$C(x_1, x_2) = P(X_1 \le F_1^{-1}(x_1), X_2 \le F_2^{-1}(x_2)) = P(F_1(X_1) \le x_1, F_2(X_2) \le x_2)$$

$$C(x_1, x_2) = P(F_1(X_1) \le x_1, (F_1 \circ T^{-1})(X_2) \le x_2) = P(F_1(X_1) \le x_1, F_1(X_1) \le x_2)$$

$$C(x_1, x_2) = P(F_1(X_1) \le \min(x_1, x_2) = M(x_1, x_2)$$

Similarly, but with more work, it can be shown that if $X_2 = \beta(X_1)$*, where* β *is strictly decreasing, then* $C(x_1, x_2) = W(x_1, x_2)$.

SKLAR THEOREM

Let *F(x,y)* be a joint distribution with continuous marginals $F_1(x)$ and $F_2(x)$. Then there exists a unique copula *C(x,y)* such that

$$F(x,y) = C(F_1(x), F_2(y))$$

Conversely if C(x, y) is a copula and $F_1(x)$ and $F_2(y)$ are two continuous univariate distributions, then

$$F(x,y) = C(F_1(x), F_2(y))$$

is a joint distribution with marginals $F_1(x)$ and $F_2(y)$.

Definition

Let $F(t)$ be a distribution function. Then the quasi-inverse of F is any function $F^{(-1)}(t)$ with domain ***I*** such that:

- If t is in Ran$F$, then $F^{(-1)}(t)$ is any number x such that $F(x) = t$, i.e., for all t in Ran$F$, $F(F^{(-1)}(t) = t$;
- If t is not in Ran$F$, then $F^{(-1)}(t) = inf\{x|\ F(x) \geq t\} = sup\{x|\ F(x) \leqslant t\}$.

Corollary:

Let $F(x,y)$, $F_1(x)$, $F_2(y)$ and C be as in Sklar's theorem. Then for any u and v in ***I****, $C(u,v) = F(F_1^{(-1)}(u), F_2^{(-1)}(v))$.*

Note: This result can be extended to n dimensions.

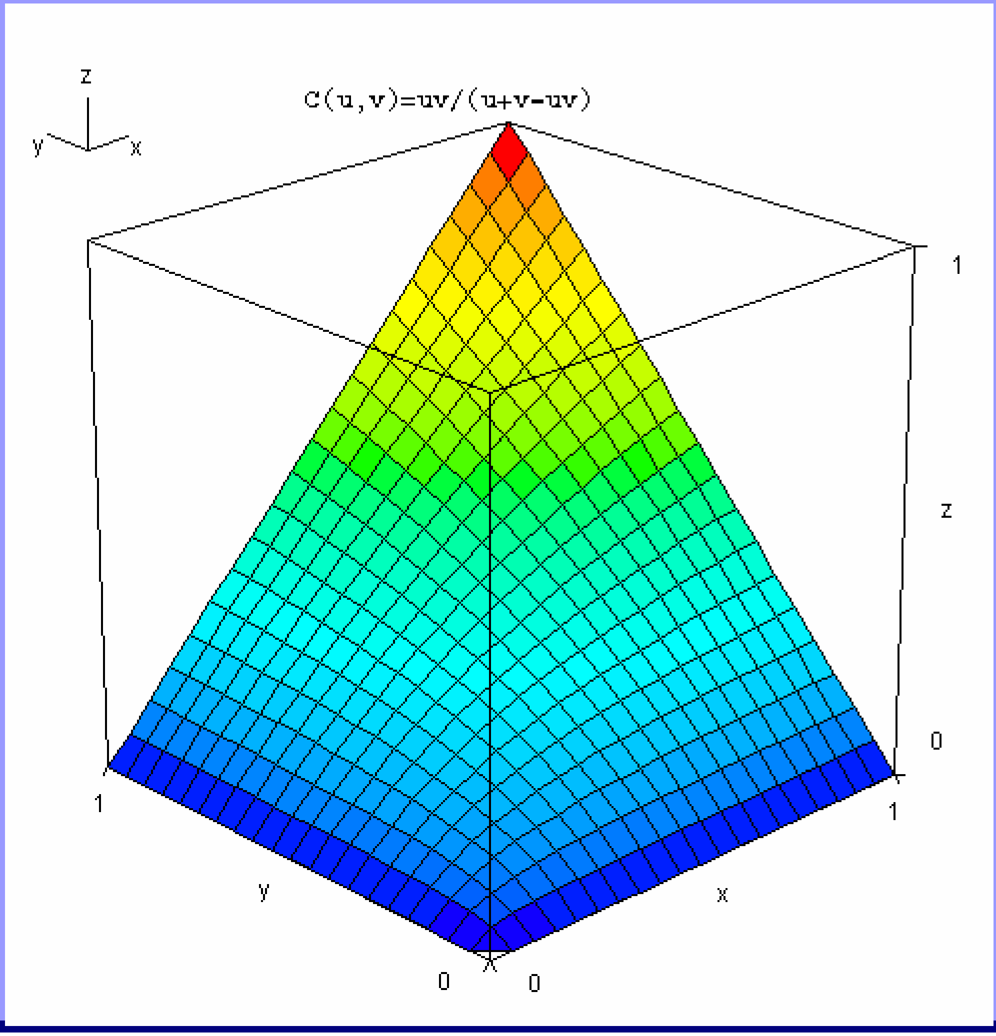
Example:

$$F(x,y)=\begin{cases}\dfrac{(x+1)(e^{y}-1)}{x+2e^{y}-1} & (x,y)\in \quad [-1,1]\times[0,\infty] \\ 1-e^{y} & (x,y)\in \quad (1,\infty]\times[0,\infty] \\ 0 & elsewhere \end{cases}$$

$$F_1(x)=\begin{cases} 0 & x<-1 \\ (x+1)/2 & x\in[-1,1] \\ 1 & x>1 \end{cases}$$

$$F_2(y)=\begin{cases} 0 & y<0 \\ 1-e^{-y} & y\geq 0 \end{cases}$$

$$C(u,v)=\frac{uv}{u+v-uv}$$



C(u,v)=uv/(u+v-uv)
z
y
x
1
z
0
1
y
0
0
x
1

Theorems:

1. *Let X and Y be continuous random variables. Then X and Y are independent iff* $C_{XY} = \Pi$.
   Proof: Assume *X* and *Y are independent. Then* $F(x,y) = F_1(x)F_2(y)$, *and* $C(u,v) = F(F_1^{(-1)}(u), F_2^{(-1)}(v)) = F(F_1^{(-1)}(u)\ F(F_2^{(-1)}(v)) = uv$.
   The other direction is clear.
2. *Let X and Y be continuous random variables with copula* C_{XY}. *If* α *and* β *are strictly increasing on RanX and RanY respectively, then* $C_{\alpha(X)\beta(Y)} = C_{XY}$.
   Proof: see Nelsen.
   Note that the joint distribution of $\alpha(X)$ *and* $\beta(Y)$ *is not the same as the joint distribution of X and Y. It is this property of copulas that will be most useful in studying the dependence structure of bivariate random variables.*

Theorems cont.

3. *Let X and Y be continuous random variables with copula C_{XY}. If α and β are strictly monotone on RanX and RanY respectively, then*
   1. *If α is strictly increasing and β is strictly decreasing, then* $C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$.
   2. *If α is strictly decreasing and β is strictly increasing, then* $C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v)$.
   3. *If α and β are both strictly decreasing, then* $C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$.

Survival Copulas

- Let (X, Y) be a random pair with distributions $F_1(x)$, and $F_2(y)$, *and with copula* C. *Then C* is the cdf of the random pair $U = F_1(X)$ and $V = F_2(Y)$.
- The cdf of the random pair $(1 – U, 1 – V)$ is $C^*(u, v)$ given by: $C^*(u, v) = u + v – 1 + C(1 – u, 1 – v)$.
- C^* *satisfies* $P(X > x, Y > y) = \overline{F}(x, y) = C^*(\overline{F}_1(x), \overline{F}_2(y))$

Example:

If X and Y are independent, then their survivor copula is given by:

$C^*(u, v) = u + v - 1 + (1 – u)(1 – v) = uv = \Pi$

Definition

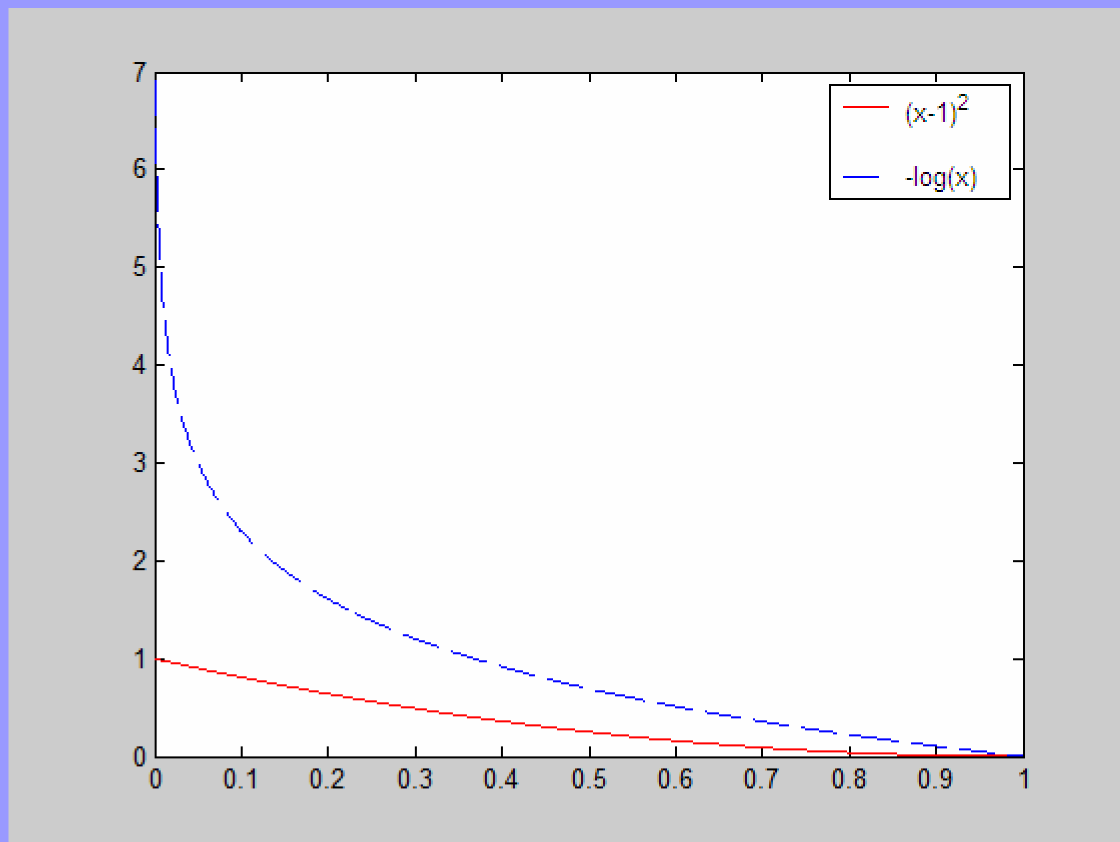
Let ϕ: [0,1] → [0, ∝] be a continuous decreasing function. Then the quasi-inverse of ϕ is any function $\phi^{(-1)}$ with the domain and range I such that

a. *It t is in Ranϕ, i.e., , then $\phi^{(-1)}(t)$ is any number in such that $\phi(x)$ = t, i.e. for all t in Ranϕ, ;*

b. *If t is not in Ranϕ, i.e., then* $\phi^{(-1)}(t) = 0$

$$\phi^{(-1)}(t) = \begin{cases} \phi^{-1}(t) & 0 \le t \le \phi(0) \\ 0 & \phi(0) \le t \le \infty \end{cases}$$

Examples:



Now consider a class Φ of functions φ: [0, 1] → [0, ∝] with the following properties:

$\varphi(1) = 0$

$\varphi'(t) < 0$ for all t in (0, 1)

$\varphi''(t) > 0$ for all t in (0, 1)

Note that the functions in Φ are convex .

The following properties of a convex function will be useful later:

Property1. $\phi(\alpha x + (1-\alpha) y) \leq \alpha\phi(x) + (1-\alpha)\phi(y)$ for α in (0, 1).

Property 2. If $x_1 \leqslant x_2$, then there exist an x such that $\varphi(x) = \varphi(x_1) - \varphi(x_2)$.

Constructing Copulas

Consider the following function of two variables:

$$C(u,v)=\begin{cases}\varphi^{(-1)}(\varphi(u)+\varphi(v)) & \text{if } \varphi(u)+\varphi(v)\le\varphi(0)\\ 0 & \text{otherwise}\end{cases}$$

Fact: *C(u,v)* as defined above satisfies the conditions of a copula and therefore is a copula. Copulas of this form are called Archimedean copulas. The function

φ Is called a generator of the copula. If $\varphi(0) = \propto$, then we call ϕ a strict generator.

Elementary properties.

1. C is symmetric: $C(u, v) = C(v, u)$; u, v in $\mathbf{I}$.
2. C is associative: $C(C(u,v),w) = C(u, C(v, w))$.
3. If φ is a generator of C then for $c > 0$, $c\varphi$ is also a generator of C.
4. C is strict iff $C(u, v) > 0$, for all (u, v) in $(0, 1]$.

The proof of these statements are straightforward!

Examples:

Copula: $C_\theta(u, v)$	Generator: $\varphi_\theta(t)$	θ ε
$\mathrm{Max}\{[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0\}$ "*Clayton Family*	$(t^{-\theta} - 1)/\theta$	$[-1, \infty)\backslash\{0\}$
$\exp\{-[(-\ln(u))^\theta + (-\ln(v))^\theta]^{1/\theta}\}$ *"Gumbel-Hougaard Family"*	$(-\ln(t))^\theta$	$[1, \infty)$
$uv/[1 - \theta(1 – u)(1 – v)]$ "*Ali-Mikhail-Haq Family*"	$\ln[(1 - \theta(1 – t))/t]$	$[-1, 1]$

Theorem

*Let φ be a continuous, strictly decreasing function from **I** to [0, $\propto$] such that $\varphi(1) = 0$, and $\varphi^{(-1)}$ be the quasi-inverse of φ. Then the function C from **I**^{2} to **I** given by $C(u, v) = \varphi^{(-1)}(\varphi(u) + \varphi(v))$ is a copula iff φ is convex.*

Proof: see Nelsen.

Example:

*Let $\varphi(t) = (-\ln(t))^{\theta}$, where $\theta \geq 1$. Then clearly φ is continuous, strictly decreasing and $\varphi(1) = 0$. Also $\varphi''(t) \geq 0$ on **I**, so φ is convex. So for the copula we get:* $C_\theta(u, v) = \exp\{-[(-\ln(u))\theta + (-\ln(v))\theta]1/\theta\}$, which is the *Gumbel-Hougaard Family.*

Also note that $C_1 = \Pi$, and $C\propto = M$.

***Level Curves**:*

The level curves of a copula are given by $\{(u, v) \varepsilon \mathbf{I}^2 | C(u, v) = t, t \geq 0\}$. For Archimedean copulas, $t > 0$, this just the curve: $\varphi(u) + \varphi(v) = \varphi(t)$, which connects the points (1, t) and (t, 1). When t = 0, the set is called the zero set of C, denoted by Z(C).

Theorem

The level curves of an Archimedean copula are convex.

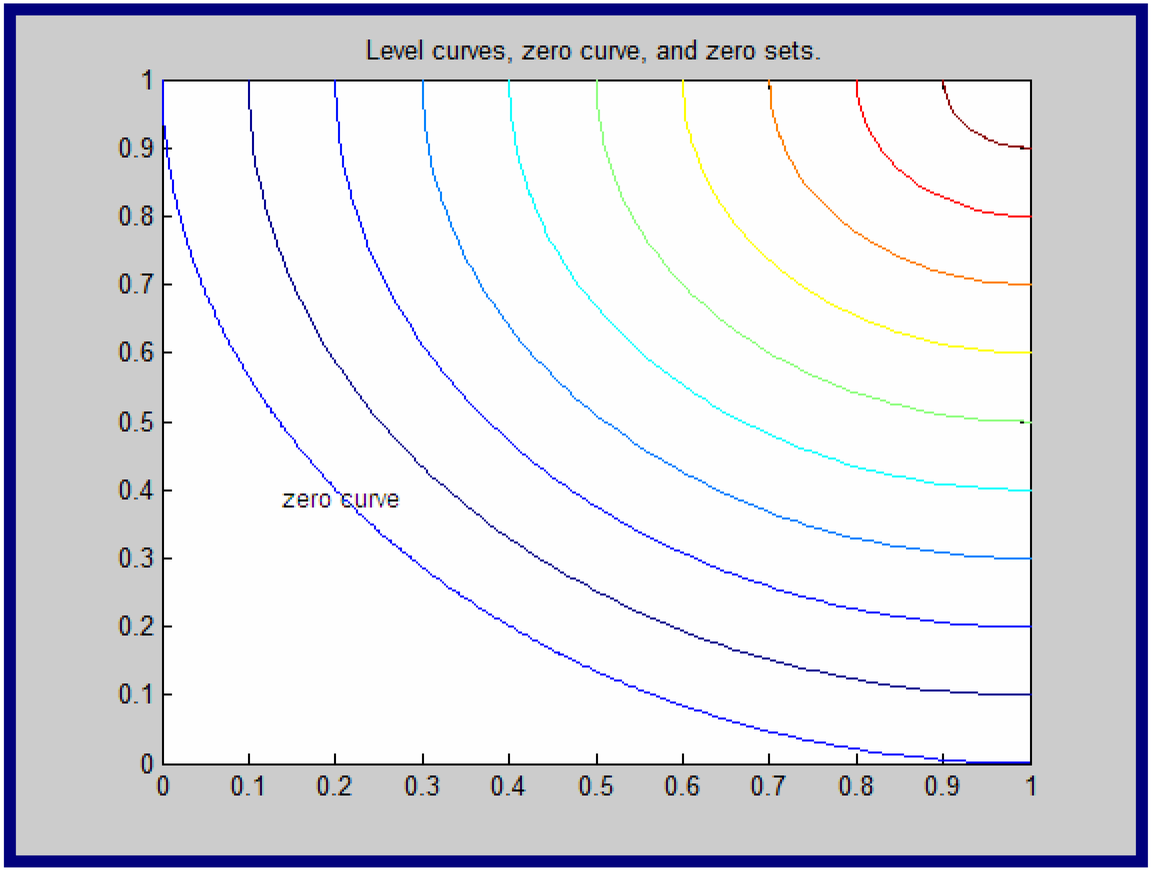
F-measure

Let X and Y be random variables in R with bivariate distribution F. Let A be a subset of R^2. Then the F-measure of A is defined by $P[(X, Y) \varepsilon A]$.

We can use this definition to determine the C-measure of the level curves of an Archimedean copula C.

$$C(u, v) = \max(1-[(1-u)^2 + (1-v)^2]^{1/2}, 0)$$

$$\varphi(t) = (1-t)^2$$



Note: it is possible for different Archimedean copulas to have the same zero set, as the following example shows:

$$\varphi_1(t) = \arctan\frac{1-t}{1+t}; \varphi_2(t) = \ln\frac{\sqrt{2}+1-t}{\sqrt{2}-1+t}$$

$$C_1(u,v) = \max(\frac{uv+u+v-1}{1+u+v-uv},0); C_2(u,v) = \max(\frac{uv+u+v-1}{3-u-v+uv},0)$$

C_1 and C_2 both have the same zero curve $v = (1 - u)/(1 + u)$, from which it follows that both have the same zero set.

***Theorem*:**

Let C be an Archimedean copula generated by φ.

1. *For t in (0, 1), the C-measure of the level curve $\varphi(u) + \varphi(v) = \varphi(t)$, is given by*

$$\varphi(t)\left[\frac{1}{\varphi'(t^-)} - \frac{1}{\varphi'(t^+)}\right]$$

   in particular if $\varphi'(t)$ exists, then the C-measure is 0.

2. *If C is not strict, i.e., $\varphi(0)$ is finite, then the C-measure of the zero curve is equal to*

$$-\frac{\varphi(0)}{\varphi'(0^+)}$$

Example:

For the copula generated by $\varphi(t) = (1 - t)^2$, we have:

- *The C-measure of level curves $C(u, v) = t$, t in $(0, 1)$ is 0.*
- *The C-measure of the zero curve is: $-\varphi(0)/\varphi`(0) = 1/2$.*

Theorem

Let C be an Archimedean copula generated by φ. Let $K_C(t)$ denote the C-measure of the set $\{(u, v) \varepsilon \mathbf{I}^2 | C(u, v) \leq t\}$. Then for any t in **I**,

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}$$

Corollary

Let U and V be uniform (0, 1) random variables whose joint distribution function is the Archimedean copula C generated by φ, a continuous strictly decreasing convex function from **I** to $[0, \infty]$. *Then the function K_C given above is the distribution function of the random variable C(U, V).*

The next theorem extends these results.

Theorem

Under the hypothesis of the previous lemma, the joint distribution function $H(s,t)$ of the random variables $S = \varphi(U)/(\varphi(U) + \varphi(V))$ and $T = C(U,V)$ is given by $H(s,t) = sK_C(t)$.

A result of this theorem is the following algorithm for generating random samples (u,v) whose joint distribution function is an Archimedean copula C with generator φ.

1. *Generate two independent standard uniform random numbers s and q.*
2. *Set* $t = K_C^{(-1)}(q)$
3. *Set* $u = \varphi^{(-1)}(s\varphi(t))$ *and* $v = \varphi^{(-1)}((1 - s)\varphi(t))$
4. The desired pair is (u,v).