Stripe Copula[1]

Construction of Stripe Copula

Construction of the "Stripe copula" starts by drawing a diagonal stripe across the unit square shown in Figure 1. The parameter "d" controls the width of the stripe, which can range from zero (perfect correlation), to covering the entire unit square uniformly (independence). The probability density in the stripe is uniform, normalized to unity, and vanishes outside the stripe. The uniform probability density of the stripe in Figure 1 is $1/(2d\text{-}d^2)$, the reciprocal of the stripe's area. By considering four distinct geometrical cases expressions for the probability density enclosed by the blue lines, $A(u,v)$ shown in the figure, may be easily found as a function of (u,v). All of these cases are covered by the formula

$$A(u,v)=\frac{u'v'-\tfrac{1}{2}a^2-\tfrac{1}{2}b^2}{d(2-d)} \tag{1}$$

where

$$\begin{aligned} u' &= \min[u, v+d] \\ v' &= \min[v, u+d] \\ a &= \max[u'-d, 0] \\ b &= \max[v'-d, 0] \end{aligned} \tag{2}$$

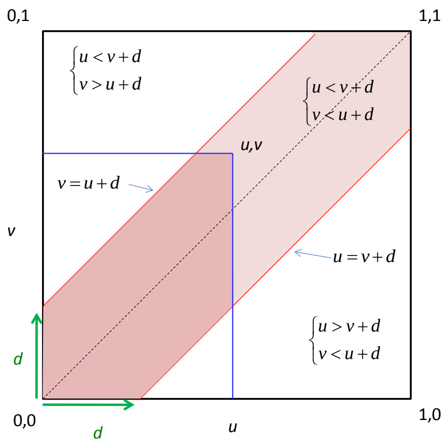


Figure 1 The diagonal stripe has uniform probability density, normalized to unity. Probability density vanishes outside the stripe. *A*(*u*,*v*) is the function which gives the probability enclosed by the rectangle (0,0)/(u,v). *A*(*u*,*v*) is a pseudo-copula because the marginal distributions are not uniform.

The function A is a pseudo-copula because it satisfies the requirements of a copula except that the margins are not uniform. The marginal distributions, $A(u,1)=f^{-1}(u)$ and $A(1,v)=f^{-1}(v)$, are non-uniform since $f^{-1}(z)\neq z$. z is a dummy argument which can be either u, or v. The functions $f^{-1}(z)$ and $f(z)$ are:

Case $d\le\tfrac{1}{2}$

$$f^{-1}(z)=\frac{1}{d(2-d)}\times\begin{cases} \tfrac{1}{2}z^2+zd & 0\le z\le d \\ 2zd-\tfrac{1}{2}d^2 & d\le z\le 1-d \\ z-\tfrac{1}{2}(z-d)^2-\tfrac{1}{2}(1-d)^2 & 1-d\le z\le 1 \end{cases} \tag{3}$$

[1] C. Glenn Shirley, October 2012.

Case $d \ge \frac{1}{2}$

$$f^{-1}(z) = \frac{1}{d(2-d)} \times \begin{cases} \frac{1}{2}z^2 + zd & 0 \le z \le 1-d \\ z - \frac{1}{2}(1-d)^2 & 1-d \le z \le d \\ z - \frac{1}{2}(z-d)^2 - \frac{1}{2}(1-d)^2 & d \le z \le 1 \end{cases} \tag{4}$$

To construct a copula we need the inverse of this function.

Case $d \le \frac{1}{2}$

$$f(z) = \begin{cases} -d + \sqrt{d^2 + 2(2d-d^2)z} & 0 \le z \le \dfrac{3d}{4-2d} \\ (1-\frac{1}{2}d)z + \frac{1}{4}d & \dfrac{3d}{4-2d} \le z \le \dfrac{4-5d}{4-2d} \\ 1 + d - \sqrt{2(2d-d^2)(1-z) + d^2} & \dfrac{4-5d}{4-2d} \le z \le 1 \end{cases} \tag{5}$$

Case $d \ge \frac{1}{2}$

$$f(z) = \begin{cases} -d + \sqrt{d^2 + 2(2d-d^2)z} & 0 \le z \le \dfrac{1-d^2}{2(2d-d^2)} \\ (2d-d^2)z + \frac{1}{2}(1-d)^2 & \dfrac{1-d^2}{2(2d-d^2)} \le z \le 1 - \dfrac{1-d^2}{2(2d-d^2)} \\ 1 + d - \sqrt{2(2d-d^2)(1-z) + d^2} & 1 - \dfrac{1-d^2}{2(2d-d^2)} \le z \le 1 \end{cases} \tag{6}$$

So the stripe copula we are seeking is

$$St(x,y) = A\big(f(x), f(y)\big) \tag{7}$$

This is a copula because it has all the properties of a copula, including uniform margins:

$$St(x,1) = A\big(f(x), f(1)\big) = A\big(f(x),1\big) = f^{-1}\big(f(x)\big) = x \tag{8}$$

The equation of the upper line bounding the area of finite probability density for this copula is

$$y = \begin{cases} f^{-1}\big(f(x)+d\big) & 0 \le x \le f^{-1}(1-d) \\ 1 & f^{-1}(1-d) \le x \le 1 \end{cases} \tag{9}$$

and of the lower line

$$y = \begin{cases} 0 & 0 \le x \le f^{-1}(d) \\ f^{-1}\big(f(x)-d\big) & f^{-1}(d) \le x \le 1 \end{cases} \tag{10}$$

The low tail dependence is

$$LT = \lim_{x \to 0+} \frac{St(x,x)}{x} = \lim_{x \to 0+} \frac{f^2(x)}{xd(2-d)} = \lim_{x \to 0+} \frac{2-d}{d} x = 0 \quad (d > 0) \tag{11}$$

so there is no asymptotic low tail dependence unless d vanishes.

Algorithms for Monte-Carlo synthesis of random points in geometrical copulas such as the stripe copula can be derived using geometrical arguments starting with the pseudo copula, A. It is possible to fill any parallelogram or triangle with uniformly distributed random points using every point generated. This is done by weighting the basis vectors which define the parallelogram with a pair of independent uniformly distributed random numbers. For a triangle, points in the "wrong half" of a parallelogram are reflected into the triangle of interest. The slice in Figure 1 can be decomposed into rectangles and triangles, and random points placed in them according to probabilities determined by area ratios of the rectangles and triangles. This produces uniformly distributed points within the stripe. Most importantly, it is possible to limit the region of the stripe over which these points are generated. Generated points may be mapped into the copula using Eqs. (3) and (4), so that if u, and v are generated for the base pseudo-copula, then the corresponding points of the copula are

$$x = f^{-1}(u) \quad y = f^{-1}(v) \tag{12}$$

Details are given below. Figure 2 shows examples of synthesized probability maps of the stripe copula. The density of random points indicates the probability density of the copula. Notice that both the shape and the density of points have been remapped from the initial probability density function (pseudo-copula) of Figure 1. Estimates of Kendall's tau are easily computed.

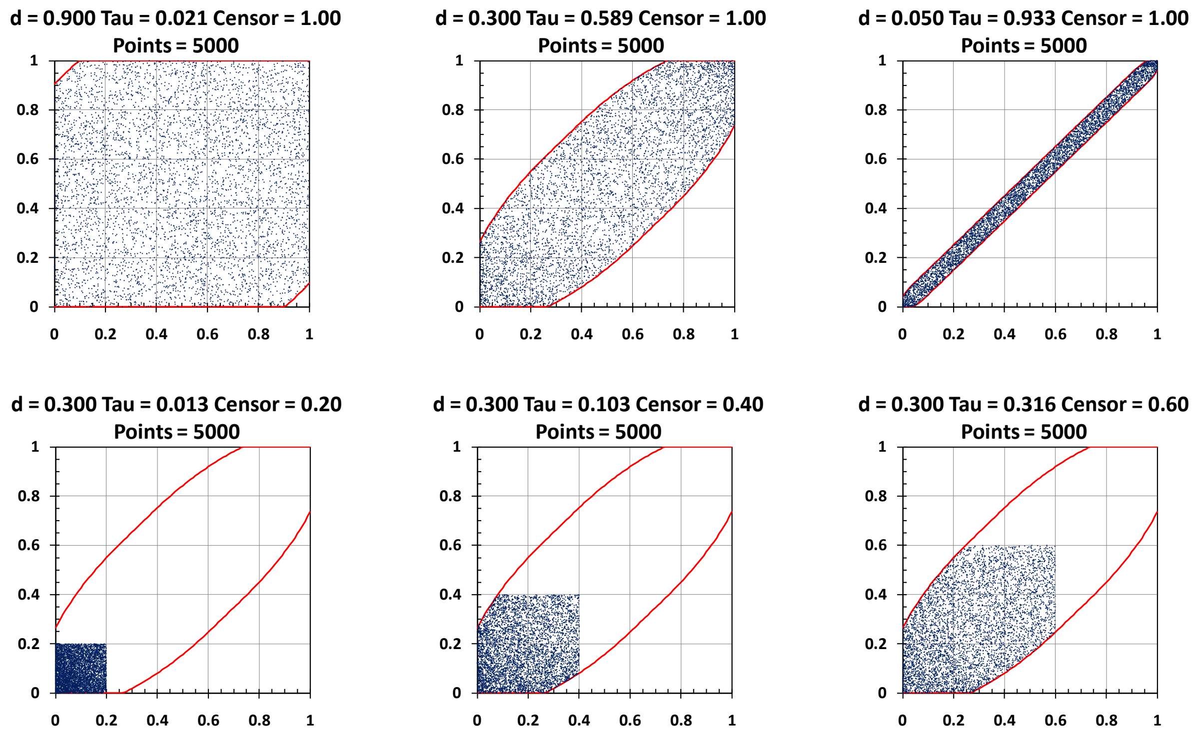
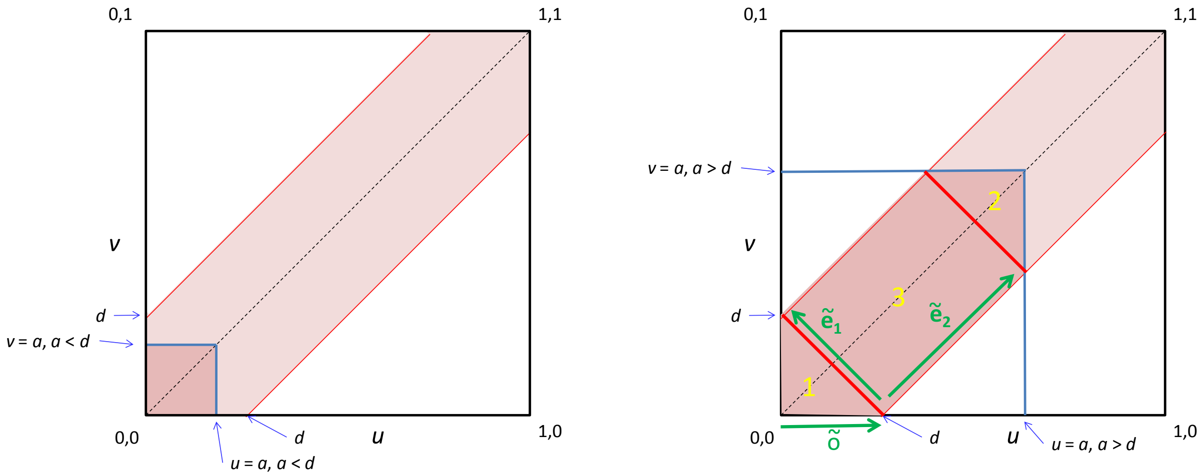


Figure 2 Synthesized probability maps of the stripe copula. Kendall's tau was computed from the synthesized data. Top Row: The stripe copula spans independence (d ≅ 1) to perfect correlation (d ≅ 0). Bottom Row: Synthesized points can be concentrated near the origin (or anywhere).

Synthesis from Stripe Copula

An algorithm for completely efficient synthesis (with no rejected points) of uniformly distributed random points inside the region of non-vanishing probability density of the pseudo-copula in Figure 3 and enclosed by $u \times v = [0,a] \times [0,a]$ is a little intricate, but the derivation follows straightforward principles which can be applied to any geometrical copula constructed in this manner. First, an algorithm to fill the diagonal stripe of the base pseudo-copula shown in Figure 3 with a uniform density of random points is derived based on decomposition of the region $[0,\ a]^2$ into triangles and rectangles. Then, the resulting points are mapped into the copula using the marginal distribution functions.



Case A. $a \le d$ Case B. $a \ge d$

Figure 3 Geometrical constructions on the stripe pseudo-copula to derive a Monte-Carlo sampling algorithm for a censored sub-population of the stripe copula.

For the diagonal stripe of the pseudo-copula, two cases may be identified. Case A, merely fills a small square with uniform random points. Case B divides the region to be filled into two halves of a square, regions 1 and 2, separated by a rectangle, 3. A uniformly distributed random number, u_1, is generated to decide whether to place the point in 3, or in the divided square, 1 and 2. This decision is based on the area ratio of rectangle 3 versus the divided square, 1 and 2. If the point goes into 3, two uniformly distributed random numbers u_2 and u_3 are used to place a point in 3. This is done in terms of basis vectors which span 3. On the other hand, if the point is to be placed in the divided square, 1 and 2 the point is placed in a d x d square, but if $u_2 + u_3$ exceeds unity the point is displaced by (a,a) so that it falls into the triangle 2.

Case A $0 \le a \le d$

Generate two random numbers from the uniform distribution on [0,1], u_1 , u_2 and place a point at

$$u = au_1,\ \ v = au_2 \tag{13}$$

Case B $d \le a \le 1$

Generate three random numbers from the uniform distribution on [0,1], u_1 , u_2 and u_3.

If $0 \le u_1 \le d/(2a-d)$ and if $u_2 + u_3 < 1$ place a point at (case B, region 1)

$$u = du_2,\ \ v = du_3 \tag{14}$$

If $0 \le u_1 \le d/(2a-d)$ and if $u_2 + u_3 \ge 1$ place a point at (case B, region 2)

$$u = a + d(u_2 - 1) \quad v = a + d(u_3 - 1) \tag{15}$$

If $d/(2a-d) \le u_1 \le 1$ then place a point at (case B region 3)

$$\begin{aligned} u &= d(1-u_2) + (a-d)u_3 \\ v &= du_2 + (a-d)u_3 \end{aligned} \tag{16}$$

The generated point is mapped to the copula by

$$x = f^{-1}(u) \quad y = f^{-1}(v) \tag{17}$$

where f^{-1} is given by Eqs. (3) and (4).

We have used the fact that the sum of areas 1 and 2 in the figure is a fraction $d/(2a-d)$ of the sum of areas 1, 2, and 3 in the figure. To derive Eq. (16) note that e_1 and e_2 span region 3 in case B, and the point $[d,0]$ is $\tilde{o}$,

$$\begin{aligned} \tilde{e}_1 &= (-\hat{i} + \hat{j})d \\ \tilde{e}_2 &= (\hat{i} + \hat{j})(a-d) \\ \tilde{o} &= \hat{i}\, d \end{aligned} \tag{18}$$

where $\hat{i}$ and $\hat{j}$ are orthogonal unit vectors spanning $[0,1]^2$. So for region 3, we place a point at

$$\tilde{r} = \tilde{o} + \tilde{e}_1 u_2 + \tilde{e}_2 u_3 = \left(d - du_2 + (a-d)u_3\right)\hat{i} + \left(u_2 d + (a-d)u_3\right)\hat{j} \tag{19}$$

from which Eq. (16) follows.