Wedge Copula[1]

This document explains the construction and properties of a particular"geometrical" copula used to fit dependency data from the eDRAM case study done at Portland State University. The probability density function of the "Wedge copula" resulting copula is shown in Figure 1.

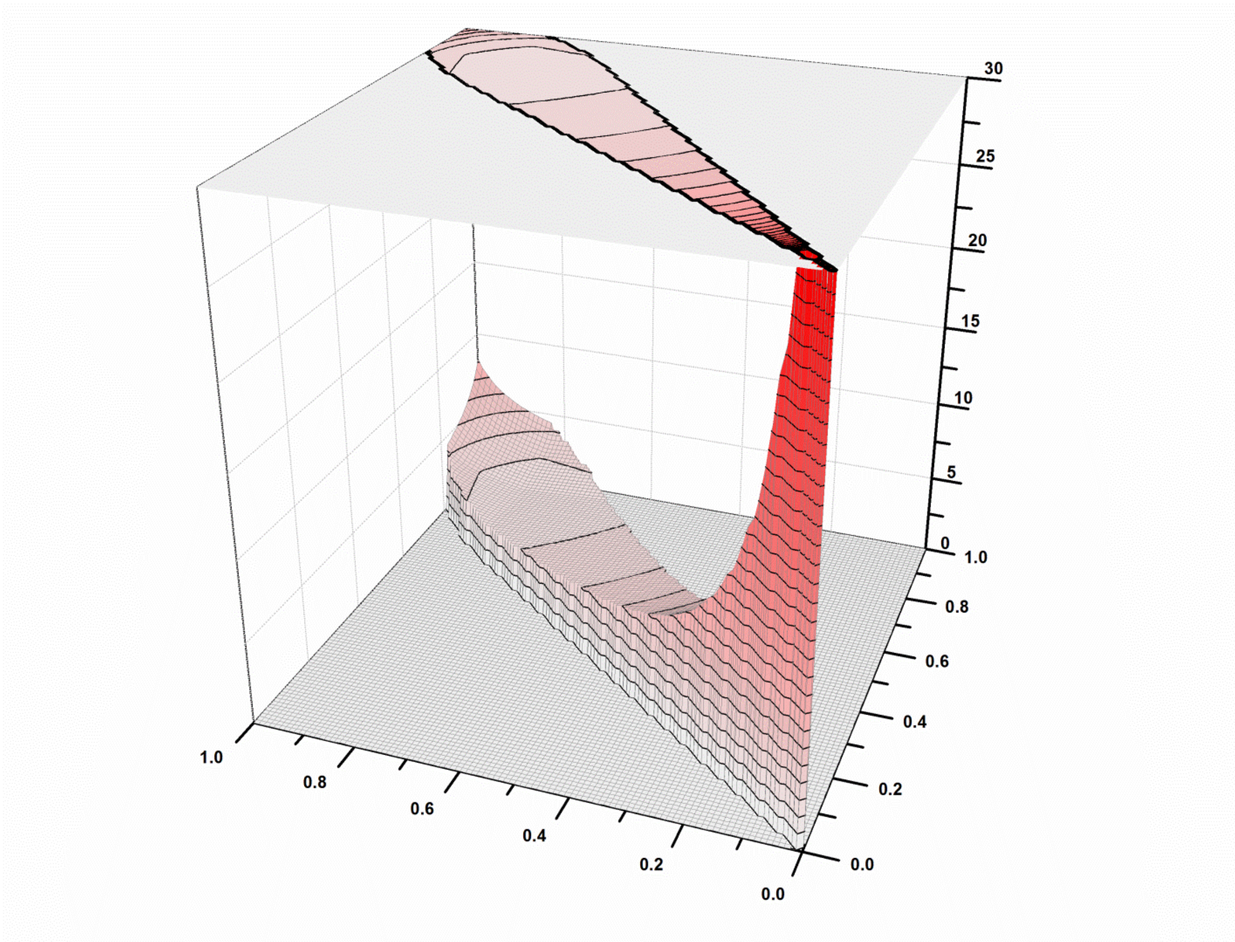

Figure 1 Probability density function of Wedge copula for c = 1.142 (best fit to Nominal skew of DRAM case study).

A wedge-shaped copula C is defined by starting with a pseudo-copula $A(u,v)$, consisting of a shaded wedge-shaped region, symmetrical about the (0,0)/(1,1) diagonal, shown in Figure 2. This shaded region has area $(1 - c)/c$ where c is defined in the figure. If the wedge is a region of uniform probability density normalized to unity within the unit square, and the probability density outside the wedge vanishes, then the wedge's uniform probability density is $c/(1 - c)$.

[1] C. Glenn Shirley, October 2012.

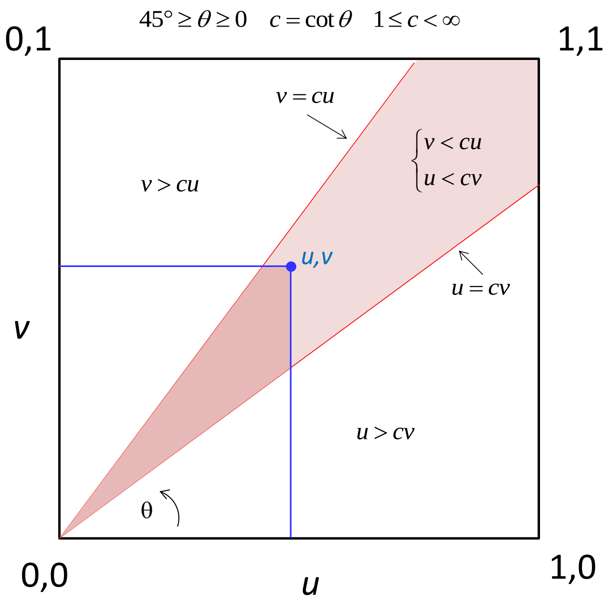

Figure 2 The base pseudo-copula *A*(*u*,*v*) is a function on [0,1]^2 which is defined by a wedge-shaped region of uniform probability density, normalized to unity within the unit square. Regions outside the wedge have vanishing probability density. The size of the region is controlled by the parameter 1 ≤ *c* ≤ ∞ which ranges from *c* = 1 for perfect correlation and to *c* = ∞ for independence

Geometrical considerations show that the probability density enclosed by the blue (0,0)/(*u*,*v*) rectangle in Figure 2 for (*u*,*v*) anywhere in $[0,1]^2$, is

$$A(u,v)=\frac{c}{c-1}\left(u'v'-\frac{u'^2}{2c}-\frac{v'^2}{2c}\right)\qquad u'=\min\left[u,cv\right],\, v'=\min\left[v,cu\right] \tag{1}$$

On the margins, this is

$$A(u,1)=f^{-1}(u)\qquad A(1,v)=f^{-1}(v) \tag{2}$$

where (*z* is a dummy argument)

$$f^{-1}(z)=\begin{cases}\dfrac{(c+1)z^2}{2} & 0\le z\le c^{-1}\\ z-\dfrac{(1-z)^2}{2(c-1)} & c^{-1}\le z\le 1\end{cases} \tag{3}$$

which is a monotonically increasing function. Notice that *A*(*u*,*v*) satisfies the requirements of a copula, except that the marginal distributions, Eqs. (2), are not uniform. To construct a "wedge copula", with uniform margins we need the inverse of Eq. (3):

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

So the wedge copula corresponding to Figure 2 is

$$C(x,y)=A\big(f(x),f(y)\big) \tag{5}$$

which satisfies the requirement that it have a uniform marginal distribution because

$$C(x,1) = A\big(f(x), f(1)\big) = A\big(f(x),1\big) = f^{-1}\big(f(x)\big) = x \tag{6}$$

and the same for *y*, since *x* and *y* are cdfs of uniform distributions. Since

$$\begin{aligned} x &= f^{-1}(u) \\ y &= f^{-1}(v) \end{aligned} \tag{7}$$

and since the lower boundary of the probability area of *A* is defined by *u* = *cv* then the corresponding boundary of *C* is given by

$$y = f^{-1}\left(c^{-1} f(x)\right) \qquad 0 \le x \le 1 \tag{8}$$

Similarly, the upper boundary of the probability area of *A* is defined by

$$v = \begin{cases} cu & 0 \le u \le c^{-1} \\ 1 & c^{-1} \le u \le 1 \end{cases} \tag{9}$$

and the corresponding upper boundary of the probability area of *C* is

$$y = \begin{cases} f^{-1}\left(cf(x)\right) & 0 \le x \le \tfrac{1}{2}(c+1)/c^2 \\ 1 & \tfrac{1}{2}(c+1)/c^2 \le x \le 1 \end{cases} \tag{10}$$

The zone of non-vanishing probability of the copula *C* is defined by simultaneous satisfaction of

$$y > f^{-1}\left(c^{-1} f(x)\right) \qquad 0 \le x \le 1 \tag{11}$$

$$y < \begin{cases} f^{-1}\left(cf(x)\right) & 0 \le x \le \tfrac{1}{2}(c+1)/c^2 \\ 1 & \tfrac{1}{2}(c+1)/c^2 \le x \le 1 \end{cases} \tag{12}$$

For the test application the important region is the region near the origin. In this region the boundaries of the regions with finite probability given by Eqs. (8) and (10) are straight lines. For the lower boundary

$$y = c^{-2} x \qquad 0 \le x \le \frac{c+1}{2c^2} \tag{13}$$

and for the upper boundary

$$y = c^2 x \qquad 0 \le x \le \frac{c+1}{2c^4} \tag{14}$$

The low tail dependence of the wedge copula is

$$LT = \lim_{x\to 0+} \frac{C(x,x)}{x} = \lim_{x\to 0+} \frac{f^2(x)}{x} = \lim_{x\to 0+} \frac{1}{x}\frac{2x}{c+1} = \frac{2}{c+1} \tag{15}$$

So the wedge copula has asymptotic tail dependence except in the limit of independence ($c \to \infty$). In the opposite limit ($c \to 1$), the asymptotic dependence becomes unity because C becomes M, the Frechet upper bound, corresponding to perfect correlation.

Points of the wedge copula may easily be synthesized by generating uniformly distributed (u,v) points in the wedge-shaped finite uniform probability density area of the base pseudo-copula in Figure 2, and then mapping them to the space of the copula using $x = f(u)$, and $y = f(v)$ where f is given by Eq.(4). The algorithm for generating (u, v) points is given in the Appendix. The probability density maps of the wedge copula in Figure 3 were generated using this method. Notice that it is possible to restrict the generation of synthesized points to sub-domains of the copula, particularly the region near the origin.

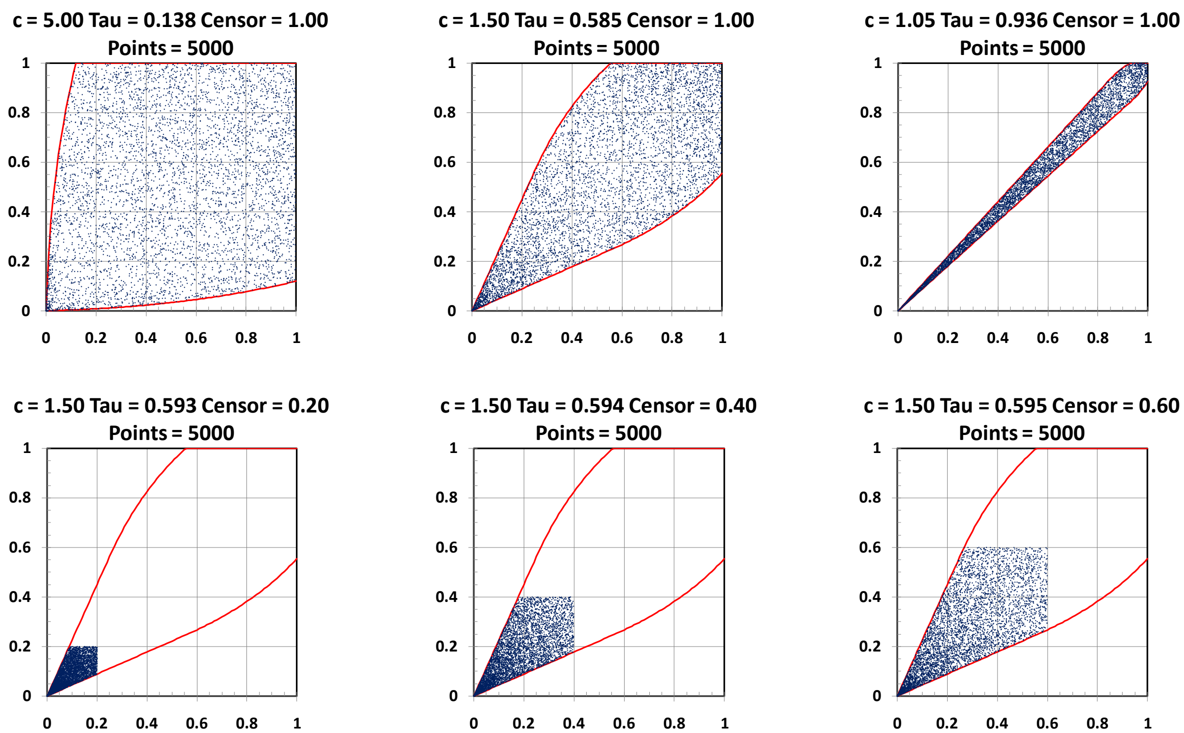

Figure 3 Synthesized probability maps of the wedge copula. Kendall's tau was computed from the synthesized data. Top Row: The wedge copula spans independence ($c \cong \infty$) to perfect correlation ($c \cong 1$). Bottom Row: Synthesized points can be concentrated near the origin.

Values of Kendall's tau given in Figure 3 for the synthesized data were estimated to good precision from synthesized data. It is also possible to derive an analytical expression for Kendall's tau for the copula as a function of the parameter c, and of the censor fraction a, using the expression for tau of a truncated part of a copula[2].

$$\tau_{\text{Subpopulation}}(c,a) = \frac{4}{We^2(a,a;c)} \int_0^a dx \int_0^a dy We(x,y;c) \frac{\partial^2 We(x,y;c)}{\partial x \partial y} - 1 \tag{16}$$

The result, derived in the Appendix is

$$\tau_{\text{Subpopulation}}(c,a) = \frac{2c+1}{3c^2} \tag{17}$$

[2] $\tau(u,v) = \frac{4}{C^2(u,v)} \int_0^u \int_0^v dxdy C(x,y) \frac{\partial^2 C(x,y)}{\partial x \partial y} - 1$

The inverse of Eq. (17) is

$$c = \frac{1+\sqrt{1+3\tau}}{3\tau}. \tag{18}$$

The wedge copula has the attractive property that the subpopulation tau is independent of a, the degree of censoring. This is not true of copulas in general. The small variation of tau in Figure 3 for constant c is due to sampling variation of the Monte-Carlo estimate of tau. Since tau may be computed directly from data, Eq. (17) provides a way to estimate the parameter c of the copula.

Since it is convenient if the subpopulation tau is independent of the degree of censoring, it is useful to know more general conditions under which this holds so that copulas with this property can be identified. A sufficient condition for this to be true is that a copula be expressed as $C(x,y) = A[f(x),f(y)]$ where A satisfies, for $a \le 1$, $A(a\times u,a\times v) = a^2 \times A(u, v)$. This is shown in the Appendix. The geometrical interpretation is that all sub-regions $[0,a]^2$ of the base pseudo-copula are geometrically self-similar.

Wedge Copula

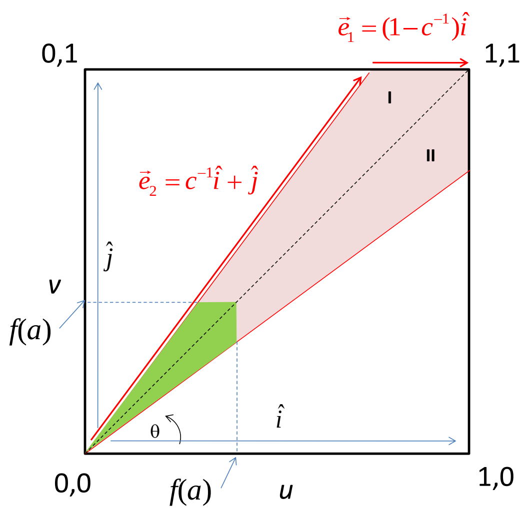

Figure A 1 Basis vectors (red) for sampling the area of uniform probability in the pseudo-copula, *A*(*u*,*v*). A subset of the space, such as the green area, may be sampled by scaling the basis vectors.

To place a random point r in triangle I in the figure in (u,v) space, sample u_1 and u_2 independently from the uniform distribution on [0,1], rejecting the sample if $u_1 > u_2$.

$$\begin{aligned}\vec{r} &= u_1\vec{e}_1 + u_2\vec{e}_2 \\ &= [u_1 + (u_2 - u_1)/c]\hat{i} + u_2\hat{j}\end{aligned} \tag{19}$$

In Eq. (19) $\vec{e}_1$ and $\vec{e}_2$ are basis vectors which span triangle I in the figure. Decomposition into the orthogonal unit vectors spanning the unit square in the figure gives the second equation. By symmetry, to place a random point r point in triangle II in the figure, reject the sample if $u_2 > u_1$, and place the point at

$$\vec{r} = u_1\hat{i} + [u_2 + (u_1 - u_2)/c]\hat{j} \tag{20}$$

To sample a subpopulation of the copula such as the sub area shown in Figure A 1, the uniform random variables u_1 and u_2 should be independently sampled from $[0,f(a)]$.

So the algorithm to sample a region $x \times y = [0,a]\times[0,a]$ of the wedge copula is (a = 1 samples the entire copula):

1. Sample two independent uniform numbers, u_1 and u_2 from a uniform distribution on $[0,f(a)]$.

2. If $u_2 \geq u_1$ then place a point at

$$x = f^{-1}\left(u_1 + \frac{u_2 - u_1}{c}\right) \qquad y = f^{-1}(u_2) \tag{21}$$

3. Else if $u_2 < u_1$ then place a point at

$$x = f^{-1}(u_1) \qquad y = f^{-1}\left(u_2 + \frac{u_1 - u_2}{c}\right) \tag{22}$$

where f^{-1} is given by Eq. (3).

Subpopulation Tau

We seek an expression for tau of a subpopulation of a copula in the region $J^2 = [0,u_a]\times[0,v_b]$. To do this, we write down an expression for the copula of the subpopulation, and substitute into the formula for tau, The probability density function for the region J^2 is

$$D'(u,v) = \frac{C(u,v)}{C(u_a,v_b)} \tag{23}$$

This may be converted into a copula by using the marginal distribution functions

$$u' = \frac{C(u,v_b)}{C(u_a,v_b)} = f(u) \qquad\qquad v' = \frac{C(u_a,v)}{C(u_a,v_b)} = g(v) \tag{24}$$

Notice that these are not uniform, $f(u) \neq u$, and $g(v) \neq v$, because $u_a \neq 1$ and $v_b \neq 1$.

The copula for the region J^2, in terms of the transformed variables, is

$$D(u',v') = \frac{C(u,v)}{C(u_a,v_b)} = \frac{C\left(f^{-1}(u'), g^{-1}(v')\right)}{C(u_a,v_b)} \qquad [u',v'] \in [0,1]^2 \tag{25}$$

So the subpopulation tau is

$$\begin{aligned}
\tau_{\text{Subpopulation}} &= 4\int_0^1\int_0^1 D(u',v')\frac{\partial D(u',v')}{\partial u'\partial v'}du'dv' - 1 \\
&= 4\int_0^1\int_0^1 D(u',v')\frac{\partial D(u',v')}{\partial u\partial v}\frac{\partial u}{\partial u'}\frac{\partial v}{\partial v'}du'dv' - 1 \\
&= 4\int_0^{u_a}\int_0^{v_b} D(u',v')\frac{\partial D(u',v')}{\partial u\partial v}dudv - 1 \\
&= \frac{4}{C^2(u_a,v_b)}\int_0^{u_a}\int_0^{v_b} C(u,v)\frac{\partial C(u,v)}{\partial u\partial v}dudv - 1
\end{aligned} \tag{26}$$

Sufficient Condition for Censor-Independence of Subpopulation Tau

Suppose that a copula is expressed as

$$C(x,y) = A\left[f(x), f(y)\right] \tag{27}$$

where A is a pseudo-copula satisfying, for $a \le 1$,

$$A\left(a\times u, a\times v\right) = a^2 A(u,v) \tag{28}$$

The subpopulation tau of C is

$$\begin{aligned}
\tau_{\text{Subpopulation}} &= \frac{4}{C^2(\alpha,\alpha)}\int_0^\alpha du\int_0^\alpha dv C(u,v)\frac{\partial C(u,v)}{\partial u\partial v}-1 \\
&= \frac{4}{A^2\left[f(\alpha), f(\alpha)\right]}\int_0^\alpha du\int_0^\alpha dv A\left[f(u), f(v)\right]\frac{\partial A\left[f(u), f(v)\right]}{\partial u\partial v}-1 \\
&= \frac{4}{f^4(\alpha)A(1,1)}\int_0^{f(\alpha)} dx\int_0^{f(\alpha)} dy\frac{\partial u\partial v}{\partial x\partial y}A(x,y)\frac{\partial A(x,y)}{\partial u\partial v}-1 \\
&= \frac{4}{f^4(\alpha)}\int_0^{f(\alpha)} dx\int_0^{f(\alpha)} dy A(x,y)\frac{\partial A(x,y)}{\partial x\partial y}-1
\end{aligned} \tag{29}$$

Now set $x = f(\alpha)\times x'$ and $y = f(\alpha)\times y'$, so

$$\begin{aligned}
\tau_{\text{Subpopulation}} &= \frac{4}{f^4(\alpha)}\int_0^1 f(\alpha)dx'\int_0^1 f(\alpha)dy' A\left[(f(\alpha)x', f(\alpha)y')\right]\frac{\partial A\left[f(\alpha)x', f(\alpha)y'\right]}{\partial x'\partial y'}\frac{1}{f^2(\alpha)}-1 \\
&= \frac{4}{f^4(\alpha)}\int_0^1 f(\alpha)dx'\int_0^1 f(\alpha)dy' f^2(\alpha)A(x',y')\frac{f^2(\alpha)\partial A(x',y')}{\partial x'\partial y'}\frac{1}{f^2(\alpha)}-1 \\
&= 4\int_0^1 dx'\int_0^1 dy' A(x',y')\frac{\partial A(x',y')}{\partial x'\partial y'}-1
\end{aligned} \tag{30}$$

which is independent of the degree of censoring, QED. Moreover, the subpopulation tau for any degree of censoring is the same as the population tau.

Tau of Wedge Copula

Kendall's tau for the wedge copula is

$$\tau(c) = 4\int_0^1 dx\int_0^1 dy We(x,y;c)\frac{\partial^2 We(x,y;c)}{\partial x\partial y}-1 = 4I-1 \tag{31}$$

Since

$$\begin{aligned}
&We(x,y) = A(u,v) \\
&u = f(x), v = f(y)
\end{aligned} \tag{32}$$

we have

$$\begin{aligned} I &= \int_0^1 dx \int_0^1 dy We(x,y) \frac{\partial^2 We(x,y)}{\partial x \partial y} \\ &= \int_0^1 du \frac{dx}{du} \int_0^1 dv \frac{dy}{dv} A(u,v) \frac{\partial^2 A(u,v)}{\partial u \partial v} \frac{du}{dx} \frac{dv}{dy} \\ &= \int_0^1 du \int_0^1 dv A(u,v) \frac{\partial^2 A(u,v)}{\partial u \partial v} \end{aligned} \tag{33}$$

so the evaluation of tau can be done entirely in terms of the pseudo copula A. Evaluation of the integral is facilitated by noticing: 1) The second mixed derivative of $A(u,v)$ is the pdf of A, which is a constant equal to $c/(c\text{-}1)$ inside the wedge and zero elsewhere, and 2) By symmetry across the diagonal, the desired integral is twice the integral of the shaded zone in Figure A 2. 3) Because $A(u,v)$ vanishes outside the wedge, the v-integration limits may be changed so that for each u, v ranges from u/c to u. So

$$I = 2\int_0^1 du \int_{u/c}^{u} dv A(u,v) \frac{c}{c-1} \tag{34}$$

Inside the wedge, where the argument of Eq. (34) is evaluated, we have, from Eq. (1)

$$A(u,v) = \frac{c}{c-1}\left(uv - \frac{u^2}{2c} - \frac{v^2}{2c} \right) \qquad \text{(Inside wedge.)} \tag{35}$$

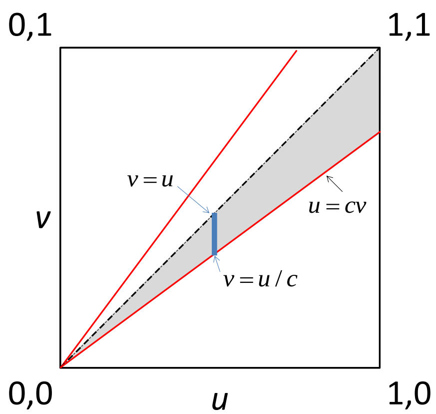

Figure A 2 Integration limits for *I*.

So

$$I = 2\int_0^1 du \int_{u/c}^{u} dv \left(\frac{c}{c-1} \right)^2 \left(uv - \frac{1}{2c}(u^2 + v^2) \right) = 2\left(\frac{c}{c-1} \right)^2 \left(I_a + I_b + I_c \right) \tag{36}$$

where

$$I_a = \int_0^1 du \int_{u/c}^{u} dv\, uv = \int_0^1 du\, u \times \tfrac{1}{2} v^2 \Big|_{u/c}^{u} = \frac{1}{2}\left(1 - \frac{1}{c^2}\right) \int_0^1 du\, u^3 = \frac{1}{8}\left(1 - \frac{1}{c^2}\right) \tag{37}$$

$$I_b = -\frac{1}{2c}\int_0^1 du \int_{u/c}^{c} dv\, u^2 = -\frac{1}{2c}\int_0^1 u^2 du \times v\Big|_{u/c}^{u} = -\frac{1}{2c}\left(1 - \frac{1}{c}\right)\int_0^1 u^3 du = -\frac{1}{8c}\left(1 - \frac{1}{c}\right) \tag{38}$$

$$I_c = -\frac{1}{2c}\int_0^1 du \int_{u/c}^{u} dv\, v^2 = -\frac{1}{2c}\int_0^1 du\left(\tfrac{1}{3}v^3\Big|_{u/c}^{u}\right) = -\frac{1}{6c}\left(1-\frac{1}{c^3}\right)\int_0^1 du\, u^3 = -\frac{1}{24c}\left(1-\frac{1}{c^3}\right) \tag{39}$$

So

$$\begin{aligned} I_a + I_b + I_c &= \frac{1}{8}\left(1-\frac{1}{c^2}\right) - \frac{1}{8c}\left(1-\frac{1}{c}\right) - \frac{1}{24c}\left(1-\frac{1}{c^3}\right) \\ &= \frac{1}{8c^4}\left(c^4 - \cancel{c^2} - c^3 + \cancel{c^2} - \frac{1}{3}c^3 + \frac{1}{3}\right) \\ &= \frac{1}{24c^4}\left(3c^4 - 4c^3 + 1\right) \\ &= \frac{(c-1)^2(3c^2+2c+1)}{24c^4} \end{aligned} \tag{40}$$

and from Eq. (36)

$$I = \frac{3c^2+2c+1}{12c^2} \tag{41}$$

Substitution of Eqs. (41) into Eq. (31) gives

$$\tau_{\text{Subpopulation}} = \frac{2c+1}{3c^2} \tag{42}$$

Probability Density Function

The probability density is non-vanishing only for x,y pairs which satisfy

$$y > f^{-1}\left(c^{-1}f(x)\right) \qquad 0 \le x \le 1 \tag{43}$$

$$y < \begin{cases} f^{-1}\left(cf(x)\right) & 0 \le x \le \tfrac{1}{2}(c+1)/c^2 \\ 1 & \tfrac{1}{2}(c+1)/c^2 \le x \le 1 \end{cases} \tag{44}$$

Inside this region we need the probability density function given by

$$c(x,y) = \frac{\partial^2 C}{\partial x \partial y} = \frac{\partial^2 A}{\partial u \partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} = \frac{c}{c-1}\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \tag{45}$$

From Eq. (4) we have

$$\frac{\partial f}{\partial x} = \begin{cases} f_l(x) = \dfrac{1}{\sqrt{2(c+1)}}\dfrac{1}{\sqrt{x}} & 0 \le x \le \dfrac{1+c}{2c^2} \\ f_h(x) = \dfrac{c-1}{\sqrt{(c-1)^2 + 2(c-1)(1-x)}} & \dfrac{1+c}{2c^2} \le x \le 1 \end{cases} \tag{46}$$