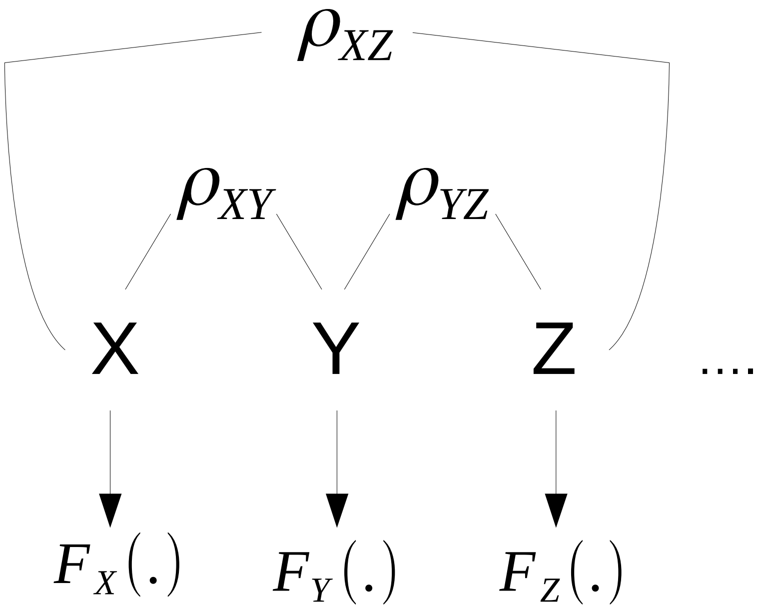
Simulation of random variables with prescribed marginals and given Pearson‘s correlation coefficients

Barbora Kessel

Helmholtz Centre for Infection Research, Braunschweig
Georg-August University, Göttingen

Problem

Generate (X,Y,Z,...) such that:



- finite second moments
- $if\ U \sim Unif(0,1), then\ F_*^{-1}(U) \sim F_*, * = X, Y, Z, \ldots$
- continuous marginals

Motivation

Monte Carlo evaluation of uncertainties

Input quantities → $f(X, Y, Z, ...)$ → Ouput quantity

https://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf

GUM (Guide to the expression of uncertainty in measurement), Suppl. 1

Dukic, Marić (2013)

On minimum correlation in construction of multivariate distributions

Physical Review E, 87, 032114

Fréchet-Hoeffding bounds

$E(X \cdot Y)$ ←

$F_{(X,Y)}(x,y) = C(F_X(x), F_Y(y))$

$$\rho_{XY} = \frac{\int_0^1 \int_0^1 F_X^{-1}(u) F_Y^{-1}(v)\, dC(u,v) - EX\, EY}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$W_2(u,v) \leq C(u,v) \leq M_2(u,v)$$

$U \sim Unif(0,1)$ $\quad W_2 \downarrow (U, 1-U) \quad M_2 \downarrow (U, U)$

Possible correlations: $\rho \in [\rho_{min}, \rho^{max}] \subseteq [-1,1]$

$$\rho_{min} = \frac{\int_0^1 F_X^{-1}(u) F_Y^{-1}(1-u)\, du - EX\, EY}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$\rho^{max} = \frac{\int_0^1 F_X^{-1}(u) F_Y^{-1}(u)\, du - EX\, EY}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Working example

$$X \sim t_5, Y \sim Unif(0,1)$$

$$\rho_{min} = -0.9284$$

$$\rho^{max} = 0.9284$$

Target correlation: 0.54

Algorithm 1, Dukic, Marić (2013)

$\rho \geq 0$

U, V ~ Unif(0,1), independent, $a \in [0,1]$ such that $a * \rho^{max} = \rho$

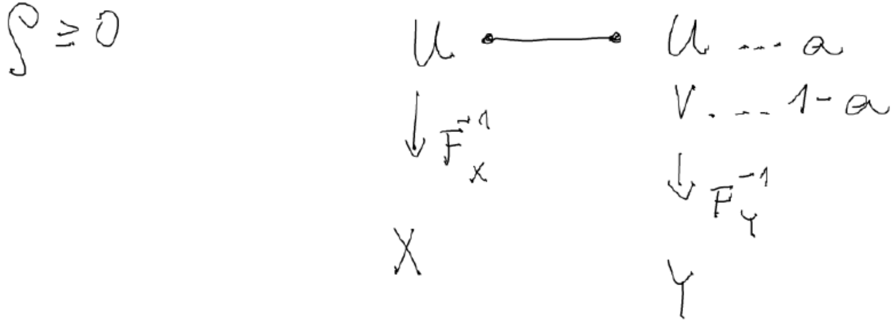
1. generate (U,U) with probability a → $X = F_X^{-1}(U), Y = F_Y^{-1}(U)$
2. generate (U,V) with probability $1-a$ → $X = F_X^{-1}(U), Y = F_Y^{-1}(V)$

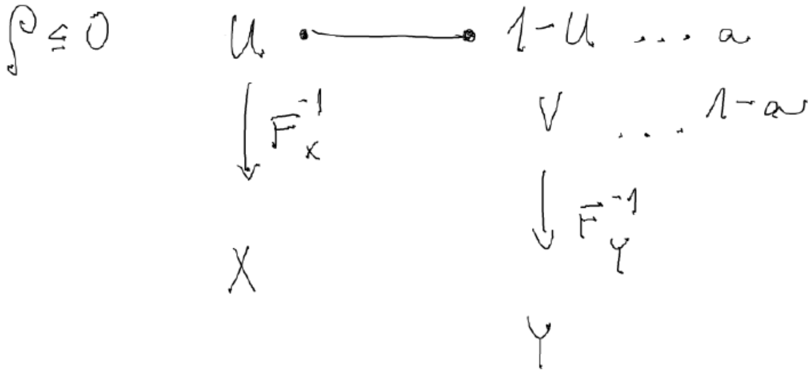
$\rho \leq 0$

U, V ~ Unif(0,1), independent, $a \in [0,1]$ such that $a * \rho_{min} = \rho$

1. generate (U,1-U) with probability a → $X = F_X^{-1}(U), Y = F_Y^{-1}(1-U)$
2. generate (U,V) with probability $1-a$ → $X = F_X^{-1}(U), Y = F_Y^{-1}(V)$

Algorithm 1, Dukic, Marić (2013)





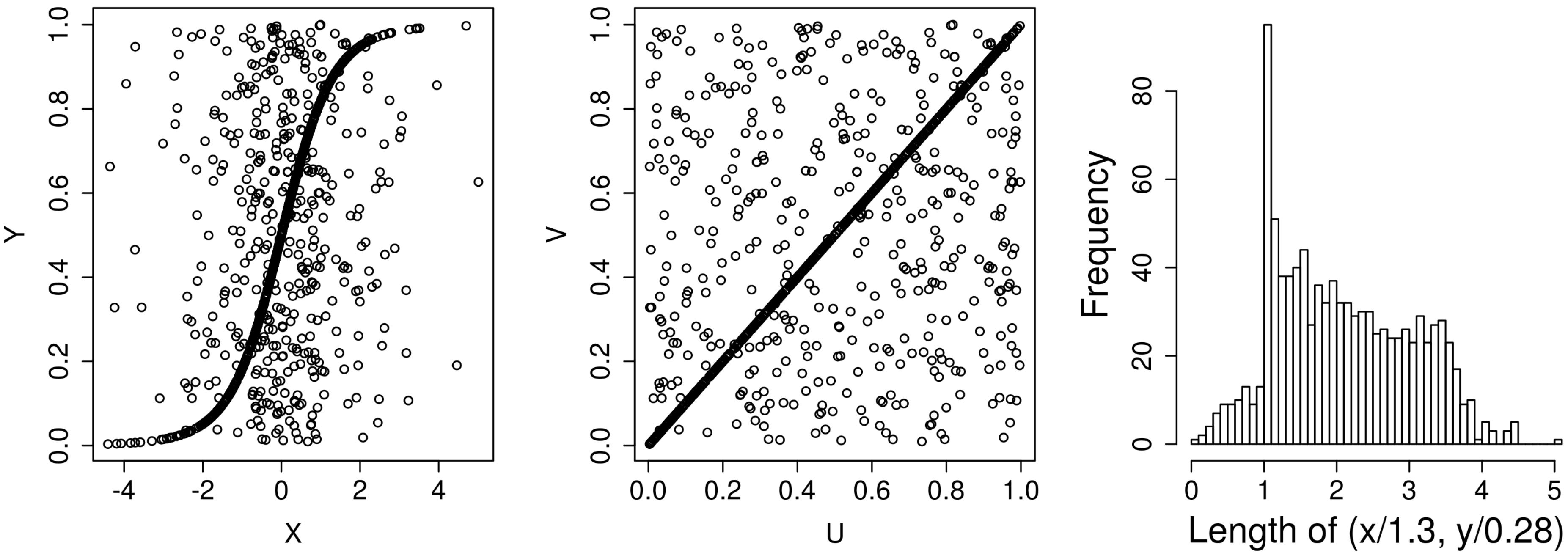
Equivalently:

The univariate distributions are coupled using

$\rho \geq 0$ $$C(u,v) = a * M_2(u,v) + (1-a) * \Pi_2(u,v)$$

$\rho \leq 0$ $$C(u,v) = a * W_2(u,v) + (1-a) * \Pi_2(u,v)$$

Target correlation 0.54



Modifications

$$C(u,v)=a*C_{normal,r_{up}}(u,v)+(1-a)*C_{normal,r_{lo}}(u,v)$$

$$a\rho^{up}+(1-a)\rho_{lo}=\rho$$

$$a\in[0,1]$$

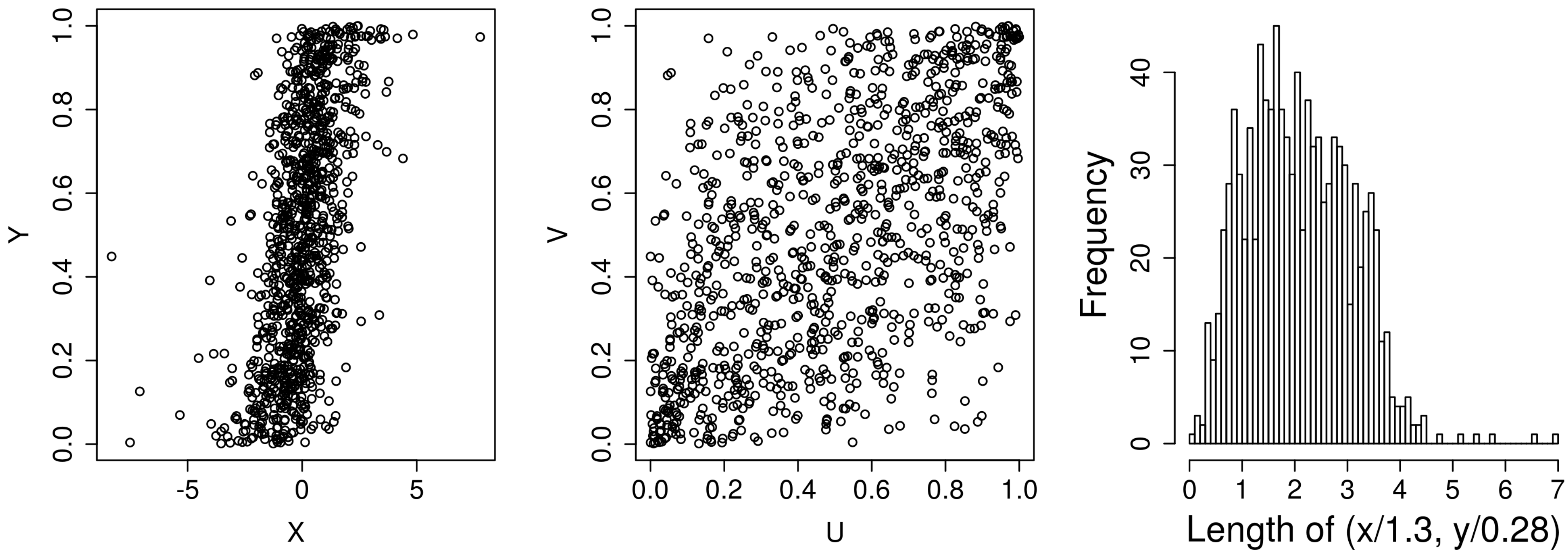
$$\rho_{lo}=\frac{\int_0^1\int_0^1 F_X^{-1}(u)F_Y^{-1}(v)\,dC_{normal,r_{lo}}(u,v)-EX\,EY}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$\rho^{up}=\frac{\int_0^1\int_0^1 F_X^{-1}(u)F_Y^{-1}(v)\,dC_{normal,r_{up}}(u,v)-EX\,EY}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Working example - convex combination of two normal copulas with r=0.7 and r=0.5

Target correlation 0.54

$$a = 0.34527, \ \ \rho_{lo} = 0.47616, \ \ \rho^{up} = 0.661057$$



Does the simple algorithm generalize to 3 or even more variables X_1, X_2, X_3,...?

Algorithm 3, Dukic, Marić (2013) [correctly formulated]

- formulated for identical marginals

- general n

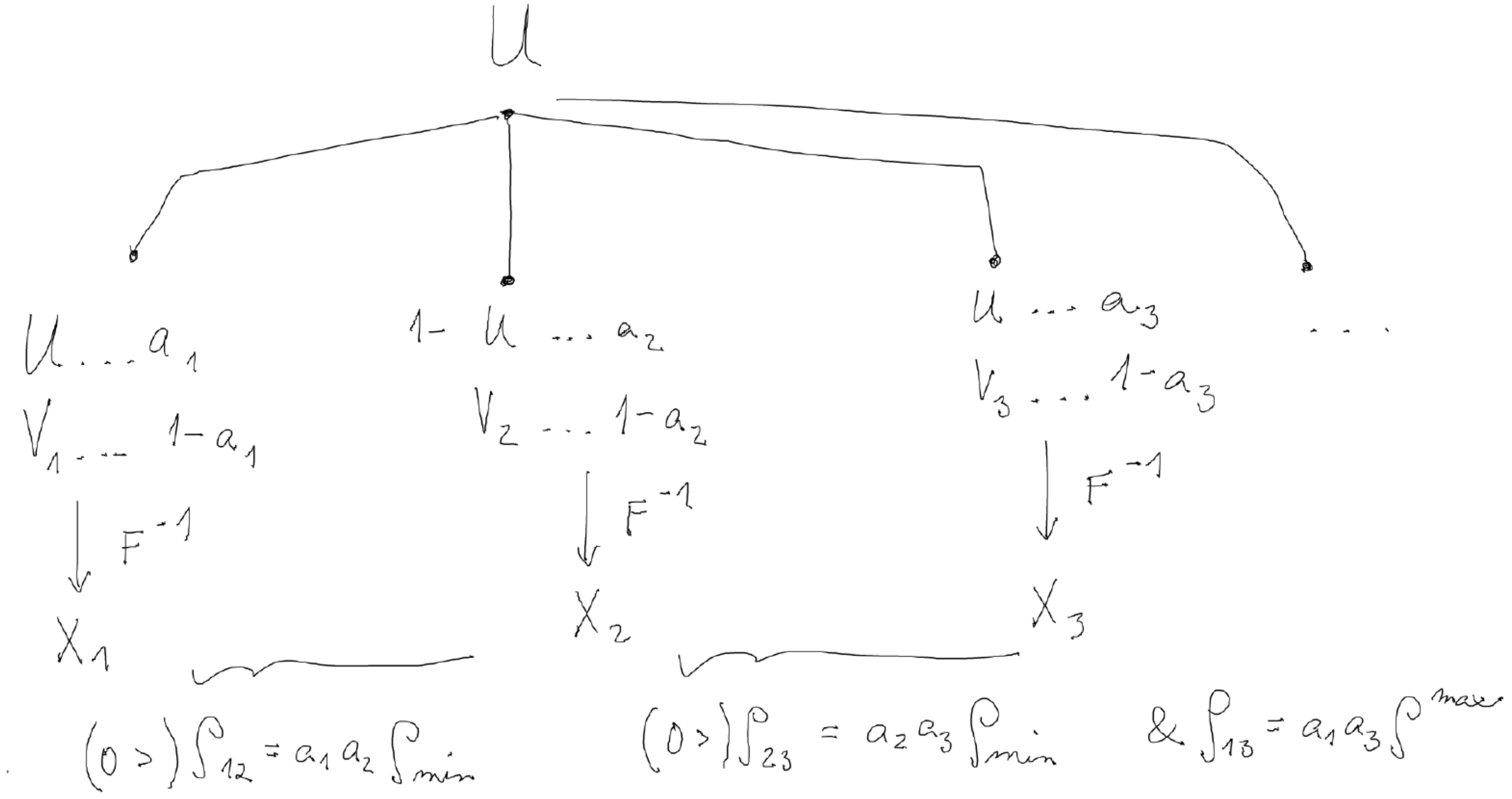
- valid correlation coefficients must be expressible as

$$\text{if } \rho_{ij} \geq 0: \quad \frac{|\rho_{ij}|}{\rho^{max}} = \alpha_i \alpha_j ,$$

$$\text{if } \rho_{ij} \leq 0: \quad \frac{|\rho_{ij}|}{\rho^{min}} = \alpha_i \alpha_j ,$$

$$\alpha_i , \alpha_j \in [-1,1]$$

Algorithm 3, Dukic, Marić (2013)



In terms of copulas...

C-lifting

Let $A(u_1,u_2), B(u_2,u_3)$ be bivariate copulas. And let $\{C_t\}$ be a family of bivariate copulas. Then

$$C(u_1,u_2,u_3)=\int_0^{u_2} C_t(\partial_2 A(u_1,t),\partial_1 B(t,u_3))\,dt$$

defines a trivariate copula with marginals $A(u_1,u_2), B(u_2,u_3)$ and $C(u_1,u_3)=\int_0^1 C_t(\partial_2 A(u_1,t),\partial_1 B(t,u_3))\,dt$

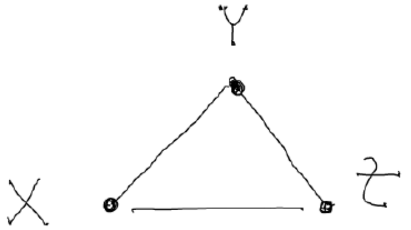
$\Updownarrow$

conditional distribution

(U_1, U_3) given $U_2 = t$

Durante et al (2007) – Remarks on two product-like constructions for copulas, Kybernetika 43
Kolesárová et al (2006) – Three copulas and compatibility, Proc. of IPMU 2006
Durante et al (2008) – Copulas: compatibility and Fréchet classes, J of Ineq and Appl, 161537

In terms of copulas...



$$X, Y: \ C_{XY} = aM_2 + (1-a)\Pi_2 \quad \longrightarrow \rho_{XY} = a\rho_{XY}^{max}$$

$$Y, Z: \ C_{YZ} = bM_2 + (1-b)\Pi_2 \quad \longrightarrow \rho_{YZ} = b\rho_{YZ}^{max}$$

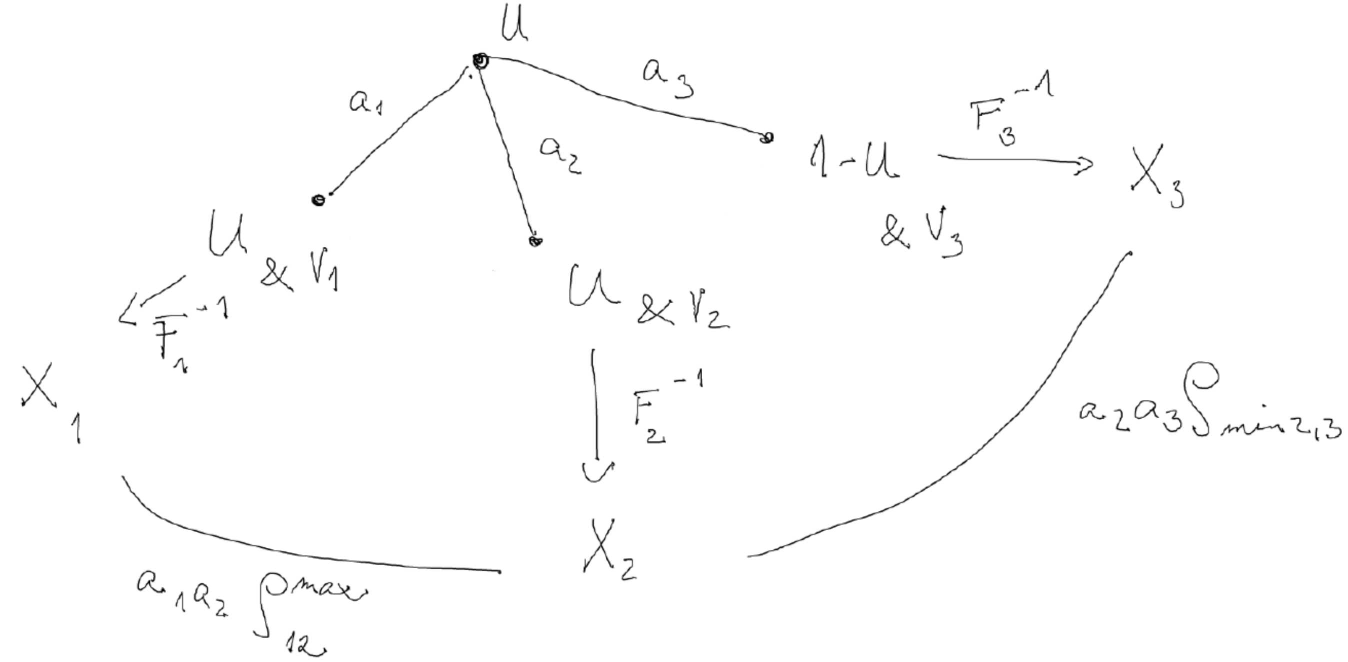
$$C_{XZ}(u,v) = C_{XY} *_{\Pi_2} C_{YZ} = \int_0^1 \Pi_2(\partial_2 C_{XY}(u,t), \partial_1 C_{YZ}(t,v))\,dt$$

$$= abM_2 + (1-ab)\Pi_2 \quad \longrightarrow \rho_{XZ} = ab\rho_{XZ}^{max}$$

$*_{\Pi_2} \longrightarrow C_{XZ|Y}$

$$(aM_2 + (1-a)\Pi_2) *_{\Pi_2} (bW_2 + (1-b)\Pi_2) = abW_2 + (1-ab)\Pi_2 \longrightarrow \text{corr} = ab\rho_{min}$$

Algorithm 3 once more



$X_1 X_3$: $a_1 a_3 \rho_{min,13}$

Natural questions...

- Can the bivariate copulas C_{XY}, C_{YZ}, C_{XZ} of the simple form (convex combinations of M_2 (W_2), Π_2) corresponding to a valid correlation structure be always joined by C-lifting into a trivariate copula?

This question was intensively studied by Hürlimann (2012) in

On Trivariate Copulas with Bivariate Linear Spearman Marginal Copulas,
Journal of Mathematics and System Science 2, 368-383

and answered with no.

Natural questions...

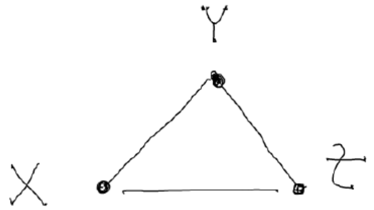
- Does there exist a copula C so that the C-lifting with it would yield C_{XZ} with the correct correlation?

???

$$C_{XY} *_{C} C_{YZ} = C_{XZ}$$

C_{XY}: $a M_2 + (1-a) \Pi_2$

C_{YZ}: $b W_2 + (1-b) \Pi_2$



$\rightarrow$ C_{XZ} is compatible with C_{XY}, C_{YZ}

$\Rightarrow$ Durante (2008) $\quad C_{XY} *_{W_2} C_{YZ} \leq C_{XZ} \leq C_{XY} *_{M_2} C_{YZ}$