

Szakdolgozat

Egymásba ágyazott arkhimédészi kopulák

Kasoly Benedek

Matematika alapképzés

Matematikai elemző specializáció

Témavezető: Dr. Zempléni András

Valószínűségelméleti és Statisztika Tanszék



Eötvös Loránd Tudományegyetem

Természettudományi Kar

2024

Tartalomjegyzék

1. Bevezetés	2
1.1. Kopulák	2
1.2. Összefüggési mérőszámok	4
2. Kétváltozós kopulák konstruálása	5
2.1. Az inverziós módszer	5
2.1.1. A Marshall-Olkin kétváltozós exponenciális eloszlás	6
2.1.2. Egyenletes eloszlás a körvonalon	6
2.2. Szinguláris kopulák	8
2.3. Kétváltozós arkhimédészi kopulák	9
3. Kopulák többdimenzióban	12
3.1. Elliptikus kopulák	12
3.2. Egymásba ágyazott arkhimédészi kopulák	12
3.2.1. A felcserélhető eset	13
3.2.2. A háromváltozós beágyazott eset	15
3.2.3. Páros négyváltozós eset	16
3.2.4. Az általános beágyazott eset	17
3.3. Vine kopulák	21
3.3.1. Szabályos vine fasorozatok	21
4. Arkhimédészi kopulastruktúra illesztése adathalmazra	23
4.1. Illeszkedésvizsgálat	23
5. Gyakorlati alkalmazás	24
5.1. Az adathalmaz elemzése	24
5.2. Különböző kopulák összehasonlítása	28
6. Konklúzió	31

1. Bevezetés

A modern korban rendkívül sok valuta, csereeszköz áll a társadalom rendelkezésére, melyek piaci értékét számos tényező befolyásolhatja. Például egy ország hivatalos fizetőeszközének árfolyamát erősen meghatározza az ország gazdasági, illetve politikai helyzete, vagy egy részvény értéke a tőzsdén erősen függ a részvényt kiadó társaság szituációjától. A közelmúltban megjelentek továbbá a kriptovaluták is, melyek piaci értékét kizárólag azok kereslete és kínálata határozza meg, így olyan tényezők is befolyásolhatják azt, mint egy közszereplő véleménye, vagy valamilyen szociális trend. Ezen valuták és értékpapírok mindegyike szabadon hozzáférhető, így kialakultak olyan portfóliók, ahol egy befektető ezekből többet is birtokol. Ahhoz, hogy ezek a befektetések minél nyereségesebbek lehessenek, szeretnénk megismerni a piac viselkedését kockázatelemzés céljából, például mi annak a valószínűsége, hogy az általunk vásárolt valuták, értékpapírok értéke egyszerre egy bizonyos százaléktérteket csökken? Erre adnak lehetőséget a többdimenziós statisztikai modellek, melyek egy viszonylag új ága az úgynevezett kopulák. Ezen modelleknek is nagyon sok családja létezik, attól függően, hogy milyen tulajdonságokkal rendelkezik az adathalmazunk, más-más kopulamodell fogja azt jól leírni. Egy ilyen család az egymásba ágyazott arkhimédészi kopulák családja, mely az egyik legfrissebb ezek közül, a dolgozatban ezzel fogunk majd részletesen foglalkozni, kitérve a többi fontosabb kopulacsaládra is. A modellezéshez az R programnyelvet fogjuk használni, a forráskódok a [12] linken elérhetőek.

1.1. Kopulák

A többdimenziós statisztika egyik legnehezebb feladata a többdimenziós eloszlás illesztése, ennek megvalósításában játszanak szerepet a kopulák. Egy többdimenziós eloszlás felbontható az egydimenziós egyenletes eloszlású peremeire és ezen peremeloszlások közötti összefüggési struktúrára, melyet kopulának nevezünk. Precízebben, a kopula egy olyan többdimenziós eloszlás eloszlásfüggvénye, melynek minden egydimenziós pereme egyenletes eloszlású a $[0, 1]$ intervallumon. Ezt az eloszlást a dolgozat további részében az $E(0, 1)$ jelöléssel jelöljük.

1. Definíció (Kopula). Ha a $C : [0, 1]^d \rightarrow [0, 1]$ d dimenziós eloszlásfüggvény peremei egyenletesek, d dimenziós kopulának nevezzük.

A definícióból nyilvánvalóan következnek a C kopula következő tulajdonságai:

Legyen $C(\mathbf{u}) = C(u_1, u_2, \dots, u_d)$, továbbá jelölje $U_1, U_2, \dots, U_d$ a megfelelő peremeloszlások valószínűségi változóit. Ekkor

- $C(\mathbf{u})$ nemcsökkenő függvény minden változójában.
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i, \forall i \in \{1, \dots, d\}$
- $\forall a_i < b_i, i \in \{1, \dots, d\}$ esetén a $P(a_1 \leq U_1 < b_1, a_2 \leq U_2 < b_2, \dots, a_d \leq U_d < b_d)$ valószínűségek nemnegatívak.

1. Tétel (Sklar tétele). Legyen F egy d dimenziós eloszlás együttes eloszlásfüggvénye, továbbá jelölje F egydimenziós peremeloszlásfüggvényeit $F_1, F_2, \dots, F_d$. Ekkor létezik olyan C kopula, amelyre

$$C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) = F(x_1, x_2, \dots, x_d), \forall x_i \in (\mathbb{R} \cup \{\infty, -\infty\}), i \in \{1, \dots, d\} \quad (1)$$

teljesül. Amennyiben a peremeloszlásfüggvények folytonosak, a C függvény egyértelmű.

Igaz továbbá ennek a tételnek a megfordítása is, azaz:

2. Tétel (Sklar tételének megfordítása). Ha C kopula és $F_1, F_2, \dots, F_d$ egydimenziós eloszlásfüggvények, akkor az (1) egyenlettel leírt F függvény egy olyan d dimenziós eloszlás eloszlásfüggvénye, mely peremeloszlásai az $F_1, F_2, \dots, F_d$ függvények.

Ahhoz hogy Sklar tételét bizonyítani tudjuk, szükségünk van arra, hogy meghatározzuk X valószínűségi változó inverzét olyan esetekben is, amikor X eloszlásfüggvénye nem szigorúan monoton.

2. Definíció (Kvantilisfüggvény). X valószínűségi változó kvantilisfüggvényének a következőt nevezzük:

$$F_X^{[-1]}(u) = \inf \{x : F_X(x) > u\}, 0 < u < 1 \quad (2)$$

Ha F_X szigorúan monoton növekvő, a kvantilisfüggvény megfelel a hagyományos értelemben vett inverznek: $F_X^{[-1]}(u) = F_X^{-1}(u)$

3. Tétel. *Tetszőleges X valószínűségi változó esetén az alábbiak teljesülnek:*

- Ha X abszolút folytonos eloszlású, a valószínűségi változót a saját eloszlásfüggvényébe helyettesítve egyenletes eloszlást kapunk $(0, 1)$ -en. Ezt a következőképpen jelöljük: $F_X(X) \sim E(0, 1)$
- Ha U egy $(0, 1)$ -en egyenletes valószínűségi változó, U -t az X valószínűségi változó kvantilisfüggvényébe helyettesítve az X valószínűségi változóval megegyező eloszlású valószínűségi változót kapunk, azaz $F_X^{[-1]}(U) = X$

Alkalmazzuk a fenti tételt egy valószínűségi vektorváltozóra. Legyen $X = (X_1, X_2, \dots, X_d)$ valószínűségi vektorváltozó. Az előző tétel alapján $F_{X_1}(X_1), F_{X_2}(X_2), \dots, F_{X_d}(X_d)$ mindegyike $E(0, 1)$ eloszlású, tehát a következő C_X függvény definiálható:

$$C_X(u_1, u_2, \dots, u_d) := P(F_{X_1}(X_1) < u_1, F_{X_2}(X_2) < u_2, \dots, F_{X_d}(X_d) < u_d)$$

A fentiek értelmében már nem is kell bizonyítani, hogy a C_X függvény kopula, a definíció pedig továbbírható a következő módon:

$$\begin{aligned} C_X(u_1, u_2, \dots, u_d) &= P(X_1 < F_{X_1}^{[-1]}(u_1), X_2 < F_{X_2}^{[-1]}(u_2), \dots, X_d < F_{X_d}^{[-1]}(u_d)) = \\ &= F_X(F_{X_1}^{[-1]}(u_1), F_{X_2}^{[-1]}(u_2), \dots, F_{X_d}^{[-1]}(u_d)) \end{aligned}$$

Legyen $x_j = F_{X_j}^{[-1]}(u_j)$, ebből adódik, hogy $F_{X_j}(x_j) = u_j$ és

$$C_X(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_d}(x_d)) = F_X(x_1, x_2, \dots, x_d)$$

ez pedig pont Sklar tételét bizonyítja □

4. Lemma. $X_1, X_2, \dots, X_d$ függetlenek akkor és csak akkor, ha $X_1, X_2, \dots, X_d$ kopulája $C(u_1, u_2, \dots, u_d) = u_1 \cdot u_2 \cdot \dots \cdot u_d$.

3. Definíció (Függetlenségi kopula). A fenti lemmában szereplő C függvényt függetlenségi kopulának nevezzük.

5. Tétel (Fréchet-Hoeffding korlátok). *Tetszőleges C kopula esetén:*

$$\max(u_1 + u_2 + \dots + u_d - (d - 1), 0) \leq C(u_1, u_2, \dots, u_d) \leq \min(u_1, u_2, \dots, u_d)$$

A $C(u_1, u_2, \dots, u_d) = \min(u_1, u_2, \dots, u_d)$ felső Fréchet-Hoeffding korlát önmagában is kopula és a tökéletes pozitív összefüggőséget írja le, $d = 2$ esetén pedig a $C(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ alsó korlát is kopula, mely a két valószínűségi változó közötti tökéletes negatív összefüggőséget határozza meg.

1.2. Összefüggési mérőszámok

Kopulák esetén az összefüggések számszerűsítésére úgynevezett rangkorrelációs mérőszámokat alkalmazunk. Erre azért van szükség, mert a rangkorrelációk közös tulajdonsága, hogy nem az eredeti értékekből, hanem a megfigyelések rangjaiból számolnak korrelációt, így az nem fognak függeni a peremeloszlásoktól. Ebben a fejezetben röviden foglalkozunk velük, mélyebb megértéshez Nelsen könyvének [3] 5. fejezetét javasoljuk.

4. Definíció (Rang). Legyenek $x_1, x_2, \dots, x_n$ megfigyelések. $x_i, i \in \{1, 2, \dots, n\}$ elem rangja k szám akkor, ha x_i a k -adik legkisebb elem ($x_k^* = x_i$).

A következő táblázatban kettő, a kopulaelméletben használt rangkorreláció definícióját, illetve mintából való becslésének módját találhatjuk meg. (X_1, Y_1) és (X_2, Y_2) ugyanolyan eloszlásúak mint (X, Y) , illetve függetlenek is tőle. $x_1, x_2, \dots, x_n$ és $y_1, y_2, \dots, y_m$ jelöli a tapasztalati mintát, $\bar{x}$ és $\bar{y}$ a minták átlagait, $r(x_i)$ és $r(y_i)$ az egyes elemek rangjait, $\overline{r(x)}$ és $\overline{r(y)}$ pedig a rangok átlagait. $R(X, Y)$ jelöli továbbá a szokásos Pearson-féle lineális korrelációt.

Korreláció	Definíció	Becslés mintából
Kendall-féle τ	$E(\text{sgn}[(X_1 - X_2)(Y_1 - Y_2)])$	$\frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)$
Spearman-féle ρ	$R(F_X(X), F_Y(Y))$	$\frac{\sum_{i=1}^n \sum_{j=1}^m (r(x_i) - \overline{r(x)})(r(y_i) - \overline{r(y)})}{\sqrt{\sum_{i=1}^n (r(x_i) - \overline{r(x)})^2} \sqrt{\sum_{j=1}^m (r(y_i) - \overline{r(y)})^2}}$

A fenti korrelációkra igazak a következő tulajdonságok:

- Értékeik a $[-1, 1]$ intervallumon vannak.
- A $+1$ a tökéletes pozitív, míg a -1 a tökéletes negatív, szigorúan monoton kapcsolatot jelöli.
- A 0 -hoz közeli érték gyenge kapcsolatot jelöl.
- Értékük nem függ a peremektől, kizárólag a kopulától.
- Nem érzékenyek a kiugró értékekre.

Mivel a Kendall-féle τ korreláció sokkal könnyebben számolható, mint a Spearman-féle ρ korreláció, így a továbbiakban ezt a mérőszámot fogjuk alkalmazni a rangkorrelációk vizsgálatára. A könnyebb érthetőség kedvéért, részletezzük ennek a kiszámítási módját. Ehhez vegyük az (x_i, y_i) és (x_j, y_j) x szerint rendezett párokat, ahol $i = 1, \dots, n-1$ és $j = i+1, \dots, n$. Legyen továbbá $x_1 \leq x_2 \leq \dots \leq x_n$. Minden pár esetén megállapítjuk, hogy:

- Ha $x_i < x_j$ és $y_i < y_j$, akkor konkordáns.
- Ha $x_i < x_j$ és $y_i > y_j$, akkor diszkordáns.
- Ha $x_i \neq x_j$ és $y_i = y_j$, akkor kötés van Y -ban.
- Ha $x_i = x_j$ és $y_i \neq y_j$, akkor kötés van X -ben.
- Ha $x_i = x_j$ és $y_i = y_j$, akkor kötés van X -ben és Y -ban.

Számoljuk össze, hogy a különböző párokból mennyi van:

- C : a konkordánsok száma
- D : a diszkordánsok száma
- T_X : Az X -beli kötések száma
- T_Y : Az Y -beli kötések száma
- T_{XY} : Az X -beli és Y -beli kötések száma

A Kendall-féle τ a konkordánsok és a diszkordánsok számát hasonlítja össze $\sqrt{(C + D + T_X) \cdot (C + D + T_Y)}$ -al normálva. Így kiszámolható a $\frac{C-D}{\sqrt{(C+D+T_X) \cdot (C+D+T_Y)}}$ módszerrel, ami megmagyarázza a fenti táblázatban található képletet is.

2. Kétváltozós kopulák konstruálása

Ebben a fejezetben meg szeretnénk mutatni néhány általános módszert, melyek segítségével kétváltozós kopulákat konstruálhatunk, Roger B. Nelsen könyvét [3] alapul véve.

2.1. Az inverziós módszer

Legyen H egy kétváltozós eloszlásfüggvény, melynek folytonos peremeloszlásai F és G . Ebből megkapható a következő kopula:

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v))$$

Ennek a kopulának a segítségével új kétváltozós eloszlásokat hozhatunk létre, tetszőleges F' és G' peremekkel, Sklar tételének (1) felhasználásával:

$$H'(x, y) = C(F'(x), G'(y))$$

Ugyanezt meg tudjuk valósítani úgynevezett túlélésfüggvényekkel is, melyeket a definíción kívül nem fogunk részletezni, ezekkel a Nelsen könyv [3] 2.6 fejezete foglalkozik részletesen.

5. Definíció (Túlélésfüggvény). Annak a valószínűsége, hogy egy objektum élettartama meghalad egy adott x időt, megadható annak $\bar{F}(x) = P[X > x] = 1 - F(x)$ túlélésfüggvényével, ahol F X eloszlásfüggvényét jelöli. Kétváltozós esetben ha X és Y véletlen valószínűségi változók, melyek együttes eloszlásfüggvénye H , az együttes túlélésfüggvényt $\bar{H}(x, y) = P[X > x, Y > y]$ adja meg, melynek egyváltozós peremei éppen $\bar{F}$ és $\bar{G}$ túlélésfüggvények.

Nelsen megmutatja, hogy $\hat{C} : \mathbf{I}^2 \rightarrow I$, $\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ kopulát definiálva a következőt kapjuk:

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y))$$

Ennek segítségével kifejezhetjük a fellebb található egyenletet túlélésfüggvények segítségével is, miszerint

$$\hat{C}(u, v) = \bar{H}(\bar{F}^{(-1)}(u), \bar{G}^{(-1)}(v))$$

ahol $\bar{F}^{(-1)}(t) = F^{-1}(1 - t)$. A fentiek segítségével megmutatjuk, hogyan találhatunk kopulákat a Marshall-Olkin kétváltozós exponenciális eloszlásrendszerre, illetve a körön értelmezett egyenletes eloszlásra.

2.1.1. A Marshall-Olkin kétváltozós exponenciális eloszlás

Az egyváltozós exponenciális eloszlás központi szerepet tölt be a statisztika terén, hiszen ez írja le a standard Poisson folyamat várakozási idejének eloszlását. A következőekben definiált kétváltozós exponenciális eloszlás hasonló szerepet játszik a kétdimenziós Poisson folyamatban. Képzeljünk egy kétkomponenses rendszert, például egy repülőgépet, melynek két hajtóműve van. Ezek a komponensek ki vannak téve valamilyen behatásoknak, melyek mindig végzetesek vagy egyik, vagy mindkettő komponens számára. A repülőgépünk esetén például egyik vagy másik hajtómű meghibásodhat, de a repülőgép akár hegynek is csapódhat, egyszerre megsemmisítve ezzel mindkettő hajtóművet. Jelölje X és Y a komponensek élettartalmát, illetve írjuk fel a $\bar{H}(x, y) = P[X > x, Y > y]$ túlélésfüggvényt, mely annak a valószínűségét adja meg, hogy az első komponens x , a második pedig y idő után is működésképes. A komponensekre ható lehetséges behatásokat három független Poisson folyamattal modellezhetjük, melyek pozitív paraméterei λ_1 , λ_2 és λ_{12} attól függően, hogy a behatás csak az első, csak a második, vagy mindkettő komponensre semmisíti meg. Az idők, amik ezen behatások előfordulását írják le, meghatározhatóak Z_1 , Z_2 és Z_{12} független exponenciális eloszlású valószínűségi változókkal, melyek paraméterei sorban λ_1 , λ_2 és λ_{12} . Ezek alapján $X = \min(Z_1, Z_{12})$ és $Y = \min(Z_2, Z_{12})$, tehát minden $x, y \geq 0$ esetén

$$\bar{H}(x, y) = P[Z_1 > x]P[Z_2 > y]P[Z_{12} > \max(x, y)] = \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)]$$

A perem túlélésfüggvények $\bar{F}(x) = \exp(-(\lambda_1 + \lambda_{12})x)$ és $\bar{G}(y) = \exp(-(\lambda_2 + \lambda_{12})y)$, tehát X és Y exponenciális eloszlású valószínűségi változók $\lambda_1 + \lambda_{12}$ illetve $\lambda_2 + \lambda_{12}$ paraméterekkel. Annak érdekében, hogy megtaláljuk a $\hat{C}(u, v)$ túlélési kopulát erre az eloszlásra, kifejezzük $\bar{H}(x, y)$ függvényt $\bar{F}(x)$ és $\bar{G}(y)$ szemszögéből, felhasználva, hogy $\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y))$. Ehhez kifejezzük $\max(x, y)$ -t, mint $x + y - \min(x, y)$, így

$$\begin{aligned}\bar{H}(x, y) &= \exp(-(\lambda_1 + \lambda_{12})x - (\lambda_2 + \lambda_{12})y + \lambda_{12} \min(x, y)) \\ &= \bar{F}(x)\bar{G}(y) \min(\exp(\lambda_{12}x), \exp(\lambda_{12}y))\end{aligned}$$

Legyen $u = \bar{F}(x)$ és $v = \bar{G}(y)$, továbbá legyen $\alpha = \frac{\lambda_{12}}{\lambda_1 + \lambda_{12}}$ és $\beta = \frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}$. Ekkor $\exp(\lambda_{12}x) = u^{-\alpha}$ és $\exp(\lambda_{12}y) = v^{-\beta}$, ezáltal a $\hat{C}$ túlélési kopula meghatározható

$$\hat{C}(u, v) = uv \min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta})$$

alapján. Mivel λ_1 , λ_2 és λ_{12} pozitívak, igaz, hogy $0 < \alpha, \beta < 1$. Ez alapján a túlélési kopuláit a Marshall-Olkin kétváltozós exponenciális eloszlásnak a következő kétparaméteres kopulacsalád adja meg, melyet Marshall-Olkin kopulacsaládnak nevezünk:

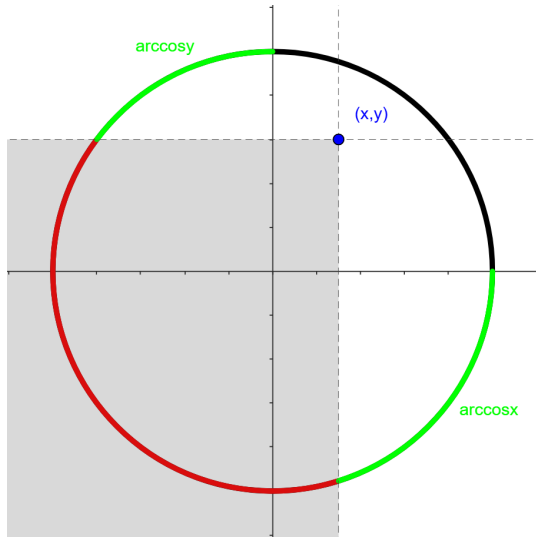
$$C_{\alpha, \beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta \end{cases}$$

2.1.2. Egyenletes eloszlás a körvonalon

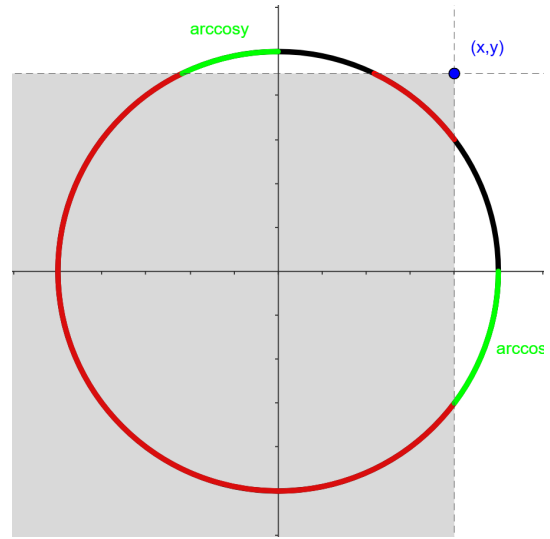
Jelölje (X, Y) egy véletlenszerűen választott pont koordinátáit az egységkörön. A kiválasztást úgy végezzük, hogy ha a pont polárkoordinátái $(1, \Theta)$, akkor a Θ véletlen változó egyenletesen oszlik el a $[0, 2\pi)$ intervallumon. Az X és Y együttes eloszlásfüggvénye $H(x, y)$ meghatározható, ha felhasználjuk, hogy $X = \cos \Theta$ és $Y = \sin \Theta$, viszont tanulságosabb, ha egy geometriai megközelítést alkalmazunk. Meg fogjuk mutatni, hogy a $[-1, 1]^2$ négyzeten belül található (x, y) pontokra H megadható a következő módon:

$$H(x, y) = \begin{cases} \frac{3}{4} - \frac{\arccos x + \arccos y}{2\pi}, & x^2 + y^2 \leq 1 \\ 1 - \frac{\arccos x + \arccos y}{\pi}, & x^2 + y^2 > 1, x, y \geq 0 \\ 1 - \frac{\arccos x}{\pi}, & x^2 + y^2 > 1, x < 0 \leq y \\ 1 - \frac{\arccos y}{\pi}, & x^2 + y^2 > 1, y < 0 \leq x \\ 0, & x^2 + y^2 > 1, x, y < 0 \end{cases}$$

ahol az $\arccos$ és $\arcsin$ függvények 0 és π közötti főértékeit használjuk. A négyzeten kívül H egyenlő lesz vagy nullával, vagy valamelyik peremével. Tegyük fel, hogy (x, y) az egységgörön, vagy azon belül van. Ekkor $2\pi H(x, y)$ az ívhossza a kör azon részének, melyet a lentebb található ábrán bal oldalt, a szürke részen belül pirossal jelöltünk. Felhasználva a kör szimmetriáját, illetve az íveket, melyek hosszát $\arccos x$ és $\arccos y$ adja meg, megkapjuk, hogy $2\pi H(x, y) = \frac{3\pi}{2} - \arccos x - \arccos y$. Ha (x, y) az egységgörön kívül helyezkedik el, de $[-1, 1]^2$ első negyedrésszében ahogy ezt a jobb oldali ábra mutatja, megkapjuk, hogy $2\pi H(x, y) = 2\pi - 2(\arccos x + \arccos y)$. Ha (x, y) valamelyik másik régióban van, $H(x, y)$ értékei hasonló módon határozhatóak meg.



(a) $H(x, y)$, ha (x, y) a körön belül helyezkedik el



(b) $H(x, y)$, ha (x, y) a körön kívül helyezkedik el

Mivel így már könnyen kifejezhetjük $H(x, y)$ együttes eloszlásfüggvényt $F(x)$ és $G(y)$ szemszögéből, így az az egy teendőnk maradt a kopula megtalálásához, hogy megtaláljuk az egységgör képét $x = F^{(-1)}(u)$, $y = G^{(-1)}(v)$ transzformáció alatt. Behelyettesítéssel azt kapjuk, hogy $\sin^2(\pi(u - \frac{1}{2})) = \cos^2(\pi(v - \frac{1}{2}))$, de hogyha α és β a $[-\frac{\pi}{2}, \frac{\pi}{2}]$ intervallumon helyezkednek el úgy, hogy $\sin^2 \alpha = \cos^2 \beta$, akkor $|\alpha| + |\beta| = \frac{\pi}{2}$. Így az egységgör képe $|u - \frac{1}{2}| + |v - \frac{1}{2}| = \frac{1}{2}$ lesz, a kopula pedig X és Y az egységgörön véletlenszerűen kiválasztott koordinátákra megadható az alábbi módon:

$$C(u, v) = \begin{cases} M(u, v), & |u - v| > \frac{1}{2} \\ W(u, v), & |u + v - 1| > \frac{1}{2} \\ \frac{u+v}{2} - \frac{1}{4}, & \text{egyébként} \end{cases}$$

ahol $M(u, v)$ a kopula felső, míg $W(u, v)$ pedig annak alsó Fréchet-Hoeffding korlátját jelöli, melyeket a 5. tételben határoztunk meg.

2.2. Szinguláris kopulák

Az előző fejezetben megmutattuk, hogyan használható fel Sklar tétele (1) az együttes eloszlásfüggvény invertálásához, ezáltal pedig kopulák megtalálásához. Most visszatérünk a kopula definíciójához és ezt felhasználva konstruálunk kopulákat, mégpedig anélkül, hogy hivatkoznánk eloszlásfüggvényekre, vagy véletlen változókra, 2-növekvő függvényeket építünk $\mathbf{I}^2$ -en egyletes peremekkel, geometriai jellegű információk felhasználásával.

6. Definíció (2-növekvő függvény). $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ 2-növekvő, ha bármely $x_1 \geq x_2$ és $y_1 \geq y_2$ esetén a következő teljesül:

$$f(x_1, y_1) + f(x_2, y_2) \leq f(x_1, y_2) + f(x_2, y_1)$$

Két példát fogunk mutatni olyan kopulákra, melyek tartója egy megadott halmazból származik. Az első példában ez a tartó egy töröttvonal, míg a másodikban körívek lesznek.

7. Definíció (Kopula tartója). A C kopula tartója $[0, 1]^2$ azon nyílt részhalmazai uniójának komplementere, ahol μ_C értéke nulla, ahol μ_C az a valószínűségi mérték, melyet C indukál.

Legyen $\theta \in [0, 1]$ és tegyük fel, hogy a θ -hoz tartozó valószínűségi tömegfüggvény egyenletes eloszlású a $(0, 0)$ és $(\theta, 1)$ pontokat összekötő szakaszon, továbbá az $1 - \theta$ valószínűségi tömegfüggvény egyenletes eloszlású a $(\theta, 1)$ és $(0, 1)$ pontokat összekötő szakaszon. Jegyezzük meg a két korlátozó esetet: amikor $\theta = 1$, ekkor a tartó $\mathbf{I}^2$ főátlója lesz, így a kapott kopula az M felső Fréchet-Hoeffding korlát, illetve amikor $\theta = 0$, ekkor a tartó $\mathbf{I}^2$ a másodlagos átló, így pedig a W alsó Fréchet-Hoeffding korlát jelöli ki a kopulát. Ezek alapján jelölje az adott töröttvonalakra koncentrálódó kopulát C_θ és legyen $C_1 = M$ és $C_0 = W$. Felhasználva, hogy C_θ tartója a két szakaszon fekszik, megtalálhatjuk $C_\theta(u, v)$ -t, ha kiszámoljuk a C_θ -beli területét a megfelelő téglalapoknak $\mathbf{I}^2$ -en. Mivel a pontok elhelyezkedése alapján a $\mathbf{I}^2$ három szegmensre bontható, így három esetet kell nézünk, attól függően, hogy az (u, v) pont hol helyezkedik el $\mathbf{I}^2$ -en belül.

Tegyük fel, hogy $u \leq \theta v$. Ekkor (u, v) a tartót kijelölő első szakasz feletti szegmensben helyezkedik el. Ekkor a C_θ -beli területe a $[0, u] \times [0, v]$ téglalapnak megegyezik a $[0, u] \times [0, 1]$ téglalap C_θ -beli területével és $V_{C_\theta}([0, u] \times [0, 1]) = u$ -ból következik, hogy $C_\theta(u, v) = u$.

Vizsgáljuk azt az esetet, amikor $u > \theta v$ és $u < 1 - (1 - \theta)v$, azaz (u, v) mindkét a tartót kijelölő szakasz alatti szegmensben helyezkedik el. Ekkor $C_\theta(u, v) = C_\theta(\theta v, v) = \theta v$, mivel a $[\theta v, u] \times [0, v]$ téglalap C_θ -beli területe nulla.

Utoljára nézzük meg, ha $u \geq 1 - (1 - \theta)v$, azaz (u, v) a tartót kijelölő második szakasz feletti szegmensben helyezkedik el. Ekkor megjegyezhetjük, hogy a C_θ -beli területe bármely olyan téglalapnak, mely egyik tartót kijelölő szakaszt sem metszi nulla kell hogy legyen. Ezáltal a $[u, 1] \times [v, 1]$ téglalap C_θ -beli területe nulla, amiből következik, hogy $C_\theta(u, v) = u + v - 1$. Ezen esetek összesítésével megkapható a $C_\theta(u, v)$ kopula:

$$C_\theta(u, v) = \begin{cases} u, & 0 \leq u \leq \theta v \leq \theta \\ \theta v, & 0 \leq \theta v < u < 1 - (1 - \theta)v \\ u + v - 1, & \theta \leq 1 - (1 - \theta)v \leq u \leq 1 \end{cases}$$

Egy másik hasonló eset, amikor a kopula tartóját két negyedkörív adja. A felső negyedkört a $u^2 + v^2 = 2u$, az alsót pedig a $u^2 + v^2 = 2v$ egyenlettel adjuk meg. Mivel a tartó szimmetrikus $\mathbf{I}^2$ főátlójára, így egy szimmetrikus kopulát fogunk konstruálni, azaz egy olyan C kopulát, amire $C(u, v) = C(v, u)$. Ahogy az előző esetben is, meg tudjuk mutatni, hogy ha (u, v) a felső tartó feletti szegmensben található, akkor $C(u, v) = u$, továbbá ha (u, v) az alsó tartó alatti szegmensben helyezkedik el, $C(u, v) = v$ lesz igaz. Így felírható, hogy $u^2 + v^2 > 2 \min(u, v)$ -re $C(u, v) = M(u, v)$.

Tegyük fel ezek után, hogy $u \leq v$ és $u^2 + v^2 \leq 2u$, azaz (u, v) a főátló felett, viszont a felső negyedkörív alatt található. Ekkor $V_C([u, v] \times [u, v]) = 0$, amiből következik, hogy $C(u, v) + C(v, u) =$

$C(u, u) + C(v, v)$, ami ekvivalens a $C(u, v) = \frac{1}{2}[\delta(u) + \delta(v)]$, ahol δ C átló szegmensét jelöli. Abban az esetben, amikor az (u, v) pont éppen a $u^+v^2 = 2u$ felső negyedköríven fekszik, a folytonosság miatt $u = C(u, v) = \frac{1}{2}[\delta(u) + \delta(v)]$, azaz $\delta(u) + \delta(v) = 2u = u^2 + v^2$. Ennek az egyenletnek a megoldásait $\delta(u) = u^2$ adja meg, amiből $C(u, v) = \min(u, v, \frac{u^2+v^2}{2})$ következik.

2.3. Kétváltozós arkhimédészi kopulák

Adathalmazok elemzésekor gyakran találkozhatunk olyan esetekkel, ahol a sejtett kopula struktúrája nem szimmetrikus, illetve nagyobb dimenziók esetén gyakran olyan kopulákat tudunk csak illeszteni az adatokra, melyek nagyon sok paramétert tartalmaznak, így rendkívül nehéz ezeket pontosan becsülni. Ezen esetek megoldására szolgálnak az arkhimédészi kopulák, melyek bevezetéséhez David Ziener 2021-es cikke [2] nyújt segítséget.

Kétváltozós esetben, ha X és Y folytonos valószínűségi változók, együttes eloszlásfüggvényük pedig H , akkor F és G peremeik akkor és csak akkor függetlenek, ha $H(x, y) = F(x)G(y)$ minden $x, y \in \mathbb{R}$ esetén. Van egy kopulacsalád, amely kielégít egy hasonló feltételt, mégpedig az úgynevezett Ali-Mikhail-Haq kopulák, melyeket a következő módon kaphatunk meg:

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1 - u)(1 - v)}$$

$\theta \in [-1, 1]$ -re. Ez a kopulacsalád úgy van megalkotva, hogy teljesüljön rá a következő:

$$\frac{1 - C_\theta(u, v)}{C_\theta(u, v)} = \frac{1 - u}{u} + \frac{1 - v}{v} + (1 - \theta) \cdot \frac{1 - u}{u} \cdot \frac{1 - v}{v}$$

Definiálva $\lambda(t) = 1 + \frac{(1-\theta)(1-t)}{t}$ ez átírható

$$\lambda(C_\theta(u, v)) = \lambda(u)\lambda(v)$$

alakba, sőt ha bevezetjük a $\psi(t) = -\log \lambda(t)$ jelölést, a következőt kapjuk:

$$\psi(C_\theta(u, v)) = \psi(u) + \psi(v)$$

Olyan kopulákat szeretnénk találni, melyek teljesítik a megkapott

$$\psi(C(u, v)) = \psi(u) + \psi(v)$$

feltételt, valamilyen ψ függvényre. Ahhoz, hogy a fenti összefüggést meg tudjuk oldani C -re, először definiálnunk kell ψ megfelelő $\psi^{[-1]}$ inverzét. Ha ez sikerül, kaptunk egy kopulát, mely kielégíti a

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v)) \quad (3)$$

összefüggést.

8. Definíció. Legyen $\psi : \mathbf{I} \rightarrow [0, \infty]$ folytonos, szigorúan csökkenő és $\psi(1) = 0$. Ekkor ψ pszeudo-inverze a következő:

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t), & 0 \leq t \leq \psi(0) \\ 0, & \psi(0) \leq t \leq \infty \end{cases}$$

Legyen $\psi^{[-1]}(t)$ a fentiek szerint definiálva. Ekkor

- $\psi^{[-1]}$ folytonos, nem növekvő a $[0, \infty]$ és szigorúan csökkenő a $[0, \psi(0)]$ intervallumon.
- $\forall t \in \mathbf{I} : \psi^{[-1]}(\psi(t)) = t$
- $\forall t \in [0, \infty] : \psi(\psi^{[-1]}(t)) = \min(t, \psi(0))$
- $\psi(0) = \infty \Rightarrow \psi^{[-1]} = \psi^{-1}$

Ahhoz, hogy bebizonyítsuk, hogy az (3) egyenletben definiált C függvény tényleg kopula, szükségünk van a következő két lemmára.

6. Lemma. Legyen $\psi : \mathbf{I} \rightarrow [0, \infty]$ folytonos és szigorúan csökkenő mégpedig úgy, hogy $\psi(1) = 0$, továbbá legyen $\psi^{[-1]}$ a pszeudo-inverz. Legyen $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ a következő módon definiálva:

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v))$$

Ebben az esetben C kielégíti a kopula feltételeit, azaz minden $(u, v) \in \mathbf{I}$ esetén

$$C(u, 0) = C(0, v) = 0$$

és

$$C(u, 1) = u, C(1, v) = v$$

7. Lemma. Legyen $\psi : \mathbf{I} \rightarrow [0, \infty]$ folytonos és szigorúan csökkenő mégpedig úgy, hogy $\psi(1) = 0$, továbbá legyen $\psi^{[-1]}$ a pszeudo-inverz. Legyen $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ a következő módon definiálva:

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v))$$

Ekkor C 2-növekvő (6. definíció) akkor és csak akkor, ha minden $v \in \mathbf{I}$ -re teljesül, hogy

$$u_1 \leq u_2 \Rightarrow C(u_2, v) - C(u_1, v) \leq u_2 - u_1$$

8. Tétel. Legyen $\psi : \mathbf{I} \rightarrow [0, \infty]$ folytonos és szigorúan csökkenő mégpedig úgy, hogy $\psi(1) = 0$, továbbá legyen $\psi^{[-1]}$ a pszeudo-inverz. Legyen $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ az eddigiekhez hasonlóan definiálva. Ekkor C egy kopula akkor és csak akkor, ha a ψ függvény konvex.

Bizonyítás. Az 6. és 7. lemma miatt elég megmutatnunk azt, hogy

$$\forall v \in \mathbf{I} : (u_1 \leq u_2 \Rightarrow C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \Leftrightarrow \psi \text{ konvex}).$$

Mivel ψ szigorúan csökkenő, ψ konvex akkor és csak akkor, ha $\psi^{[-1]}$ is konvex, azaz

$$\forall v \in \mathbf{I} : (u_1 \leq u_2 \Rightarrow C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \Leftrightarrow \psi^{[-1]} \text{ konvex}).$$

A bal oldali feltétel átírható

$$u_1 + \psi^{[-1]}(\psi(u_2) + \psi(v)) \leq u_2 + \psi^{[-1]}(\psi(u_1) + \psi(v))$$

alakba, ha $u_1 \leq u_2$. Definiálva $a := \psi(u_1)$, $b := \psi(u_2)$ és $c := \psi(v)$ függvényeket

$$\psi^{[-1]}(a) + \psi^{[-1]}(b + c) \leq \psi^{[-1]}(b) + \psi^{[-1]}(a + c). \quad (4)$$

" $\Rightarrow$ ": Feltételezzük, hogy (4) teljesül. Minden $s, t \in [0, \infty]$, $s < t$ -re legyen $a = \frac{s+t}{2}$, $b = s$, $c = \frac{t-s}{2}$. Ekkor (4) a következőt adja:

$$\psi^{[-1]} \left(\frac{s+t}{2} \right) + \psi^{[-1]} \left(\frac{t+s}{2} \right) \leq \psi^{[-1]}(s) + \psi^{[-1]}(t)$$

amely ekvivalens az alábbival:

$$\psi^{[-1]} \left(\frac{s+t}{2} \right) \leq \frac{\psi^{[-1]}(s) + \psi^{[-1]}(t)}{2}$$

Ez pont azt jelenti, hogy $\psi^{[-1]}$ belső pontjaira konvex és mivel $\psi^{[-1]}$ folytonos függvényként Lebesgue mérhető, így következik hogy konvex is.

" $\Leftarrow$ ": Legyen $\psi^{[-1]}$ egy konvex függvény, legyen továbbá $a, b, c \in \mathbf{I}$ úgy, hogy $a \geq b$ és $c \geq 0$. Definiáljuk a $\gamma := \frac{a-b}{a-b+c} \in [0, 1]$ függvényt. Ezzel kifejezve $a = (1 - \gamma)b + \gamma(a + c)$ és $b + c = \gamma b + (1 - \gamma)(a + c)$. $\psi^{[-1]}$ konvexitását felhasználva megkapjuk

$$\psi^{[-1]}(a) \leq (1 - \gamma)\psi^{[-1]}(b) + \gamma\psi^{[-1]}(a + c)$$

és

$$\psi^{[-1]}(b + c) \leq \gamma\psi^{[-1]}(b) + (1 - \gamma)\psi^{[-1]}(a + c)$$

egyenlőtlenségeket, melyeket összeadva az

$$\psi^{[-1]}(a) + \psi^{[-1]}(b + c) \leq \psi^{[-1]}(b) + \psi^{[-1]}(a + c)$$

egyenlőtlenséget kapjuk, mely pont az (4)-el egyezik meg. □

Ez alapján már definiálhatjuk az arkhimédészi kopulákat a következőképpen:

9. Definíció (Arkhimédészi kopula). Legyen $\psi : \mathbf{I} \rightarrow [0, \infty]$ folytonos és szigorúan csökkenő mégpedig úgy, hogy $\psi(1) = 0$, továbbá legyen $\psi^{[-1]}$ a pseudo-inverz. Legyen $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ a következő módon definiálva:

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v))$$

Ekkor az előző tétel szerint C egy kopula, melyet arkhimédészi kopulának nevezünk, ψ függvény pedig a kopula generátorfüggvénye.

A következő táblázat kettő, a gyakorlati felhasználásban lényeges arkhimédészi kopulacsalád tulajdonságait foglalja össze.

Kopulacsalád neve	Kopulageneráló függvény	Kendall-féle τ
Gumbel	$-\ln(u)^\theta, \theta \geq 1$	$1 - \frac{1}{\theta}$
Clayton	$u^{-\theta} - 1, \theta > 0$	$\frac{\theta}{\theta+2}$

3. Kopulák többdimenzióban

3.1. Elliptikus kopulák

Az elliptikus kopulák osztálya olyan többváltozós eloszlásokat ad a kezünkbe, melyek számos tulajdonságát osztják a többváltozós normális eloszlásnak, mely maga is ebbe az osztályba tartozik, ahogy például a többváltozós Student-féle t eloszlás is. Az elliptikus kopulák ezekből az eloszlásokból származtathatóak Sklar tétele (1) segítségével. Ezeknek a kopuláknak fontos jellemzői, hogy szimmetrikusak, ezért aszimmetrikus piaci jelenségek modellezésére nem alkalmasak, például egyes pénzügyei adatok esetén, ahol jobban összefüggnek a nagy veszteségek, mint a nagy nyereségek. Komoly hátrányuk továbbá, hogy magasabb dimenziókban nagyon sok paraméterrel rendelkeznek, melyek becslése kifejezetten nehéz feladat. Ezen kopulák leírásához P. Embrechts, F. Lindskog és A. McNeil cikke [4] fog segítséget nyújtani.

10. Definíció (Gauss kopula). Legyen $X \sim N_d(0, R)$, ahol R a korrelációs mátrix, továbbá jelölje X együttes eloszlásfüggvényét Φ_R^n . Ekkor a Gauss-kopula a következő:

$$C_R^{\text{Ga}}(u_1, u_2, \dots, u_d) = \Phi_R^n(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n))$$

ahol Φ^{-1} az egyváltozós standard normális eloszlás eloszlásfüggvényének inverzét jelöli.

11. Definíció (t_ν eloszlás). Ha X valószínűségi változó sztochasztikus reprezentációja

$$X =_d \mu + \frac{\sqrt{\nu}}{\sqrt{S}} Z$$

ahol $\mu \in \mathbb{R}^n$, $S \sim \chi_\nu^2$ és $Z \sim \mathcal{N}_n(0, \Sigma)$ függetlenek, akkor X n változós t_ν eloszlású, melynek várható értéke μ ($\nu > 1$ esetén) és kovariancia-mátrixa $\frac{\nu}{\nu-2}\Sigma$ ($\nu > 2$ esetén). Ha $\nu \leq 2$, a kovariancia nem véges.

12. Definíció (t -kopula). A fent említett X valószínűségi változó által meghatározott kopula felírható a következőképpen:

$$C_{\nu, R}^t(u_1, u_2, \dots, u_n) = t_{\nu, R}^n(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \dots, t_\nu^{-1}(u_n))$$

ahol $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$, $i, j \in \{1, \dots, n\}$, továbbá $t_{\nu, R}^n$ jelöli $\frac{\sqrt{\nu}Y}{\sqrt{S}}$ eloszlásfüggvényét, ahol $S \sim \chi_\nu^2$ és $Y \sim \mathcal{N}_n(0, R)$ függetlenek. t_ν jelöli $t_{\nu, R}^n$ peremeit, azaz $\frac{\sqrt{\nu}Y_1}{\sqrt{S}}$ eloszlásfüggvényét.

3.2. Egymásba ágyazott arkhimédészi kopulák

Az előző fejezetben megismertedtünk a kétváltozós arkhimédészi kopulákkal, terjesszük ki ezt magasabb dimenziókra is, melyhez segítséget nyújt McNeil 2008-as cikke [1]. A kiterjesztés egy klasszikus módja az úgynevezett felcserélhető konstrukció:

$$C(u_1, u_2, \dots, u_d) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \dots + \psi^{-1}(u_d)) \quad (5)$$

Bizonyítható, hogy ez tényleg egy kopulát ad bármely d dimenzió esetén akkor és csak akkor, ha ψ teljesen monoton függvény, azaz igaz rá, hogy

$$(-1)^k \frac{d^k}{dt^k} \psi(t) \geq 0 \quad (6)$$

A teljesen monoton arkhimédészi generátorokat Laplace transzformált arkhimédészi generátoroknak, röviden LP-arkhimédészinek nevezzük. Ez azért van, mert a teljesen monoton függvények osztálya

$\psi[0, \infty]$ -n egybeesik az eloszlásfüggvények Laplace(–Stieltjes) transzformáltjainak osztályával a pozitív valós számokon. Ha az arkhimédészi generátorokat Laplace transzformáltakkal reprezentáljuk, hasznos módszereket kaphatunk a magasabb dimenziós arkhimédészi kopulák szimulálására is.

A fentebb említett (5) eloszlásfüggvénnyel rendelkező véletlen vektoroknak erősen specifikus a függőségi struktúrája: bármely részhalmaza az összes ugyanolyan dimenziós részhalmazzal azonos eloszlást alkot, a páronkénti rangkorrelációik nem negatívak. Szeretnénk olyan többváltozós arkhimédészi kopulákat alkotni, amik rugalmasabbak, megengednek különböző fokú pozitív összefüggőséget a különböző kétváltozós peremeloszlások között. Ilyen modelleket tudunk készíteni, ha egymásba ágyazunk generátorfüggvényeket.

Példának vegyünk egy három dimenziós modellt, amit két LT-arkhimédészi generátorfüggvény ψ_1 és ψ_2 segítségével készítünk, a következő módon:

$$C(u_1, u_2, u_3) = \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1} \circ \psi_2(\psi_1^{-1}(u_2) + \psi_2^{-1}(u_3))). \quad (7)$$

Amellett a feltétel mellett, hogy $\psi_1^{-1} \circ \psi_2$ első deriváltja teljesen monoton, a fenti konstrukció tényleg eloszlásfüggvényt fog adni. Ha a kopulát Laplace transzformációk használatával eloszlásfüggvények keverékeként reprezentáljuk, relatíve egyszerűen tudunk belőle mintát venni, legalábbis egyes generátorfüggvények esetén. A továbbiakban kifejtjük, hogy ez hogyan hajtható végre, illetve megvizsgálunk néhány magasabb dimenziós esetet is, többek között a következő négydimenziós konstrukciót:

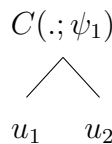
$$C(u_1, u_2, u_3, u_4) = \psi_1(\psi_1^{-1} \circ \psi_2(\psi_2^{-1}(u_1) + \psi_2^{-1}(u_2)) + \psi_1^{-1} \circ \psi_3(\psi_3^{-1}(u_3) + \psi_3^{-1}(u_4))) \quad (8)$$

illetve az alábbi rekurzív d -dimenziós esetet, ahol $d \geq 2$:

$$C_d(u_1, u_2, \dots, u_{d+1}; \psi_1, \psi_2, \dots, \psi_d) = \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1}(C_{d-1}(u_2, \dots, u_{d+1}; \psi_2, \dots, \psi_d))) \quad (9)$$

ahol $C_1(u_1, u_2; \psi)$ a kétváltozós kopulát és $C_2(u_1, u_2, u_3; \psi_1, \psi_2)$ a már említett 3 változós konstrukciót jelöli. (7)

Ahhoz, hogy könnyebben értelmezni tudjuk ezeket az egymásba ágyazott struktúrákat, ábrázolhatjuk őket egy gyökeres fagraféként. Legyenek a valószínűségi változóink $u_1, u_2, \dots$, az ezek között valamilyen módon elhelyezkedő arkhimédészi kopulastruktúrák pedig $C(\cdot; \psi_1), C(\cdot; \psi_2), \dots$, mely kopulák ugyanazon kopulacsaládból származnak, generátorfüggvényeik pedig sorban $\psi_1, \psi_2, \dots$. A fát úgy építjük fel, hogy minden olyan csúcs, melyeknek gyerekei vannak egy kopulát, amelyeknek pedig nincsenek, valószínűségi változót jelöl, a kopulák pedig a gyerekeik közötti összefüggést írják le. Egy általános kétdimenziós kopula struktúrája például a következő gráffal ábrázolható:



3.2.1. A felcserélhető eset

Legyen ψ egy monoton függvény a $[0, \infty]$ tartományon, ahol $\psi(0) = 1$ és $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$ és legyen G a $[0, \infty]$ tartományon értelmezett eloszlásfüggvény, amelynek Laplace-Stieltjes transzformáltját ψ adja meg, azaz

$$\psi(t) = \int_0^\infty e^{-tx} dG(x), \quad t \geq 0$$

Marshall és Olkin [6] megmutatta, hogyan tudunk generálni egy random $(U_1, \dots, U_d)$ vektort, melynek eloszlásfüggvényét a $C = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \dots + \psi^{-1}(u_d))$ kopula adja meg. Legyen $F(u) :=$

$\exp(-\psi^{-1}(u))$. Vegyük észre, hogy $F^\nu(u)$ egy egyváltozós eloszlásfüggvény $[0, 1]$ -en bármely $\nu > 0$ -ra. Ha V egy véletlen változó G eloszlásfüggvénnyel, és $(U_1, \dots, U_d)$ feltételesen függetlenek adott V esetén, melynek feltételes eloszlásfüggvénye $P(U_i \leq u|V = \nu) = F^\nu(u)$, akkor $(U_1, \dots, U_d)$ eloszlásfüggvénye a kívánt eloszlás lesz.

$$\begin{aligned} P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d) &= \int_0^\infty P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d | V = \nu) dG\nu \\ &= \int_0^\infty \prod_{i=1}^d F^\nu(u_i) dG\nu \\ &= \int_0^\infty e^{-\nu(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \dots + \psi^{-1}(u_d))} dG\nu \\ &= \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \dots + \psi^{-1}(u_d)) \end{aligned}$$

A fentiek alapján leírható a következő algoritmus, melynek segítségével tudunk mintát venni a kopulából:

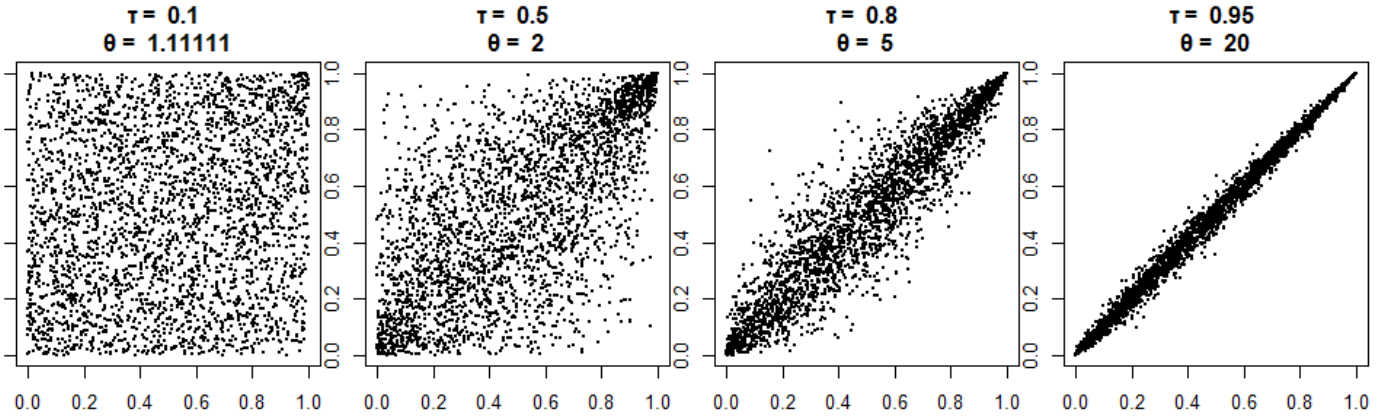
Algoritmus 1: Felcserélhető eset

Generáljunk egy V változót G eloszlásfüggvénnyel, melynek Laplace-transzformáltja ψ .

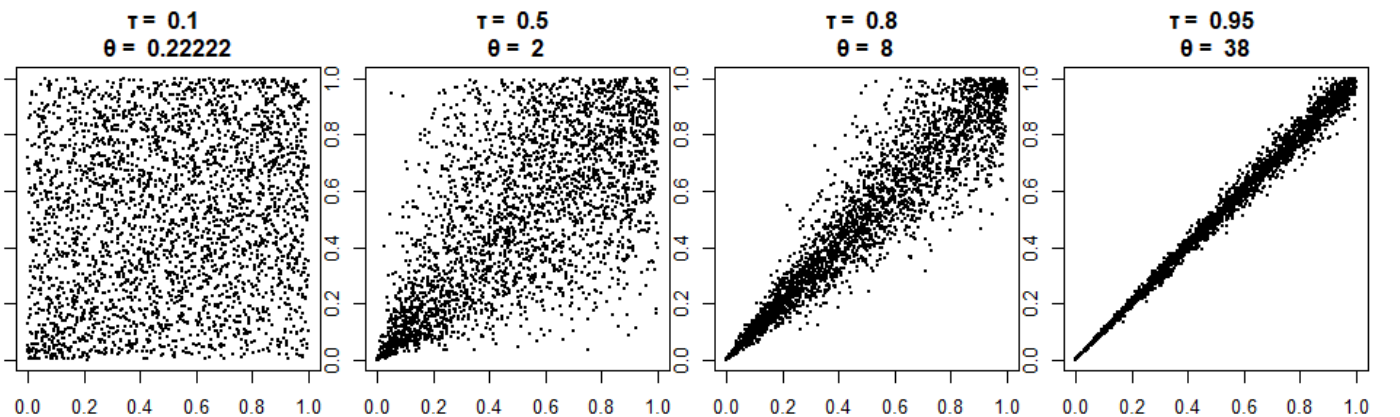
Generáljunk egymástól független egyenletes eloszlású $X_1, \dots, X_d$ változókat.

Térjünk vissza $(U_1, \dots, U_d) = (\psi(\frac{-\ln X_1}{V}), \dots, \psi(\frac{-\ln X_d}{V}))$ értékekkel.

A következő ábrák a fenti algoritmus által lettek generálva, kétdimenziós Gumbel illetve Clayton kopulákból.



2. ábra. Gumbel kopula



3. ábra. Clayton kopula

A módszer gyakorlati alkalmazásához vagy egy adott G eloszlásfüggvénnyel kell kezdenünk, amit mintavételezhetünk és amelyre kiszámíthatjuk a Laplace-transzformációt, vagy egy teljesen monoton ψ generátorral, amelyhez megkeressük azt a G eloszlásfüggvényt, melyre ψ a Laplace transzformált. Ebben a dolgozatban a második megközelítést fogjuk alkalmazni. Sajnos nem ismert általános módszer eloszlások mintavételére adott Laplace-transzformációkkal, így kénytelenek vagyunk eseti alapon tekinteni az arkhimédészi kopulákat.

A bevezetőben már felírtunk egy táblázatot, mely a két gyakorlatban leginkább használatos arkhimédészi kopulacsaládot, a Gumbel, illetve a Clayton kopulákat foglalta össze. A dolgozat további részében ezekkel a kopulacsaládokkal fogunk csak foglalkozni, a következő táblázat pedig olyan tulajdonságaikat foglalja össze, melyeket a későbbiekben fel fogunk használni.

Kopulacsalád neve	Gumbel	Clayton
$\psi(t, \theta)$	$\exp(-t^{\frac{1}{\theta}})$	$(1+t)^{-\frac{1}{\theta}}$
Paraméterter	$\theta \geq 1$	$\theta > 0$
$\psi^{-1}(t, \theta)$	$-(\ln u)^{\theta}$	$u^{-\theta} - 1$
V eloszlásfüggvénye	$\text{St}(\frac{1}{\theta}, 1, (\cos \frac{\pi}{2\theta})^{\theta}, 0)$	$\text{Ga}(\frac{1}{\theta})$
$\psi_2^{(1)}(t, \nu, \alpha)$	$\exp(-\nu t^{\frac{1}{\alpha}})$	$\exp(\nu - \nu(1+t)^{\frac{1}{\alpha}})$
Eloszlásfüggvény	$\text{St}(\frac{1}{\alpha}, 1, (\nu \cos \frac{\pi}{2\alpha})^{\alpha}, 0)$	tilted stable

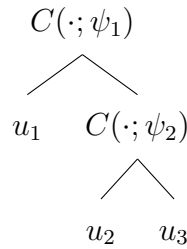
A táblázatban $\text{St}(\alpha, \beta, \gamma, \delta)$ jelöli a stabil eloszlást $\alpha, \beta, \gamma, \delta$ paraméterekkel, $\text{Ga}(\kappa)$ pedig a gamma eloszlást κ paraméterrel.

3.2.2. A háromváltozós beágyazott eset

Tegyük fel hogy ψ_1 és ψ_2 LT-arkhimédészi és legyen G_1 az az eloszlásfüggvény, melynek Laplace transzformáltja ψ_1 . Az előző fejezetben említett 3 változós konstrukció (7) így felírható a következő módon:

$$C(u_1, u_2, u_3) = \int_0^\infty e^{-\nu \psi_1^{-1}(u_1)} e^{-\nu \psi_1^{-1} \circ \psi_2(\psi_2^{-1}(u_2) + \psi_2^{-1}(u_3))} dG_1(\nu)$$

A struktúrát ábrázolhatjuk gráfként is, az előző fejezetben említett módszerrel:



Amennyiben sikerül bebizonyítani, hogy $\psi_1^{-1} \circ \psi_2$ deriváltja teljesen monoton, kimondhatjuk, hogy a fenti konstrukció értelmezhető eloszlásfüggvények keverékeként.

Legyen $F_i(u) := \exp(-\psi_i^{-1}(u))$, $i = 1, 2$. Ebben az esetben az összes $F_i^\nu(u)$ függvény létező egyváltozós eloszlásfüggvény, bármely $\nu > 0$ esetén. Az integrandus első tényezője $F_1^\nu(u_1)$, a második pedig felírható az alábbi alakban:

$$\begin{aligned}
H(u_2, u_3, \nu) &:= F_1^\nu(\psi_2(\psi_2^{-1}(u_2) + \psi_2^{-1}(u_3))) \\
&= \psi_2^{(1)}(\psi_2^{(1)-1}(F_1^\nu(u_2); \nu) + \psi_2^{(1)-1}(F_1^\nu(u_3); \nu); \nu)
\end{aligned}$$

ahol $\psi_2^{(1)}(\cdot; \nu) = F_1^\nu(\psi_2(\cdot)) = \exp(-\nu \psi_1^{-1} \circ \psi_2(\cdot))$. Mivel feltételeztük, hogy $\psi_1^{-1} \circ \psi_2$ deriváltja teljesen monoton, az előbb említett $\psi_2^{(1)}(\cdot; \nu)$ függvény egy LT-arkhimédészi kopulagenerátor, tehát a fentebb definiált H függvény kétváltozós eloszlásfüggvény $[0, 1]^2$ -en, peremeloszlásai pedig $F_1^\nu(u)$. Belátható, hogy ha f függvény teljesen monoton és g egy pozitív függvény, teljesen monoton deriválttal, akkor $f \circ g$ is teljesen monoton. Ez az állítás a mi esetünkben is igaz, ha $f(t) = \exp(-\nu t)$ és $g = \psi_1^{-1} \circ \psi_2$ helyettesítéseket alkalmazunk. Ez alapján az egyenlet átírható a következő módon:

$$C_2(u_1, u_2, u_3; \psi_1, \psi_2) = \int_0^\infty F_1^\nu(u_1) C_1(F_1^\nu(u_2), F_1^\nu(u_3); \psi_2^{(1)}(\cdot; \nu)) dG_1(\nu)$$

Ez reprezentálja a kopulát mint eloszlásfüggvények keveréke, ezáltal mintavételre alkalmassá teszi.

Algoritmus 2: Háromváltozós beágyazott eset

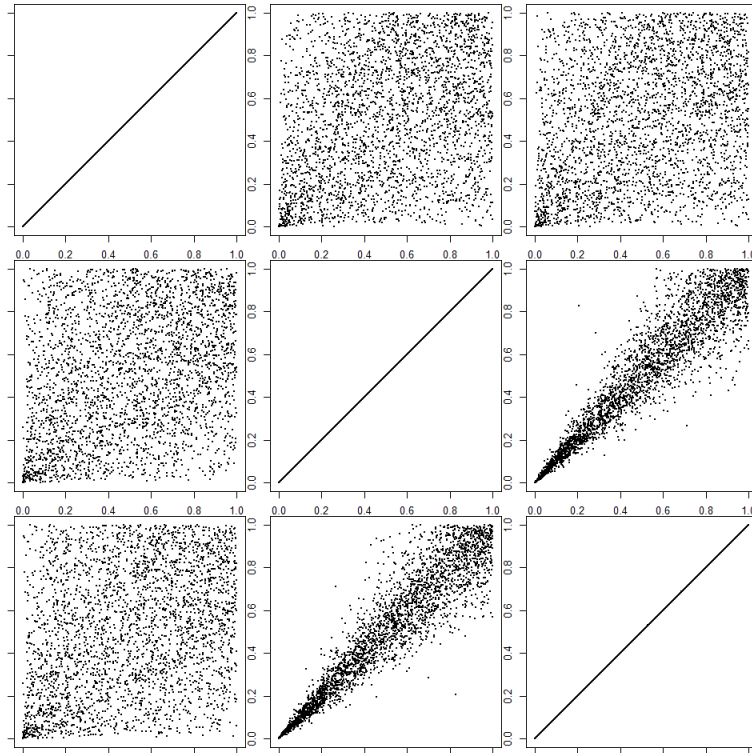
Generáljunk egy V változót G_1 eloszlásfüggvénnyel, melynek Laplace-transzformáltja ψ_1 .

Generáljunk egy standard egyenletes eloszlású X_1 változót.

Generáljunk (X_2, X_3) változókat a kétváltozós arkhimédészi kopulából az 1. algoritmus segítségével, generátorfüggvényként $\psi_2^{(1)}(\cdot, V) = \exp(-V \cdot \psi_1^{-1} \circ \psi_2(\cdot))$ -t használva.

Térjünk vissza (U_1, U_2, U_3) értékekkel, ahol $U_i = \psi_1(\frac{-\ln X_i}{V})$, $i = 1, 2, 3$

A következő ábrát a fenti algoritmussal generáltuk, Clayton kopulákból.



4. ábra. Háromváltozós egymásba ágyazott arkhimédészi kopula, $\theta_1 = 0.5, \theta_2 = 8$

3.2.3. Páros négyváltozós eset

A bevezetőben említettünk egy négyváltozós konstrukciót (8), melyet a dolgozat későbbi fejezeteiben többször is használni fogunk. A mintavételezés algoritmusának bizonyítása hasonló a háromváltozós esethez, így ezt nem részletezzük. Legyen a kopulánk a bevezetőben leírtak szerint definiálva (8), továbbá feltételezzük, hogy ψ_1 , ψ_2 és ψ_3 LT-arkhimédésziek, illetve $\psi_1^{-1} \circ \psi_2$ és $\psi_2^{-1} \circ \psi_3$ teljesen monoton deriválttal rendelkező függvények. Ekkor a következő algoritmus alkalmas lesz a kopulából való mintavételezésre:

Algoritmus 3: Páros négyváltozós eset

Generáljunk egy V_1 változót G_1 eloszlásfüggvénnyel, melynek Laplace-transzformáltja ψ_1 .

Generáljunk egy V_2 változót $G_2(\nu; V_1)$ eloszlásfüggvénnyel, melynek Laplace-transzformáltja

$$\psi_2^{(1)}(\cdot; V_1) = \exp(-V_1 \psi_1^{-1} \circ \psi_2(\cdot)).$$

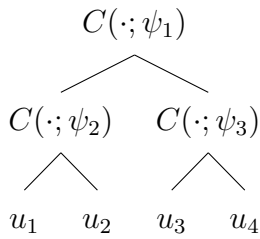
Generáljunk egy V_3 változót $G_3(\nu; V_1)$ eloszlásfüggvénnyel, melynek Laplace-transzformáltja

$$\psi_3^{(1)}(\cdot; V_1) = \exp(-V_1 \psi_1^{-1} \circ \psi_3(\cdot)).$$

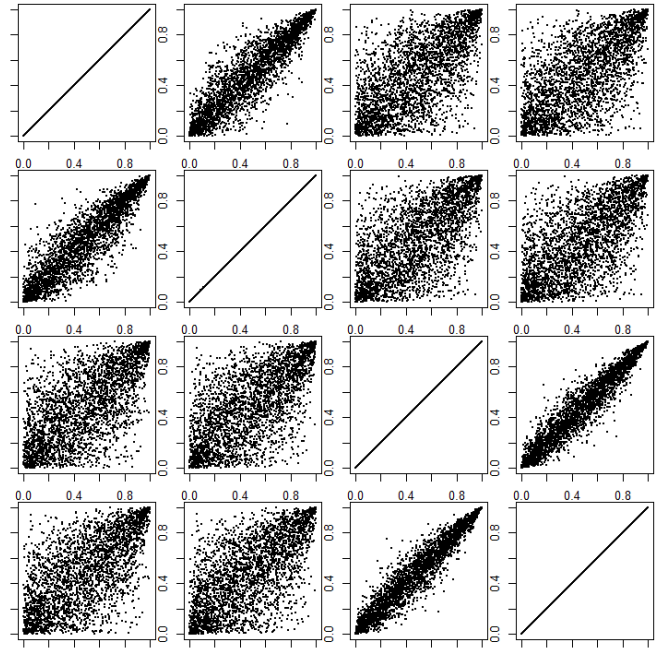
Generáljunk egymástól független egyenletes eloszlású X_1, X_2, X_3, X_4 változókat.

Térjünk vissza (U_1, U_2, U_3, U_4) értékekkel, ahol $U_i = \psi_2(-\frac{\ln(X_i)}{V_2}), i = 1, 2$ és

$$U_i = \psi_3(-\frac{\ln(X_i)}{V_3}), i = 3, 4$$



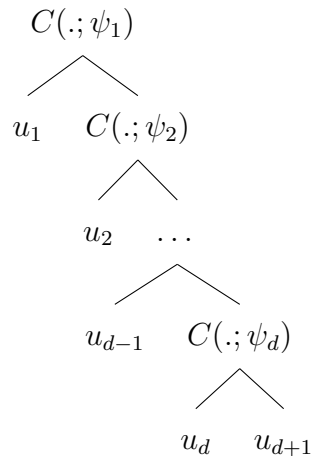
(a) A kopulastruktúra fagráf ábrázolása



(b) Minta páros négydimenziós Gumbel kopulából, $\theta_1 = 2, \theta_2 = 4, \theta_3 = 6$

3.2.4. Az általános beágyazott eset

A bevezetőben már találkoztunk olyan kopulakonstrukcióval, ahol a generátorfüggvények tetszőlegesen mélyre vannak beágyazva (9). A könnyebb értelmezhetőség érdekében először ábrázoljuk ezt a megszokott fagráfként:



A továbbiakban feltételezni fogjuk, hogy $\psi_1, \psi_2, \dots, \psi_d$ LT-arkhimédészi kopulagenerátorok, illetve

hogy a $\psi_k^{-1} \circ \psi_{k+1}$ függvénynek teljesen monoton a deriváltja $k = 1, \dots, d-1$ esetén. A következő lemma bizonyítja, hogy $\psi_k^{-1} \circ \psi_j$ deriváltja teljesen monoton, bármely $j > k$ -ra:

9. Lemma. *Ha $\psi_1^{-1} \circ \psi_2$ és $\psi_2^{-1} \circ \psi_3$ deriváltja egyaránt teljesen monoton, akkor $\psi_1^{-1} \circ \psi_3$ deriváltja is az lesz.*

Legyen $f = \psi_1^{-1} \circ \psi_2$, $g = \psi_2^{-1} \circ \psi_3$ és $h = \psi_1^{-1} \circ \psi_3$ úgy, hogy $h = f \circ g$! A derivált $h'(t) = f'(g(t))g'(t)$, $t > 0$ lesz, h' pedig így teljesen monoton, hiszen $f' \circ g$ teljesen monoton, ami adódik abból, hogy egy teljesen monoton függvénye egy olyan függvénynek, aminek a deriváltja is teljesen monoton. h' így két teljesen monoton függvény kompozíciója, tehát maga is teljesen monoton lesz. $\square$

Tegyük fel, hogy $k = 1, \dots, d-1$, $j > k$ és $\nu > 0$ esetén definiáljuk a következő kopulagenerátor-függvényeket:

$$\psi_j^k(\cdot; \nu) := \exp(-\nu \psi_k^{-1} \circ \psi_j(\cdot)) \quad (10)$$

Mivel ezek teljesen monotonok, így LT-arkhimédészi generátorok lesznek, a továbbiakban pedig azt a szerepet fogják betölteni, hogy ezek funkcionálnak majd a k által indexelt rekurzívan beágyazott belső kopulák generátoraiként. Ez a tulajdonság az alábbi lemmából adódik:

10. Lemma. *Minden $1 \leq k \leq d-2$, $k+1 < j \leq d$ és $\nu, \tilde{\nu}$ -re:*

$$\exp(-\nu \psi_{k+1}^{(k)-1}(\psi_j^k(\cdot; \tilde{\nu}); \tilde{\nu})) = \psi_j^{k+1}(\cdot; \nu) \quad (11)$$

A fenti lemma egyértelműen következik abból, hogy

$$\psi_{k+1}^{(k)-1}(\psi_j^k(\cdot; \tilde{\nu}); \tilde{\nu}) = \psi_{k+1}^{-1} \circ \psi_j \quad (12)$$

$\square$

Bizonyítható továbbá az alábbi lemma is, melyet a későbbiekben még felhasználunk:

11. Lemma. *Legyen $d \geq 1$ adott és feltételezzük, hogy $\psi_1, \psi_2, \dots, \psi_d$ LT-arkhimédészi generátorok. Legyen $F : [0, 1] \rightarrow [0, 1]$ egy szigorúan növekvő folytonos függvény úgy, hogy $\tilde{\psi}_k = F \circ \psi_k$, $k = 1, \dots, d$ szintén LT-arkhimédészi generátorok. Ekkor a bevezetőben definiált d -dimenziós kopulára (9) igaz, hogy:*

$$F(C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d)) = C_d(F(u_1), \dots, F(u_{d+1}); \tilde{\psi}_1, \dots, \tilde{\psi}_d) \quad (13)$$

12. Tétel. *Ha $\psi_1, \dots, \psi_d$ LT-arkhimédészi generátorok és $\psi_k^{-1} \circ \psi_{k+1}$ deriváltfüggvénye folyamatosan monoton $k = 1, \dots, d-1$ esetén, akkor a bevezetőben definiált $C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d)$ függvény (9) kopula, a következő keverék-reprezentációkkal:*

$$\begin{aligned} C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d) &= \\ &= \int_0^\infty F_1^{\nu_1}(u_1) C_{d-1}(F_1^{\nu_1}(u_2), \dots, F_1^{\nu_1}(u_{d+1}); \psi_2^{(1)}(\cdot; \nu_1), \dots, \psi_d^{(1)}(\cdot; \nu_1)) dG_1(\nu_1) \\ &= \int_0^\infty \dots \int_0^\infty F_1^{\nu_1}(u_1) \dots F_d^{\nu_d}(u_d) F_d^{\nu_d}(u_{d+1}) dG_d(\nu_d; \nu_{d-1}) \dots dG_2(\nu_2; \nu_1) dG_1(\nu_1), \end{aligned} \quad (14)$$

ahol G_1 Laplace-transzformáltja ψ_1 , $F_k(u) = \exp(-\psi_k^{-1}(u))$, $k = 1, \dots, d$ és $G_k(\nu; \nu_{k-1})$ az az eloszlás-függvény, melynek Laplace-transzformáltja $\psi_k^{(k-1)}(\cdot; \nu_{k-1})$, $k = 2, \dots, d$.

Bizonyítás. Ha $C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d)$ egy többváltozós eloszlásfüggvény bármely d -re mely kielégíti a tétel feltételeit, akkor egyértelműen kopula, hiszen minden peremeloszlása standard egyenletes. Megmutatjuk, hogy C_d tényleg többváltozós eloszlásfüggvény oly módon, hogy deriváljuk a kevert reprezentációit.

A definícióból és ψ^{-1} teljes monotonitásából következik, hogy

$$C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d) = \int_0^\infty e^{-\nu_1 \psi_1^{-1}(u_1)} e^{-\nu_1 \psi_1^{-1}(C_{d-1}(u_2, \dots, u_{d+1}; \psi_2, \dots, \psi_d))} dG_1(\nu_1), \quad (15)$$

ahol G_1 eloszlásfüggvény $[0, \infty)$ -n, ψ_1 Laplace-transzformálttal. A szorzat első tényezője $F_1^{\nu_1}(u_1)$, amely egy egyértelműen folytonos és szigorúan növekvő egyváltozós eloszlásfüggvény $[0, 1]$ -en. Felhasználhatjuk továbbá a már korábban kimondott 11. lemmát, miszerint a második tényező felírható

$$C_{d-1}(F_1^{\nu_1}(u_2), \dots, F_1^{\nu_1}(u_{d+1}); \psi_2^{(1)}(\cdot; \nu_1), \dots, \psi_d^{(1)}(\cdot; \nu_1)) \quad (16)$$

alakban. Ha $C_{d-1}(u_2, \dots, u_{d+1}; \psi_2^{(1)}(\cdot; \nu_1), \dots, \psi_d^{(1)}(\cdot; \nu_1))$ kopula, akkor ez egy többváltozós eloszlásfüggvény és $C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d)$ függvény a már említett kevert reprezentációval fog rendelkezni és így egy többváltozós eloszlásfüggvény lesz. Ezáltal sikeresen visszavezettük a problémát arra, hogy elég megmutatnunk, hogy $C_{d-1}(u_2, \dots, u_{d+1}; \psi_2^{(1)}(\cdot; \nu_1), \dots, \psi_d^{(1)}(\cdot; \nu_1))$ egy többváltozós eloszlásfüggvény, azaz kopula. Általánosságban a fenti gondolatmenetet ismételve a következő reprezentáció sorozatot kapjuk $k = 2, \dots, d-1$ esetén:

$$\begin{aligned} & C_{d-k+1}(u_k, \dots, u_{d+1}; \psi_k^{(k-1)}(\cdot; \nu_{k-1}), \dots, \psi_d^{(k-1)}(\cdot; \nu_{k-1})) = \\ &= \int_0^\infty F_k^{\nu_k}(u_k; \nu_{k-1}) C_{d-k}(F_k^{\nu_k}(u_{k+1}; \nu_{k-1}), \dots, F_k^{\nu_k}(u_{d+1}; \nu_{k-1}); \\ & \quad \psi_{k+1}^{(k)}(\cdot; \nu_k), \dots, \psi_d^{(k)}(\cdot; \nu_k)) dG_k(\nu_k; \nu_{k-1}), \end{aligned} \quad (17)$$

ahol $F_k(u; \nu_{k-1}) = \exp(\psi_k^{(k-1)-1}(\cdot; \nu_{k-1}))$. Minden egyes $k = 2, \dots, d-1$ lépéskor ha feltételezhetjük, hogy

$$C_{d-k}(u_{k+1}, \dots, u_{d+1}; \psi_{k+1}^{(k)}(\cdot; \nu_k), \dots, \psi_d^{(k)}(\cdot; \nu_k)) \quad (18)$$

kopula, az úgynevezett k -adik belső kopula, akkor megkapjuk az állított keverékábrázolást az előző belső kopulához

$$C_{d-k+1}(u_k, \dots, u_{d+1}; \psi_k^{(k-1)}(\cdot; \nu_{k-1}), \dots, \psi_d^{(k-1)}(\cdot; \nu_{k-1})) \quad (19)$$

A belső kopulák generátorainak alakja a 10 lemmából következik. Az induktív folyamat akkor fejeződik be, amikor eljutunk $C_1(u_d, u_{d+1}; \psi_d^{(d-1)}(\cdot; \nu_{d-1}))$ végső belső kopulához. Ez egyértelműen kopula, hiszen a generátorfüggvénye teljesen monoton és felírható a következő reprezentációja:

$$C_1(u_d, u_{d+1}; \psi_d^{(d-1)}(\cdot; \nu_{d-1})) = \int_0^\infty F_d^{\nu_d}(u_d; \nu_{d-1}) F_d^{\nu_d}(u_{d+1}; \nu_{d-1}) dG_d(\nu_d; \nu_{d-1}) \quad (20)$$

A második keverék-reprezentáció megkapható, ha leegyszerűsítjük az egymásba ágyazott reprezentációkat, tekintettel arra, hogy $k \geq 2$

$$F_k^{\nu_k}(F_{k-1}^{\nu_{k-1}}(\dots F_2^{\nu_2}(F_1^{\nu_1}(u); \nu_1); \nu_2) \dots; \nu_{k-1}) = F_k^{\nu_k}(u). \quad (21)$$

□

A $C_d(u_1, \dots, u_{d+1}; \psi_1, \dots, \psi_d)$ kopula véletlen mintavételezéséhez a következő, a kopula első keverék-reprezentációját felhasználó algoritmust használhatjuk:

Algoritmus 4: d dimenziós beágyazott eset

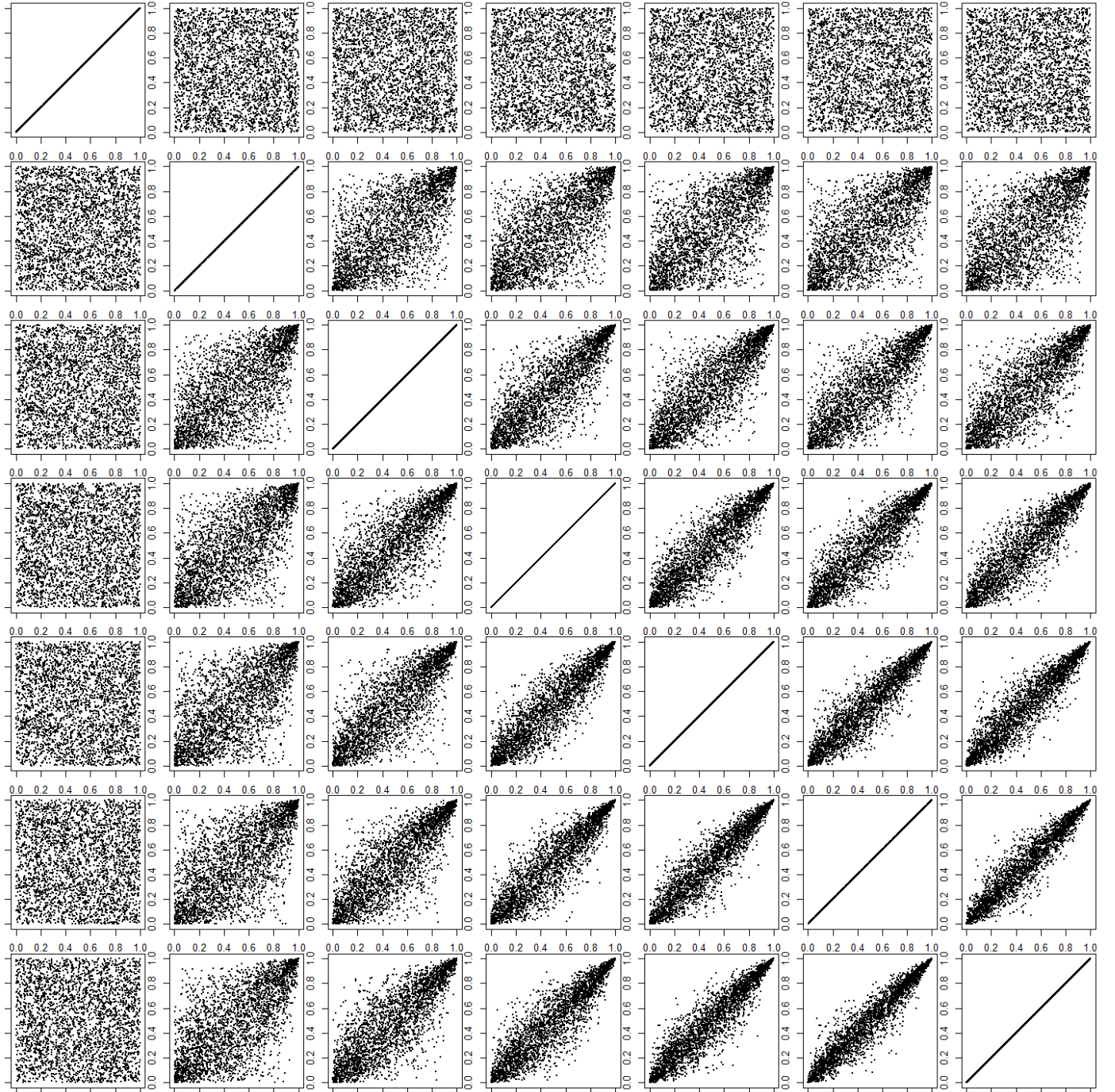
Generáljunk egy V változót G_1 eloszlásfüggvénnyel, melynek Laplace-transzformáltja ψ_1 .

Generáljunk egy standard egyenletes eloszlású X_1 változót.

Generáljunk $(X_2, \dots, X_{d+1})$ változókat a $C_{d-1}(u_2, \dots, u_{d+1}; \psi_2^{(1)}(\cdot; V), \dots, \psi_d^{(1)}(\cdot; V))$ kopulából, ugyanennek az algoritmusnak a segítségével.

Térjünk vissza $(U_1, \dots, U_{d+1})$ értékekkel, ahol $U_i = \psi_1(\frac{-\ln X_i}{V}), i = 1, \dots, d+1$

A következő ábrát a fenti algoritmussal generáltuk, hétdimenziós Gumbel kopulákból úgy, hogy a Kendall-féle τ -k értékei sorban $0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}$ legyenek.



6. ábra. $\theta_1 = 1; \theta_2 = 2; \theta_3 = 3; \theta_4 = 4; \theta_5 = 5; \theta_6 = 6$

3.3. Vine kopulák

Az előző fejezetben megismerkedhettünk az egymásba ágyazott arkhimédészi kopulastruktúrákkal. Fontos megemlítenünk, hogy létezik egy hasonló, fastruktúrával ábrázolható kopulacsalád, melyet Vine kopulának nevezünk. Ez a család a gyökeres fák helyett egy másik gráfelméleti fogalmat használ a szemléltetéshez, az úgynevezett vine-okat, amelyekkel először meg kell ismerkednünk, mielőtt a kopulacsaládot definiáljuk.

3.3.1. Szabályos vine fasorozatok

13. Definíció (Szabályos (R-) vine fasorozat). Valamely fagrafok $\mathcal{V} = (T_1, T_2, \dots, T_{d-1})$ halmaza egy szabályos vine fasorozat d elemen, ha:

- Minden $T_j = (N_j, E_j)$ fa esetén minden $a, b \in T_j, j = 1, \dots, j-1$ csúcsra létezik egy $n_1, \dots, n_k \subset N_j$ út, ahol $a = n_1, b = n_k$.
- T_1 egy fa, melynek csúcsai $N_1 = \{1, \dots, d\}$, élei pedig E_1 .
- Minden $j \geq 2$, esetén T_j egy fa, melynek csúcsai $N_j = E_{j-1}$, élei pedig E_j .
- Minden $j = 2, \dots, d-1$ és $\{a, b\} \in E_j$ esetén $|a \cap b| = 1$.

14. Definíció (Szabályos vine eloszlás). F együttes eloszlásfüggvény $\mathbf{X} = (X_1, X_2, \dots, X_d)$ d -dimenziós véletlen vektorváltozó eloszlása szabályos vine eloszlás, amennyiben meg tudunk határozni egy olyan $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ hármast, melyre igazak az alábbiak:

- **Peremeloszlások:** $\mathcal{F} = (F_1, \dots, F_d)$ folytonos invertálható peremeloszlások vektora, amelyek az $X_i, i_1, \dots, d$ valószínűségi változó peremeloszlás-függvényeit reprezentálják.
- **Szabályos vine fasorozat:** $\mathcal{V}$ egy szabályos vine fasorozat d elemen.
- **Kétváltozós kopulák:** $\mathcal{B} = \{C_e \mid e \in E_i, i = 1, \dots, d-1\}$, ahol C_e egy szimmetrikus kétváltozós kopula. E_i a $\mathcal{V}$ szabályos vine fastruktúrában található T_i fa éleinek halmaza.
- **Kapcsolat $\mathcal{V}$ és $\mathcal{B}$ között:** Minden $e \in E_i, i = 1, \dots, d-1, e = \{a, b\}$ esetén C_e az $X_{C_{e,a}}$ és $X_{C_{e,b}}$ feltételes eloszlásokhoz kapcsolódó kopula, $\mathbf{X}_{D_e} = \mathbf{x}_{D_e}$ feltétel mellett. $C_e(\cdot, \cdot)$ nem függ x_{D_e} specifikus értékétől.

15. Definíció (Egy e élen értelmezett pár-kopula és kopula-sűrűségfüggvény). Az e élhez tartozó C_e kopulát $C_{C_{e,a}C_{e,b};D_e}$, a hozzá tartozó sűrűségfüggvényt pedig $c_{C_{e,a}C_{e,b};D_e}$ jelöléssel jelöljük. Ezt a kopulát nevezzük pár-kopulának.

13. Tétel (Szabályos vine eloszlás létezése). *Tegyük fel, hogy $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ kielégíti a 14. definícióban szereplő tulajdonságokat az utolsó kivételével. Ekkor létezik egy olyan egyedi d -dimenziós F eloszlás, melynek sűrűsége*

$$f_{1,\dots,d}(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d) \cdot \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{C_{e,a}C_{e,b};D_e}(F_{C_{e,a}|D_e}(x_{C_{e,a}} \mid \mathbf{x}_{D_e}), F_{C_{e,b}|D_e}(x_{C_{e,b}} \mid \mathbf{x}_{D_e}))$$

úgy, hogy minden $e \in E_i, i = 1, \dots, d-1, e = \{a, b\}$ esetén $X_{C_{e,a}}$ és $X_{C_{e,b}}$ eloszlásfüggvényekre teljesül az

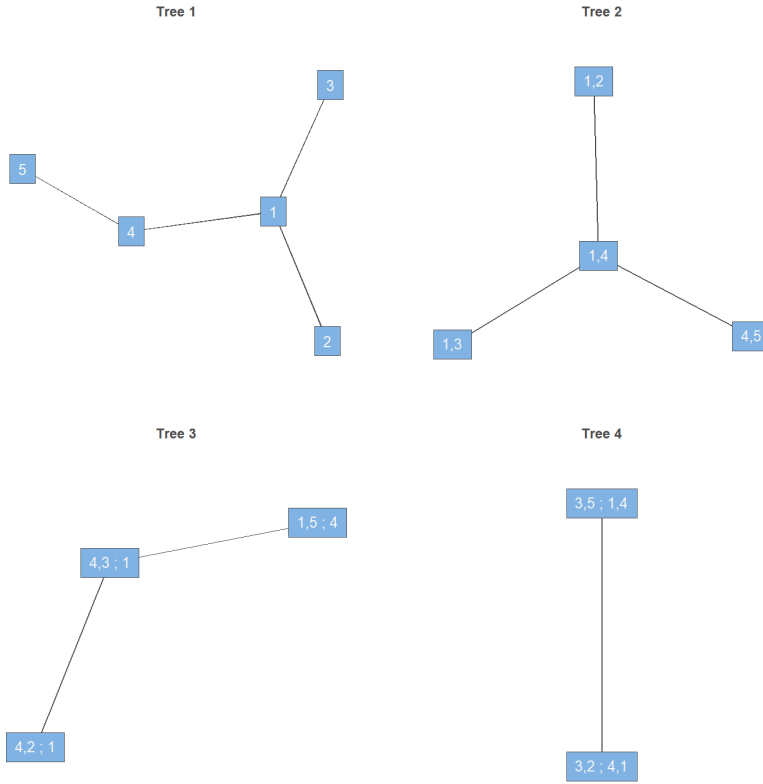
$$F_{C_{e,a}C_{e,b}|D_e}(x_{C_{e,a}}, x_{C_{e,b}} \mid \mathbf{x}_{D_e}) = C_e(F_{C_{e,a}|D_e}(x_{C_{e,a}} \mid \mathbf{x}_{D_e}), F_{C_{e,b}|D_e}(x_{C_{e,b}} \mid \mathbf{x}_{D_e}))$$

összefüggés, amennyiben $\mathbf{X}_{D_e} = \mathbf{x}_{D_e}$, F egydimenziós peremeit pedig $F_i(x_i), i = 1, \dots, d$ adja meg.

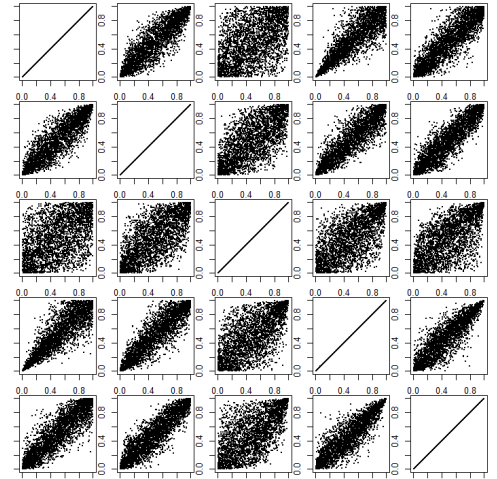
A fenti tétel lehetővé teszi a szabályos vine eloszlások meghatározását a $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ hármas segítségével ezekből pedig mindig előállítható a hozzájuk tartozó együttes sűrűség.

16. Definíció (Szabályos vine kopula). Egy szabályos vine kopula, egy olyan szabályos vine eloszlás, melynek az összes pereme egyenletesen oszlik el a $[0, 1]$ intervallumon.

A vine kopulák sokszínűségét és rugalmasságát a következő ábra mutatja be, ahol egy vine kopula vine fasorozatát, illetve a kopulából vett mintát láthatjuk. A mintavétel ábrázolásából látszik, hogy mennyire sokszínű kétdimenziós peremeket képes a kopula modellezni. A peremeloszlások ebben az esetben Gauss, Clayton, illetve Gumbel kopulákból származnak.



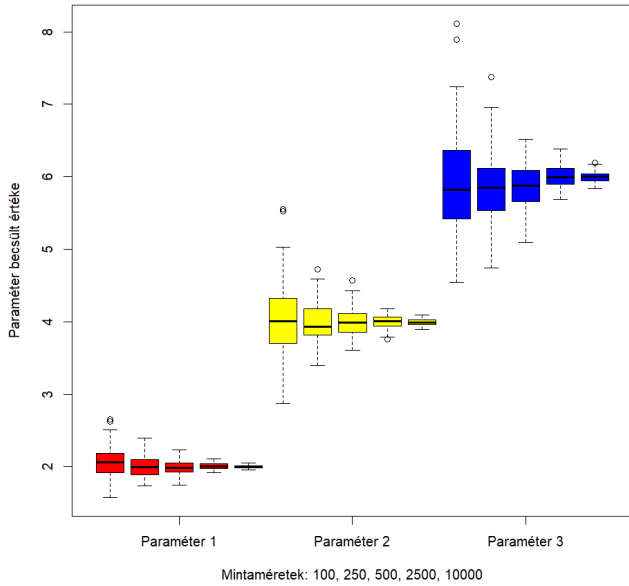
(a) A kopula szabályos vine fasorozata



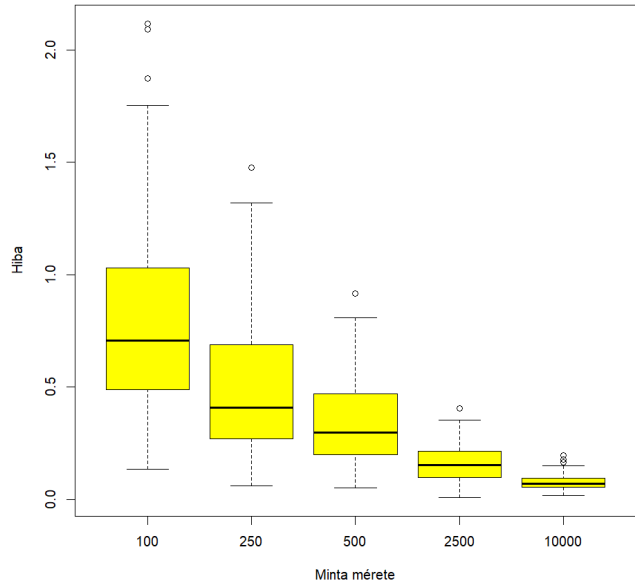
(b) Minta a kopulából

4. Arkhimédészi kopulastruktúra illesztése adathalmazra

Szeretnénk az adathalmazunkra a dolgozatban eddig taglalt egymásba ágyazott arkhimédészi kopulastruktúrát illeszteni, majd megvizsgálni, hogy ez mennyire írja le jól az eredeti adatainkat, más modellekhez képest. Ehhez a HAC R csomag [10] nyújt segítséget, melyben megtalálható egy eljárás, ami egy tetszőleges adathalmazra becsül egymásba ágyazott arkhimédészi kopulastruktúrát. Vizsgáljuk először ennek a pontosságát és hatékonyságát. Ehhez először generáljunk egy saját adathalmazt, melynek ismerjük a szerkezetét, majd alkalmazzuk erre a metódust és vizsgáljuk, hogy mennyire pontosan adja vissza az eredeti paramétereinket. A mintavételezést páros négydimenziós Gumbel kopulastruktúrából végeztük, ahol a kopulák paraméterei sorban $\theta_1 = 2, \theta_2 = 4, \theta_3 = 6$, fagráf ábrázolása pedig a 3.2.3. fejezetben leírtak szerint épül fel. A szimulációt különböző elemszámú mintákra hajtottuk végre, esetenként 100 ismétléssel. Az algoritmus minden alkalommal sikeresen adta vissza a fastruktúrát. A paraméterek becslésének pontossága az alábbi ábrákon látható, az elsőn a paraméterek értékei, a másodikon pedig az összhiba értéke, melyet az eltérések kettes normájával számoltuk.



(a) A becsült paraméterek értékei



(b) A becslés hibája a paraméterekre nézve

Az ábrákról leolvasható a becslés pontosságának konvergenciája, amely a négyzetgyök rendűnél is lassabbnak, illetve aszimptotikusan torzítatlannak tűnik.

4.1. Illeszkedésvizsgálat

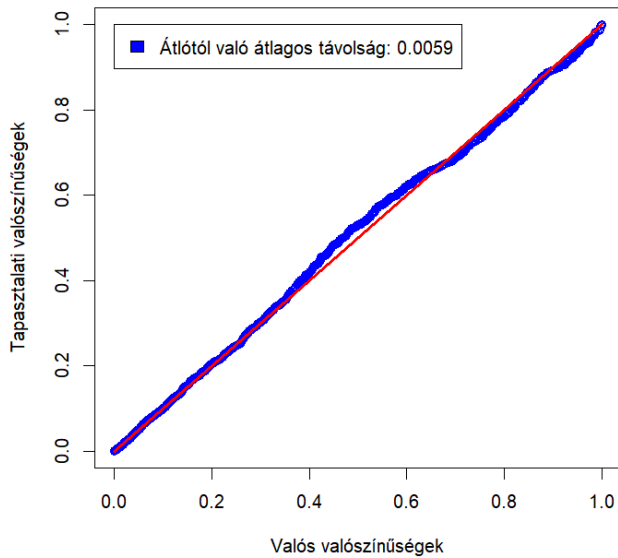
Ahhoz, hogy a gyakorlatban használni tudjuk az előző fejezetben taglaltakat, meg kell határoznunk, hogy mennyire írja le jól az illesztett struktúra az eredeti adathalmazt. Ehhez először vegyük az adathalmazhoz tartozó $\hat{C}(u_1, u_2, \dots, u_d)$ tapasztalati kopulát, melyet az alábbi módon kaphatunk meg:

$$\hat{C}(u_1, u_2, \dots, u_d) = n^{-1} \sum_{i=1}^n \prod_{j=1}^d \mathbf{I} \left\{ \hat{F}_j(X_{ij}) \leq u_j \right\},$$

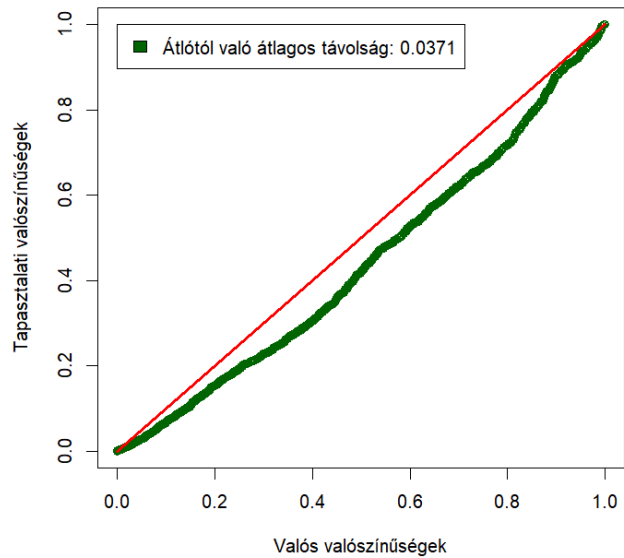
ahol $\hat{F}(\cdot)$ az X_j valószínűségi változó becsült eloszlásfüggvényét jelöli.

Ezek után becsüljük meg a kopulastruktúrát az előző fejezetben foglaltak szerint, majd vegyünk az adathalmazunk méretével megegyező darabszámú mintát ebből a struktúrából. A mintavételezés menetére már a dolgozat előző fejezeteiben mutattunk módszereket különböző esetekre. A szimulált

adatoknak szintén vegyük a tapasztalati kopuláját, aztán ennek a sorba rendezett értékeit ábrázoljuk az eredeti adatokra vett tapasztalati kopula sorba rendezett értékei függvényében. Minél jobban simulnak a kapott pontok a $(0,0)$, $(1,1)$ pontok által meghatározott átlóra, annál pontosabb a becslésünk. A következő ábrákon láthatjuk hogyan néz ez ki, ha jól illetve, ha rosszul becsljük a kopulát. Mindkét esetben az adathalmazt az előző fejezet négydimenziós struktúrájából generáltuk, az első esetben az eddig használt metódust használva határoztuk meg a kopulát, míg a másodikban ránézésre próbáltunk valamilyen kopulastruktúrát illeszteni az adatokra.



(a) A kopulabecslő algoritmus pontossága



(b) Saját becslésünk pontossága

Látható, hogy az első ábra pontjai sokkal jobban illeszkednek a pirossal jelölt átlóra, mint a második ábráé, ebből pedig következik, hogy a HAC [10] csomag kopulabecslő algoritmus a lényegesen pontosabb eredményt adott, mint amikor ránézésre próbáltuk a struktúrát meghatározni.

Az átlóhoz illeszkedés mértékét számszerűsíthetjük is, például a pontok átlótól való távolságának átlagával. A fenti ábráknál ezeket az értékeket is jelöltük.

5. Gyakorlati alkalmazás

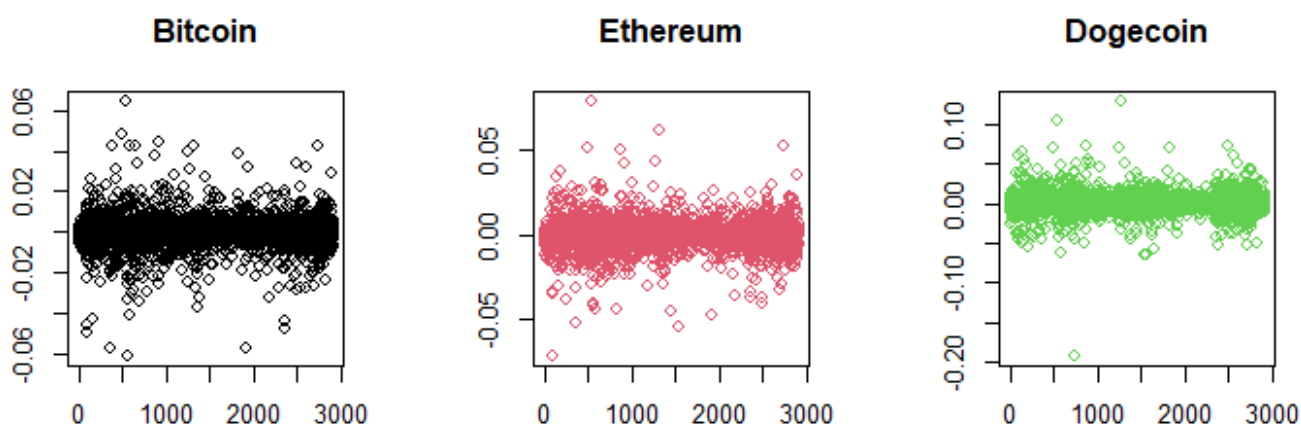
Az előző fejezetekben bemutattuk az egymásba ágyazott arkhimédészi kopulák elméleti hátterét, felépítését, illetve mintavételezését néhány specifikus esetben. A továbbiakban azt szeretnénk vizsgálni, hogy van-e gyakorlati előnye ezeknek a struktúráknak más, ismertebb kopulacsaládokkal szemben. A gyakorlatban a kopulákat leginkább pénzügyi adatok elemzésére használják, mi is így egy ilyen típusú adathalmazt kerestünk. Választásunk különböző kriptovaluták 3 órás logaritmikus-hozam adataira esett, a 2023-as évben, mely adatok a Yahoo Finance weboldalán [11] megtalálhatóak. Ezeket az adatokat fogjuk először alapvető statisztikai módszerekkel elemezni, majd megpróbálni a dolgozat előzőleg taglalt struktúráit is alkalmazni rájuk.

5.1. Az adathalmaz elemzése

Ahhoz, hogy bonyolultabb módszereket alkalmazhassunk az adathalmazunkon, mélyebben meg kell ismernünk annak belső szerkezetét. Az adatok 11 kriptovaluta árfolyamait írják le, az amerikai dollár értékében. Ezen kriptovaluták a következők:

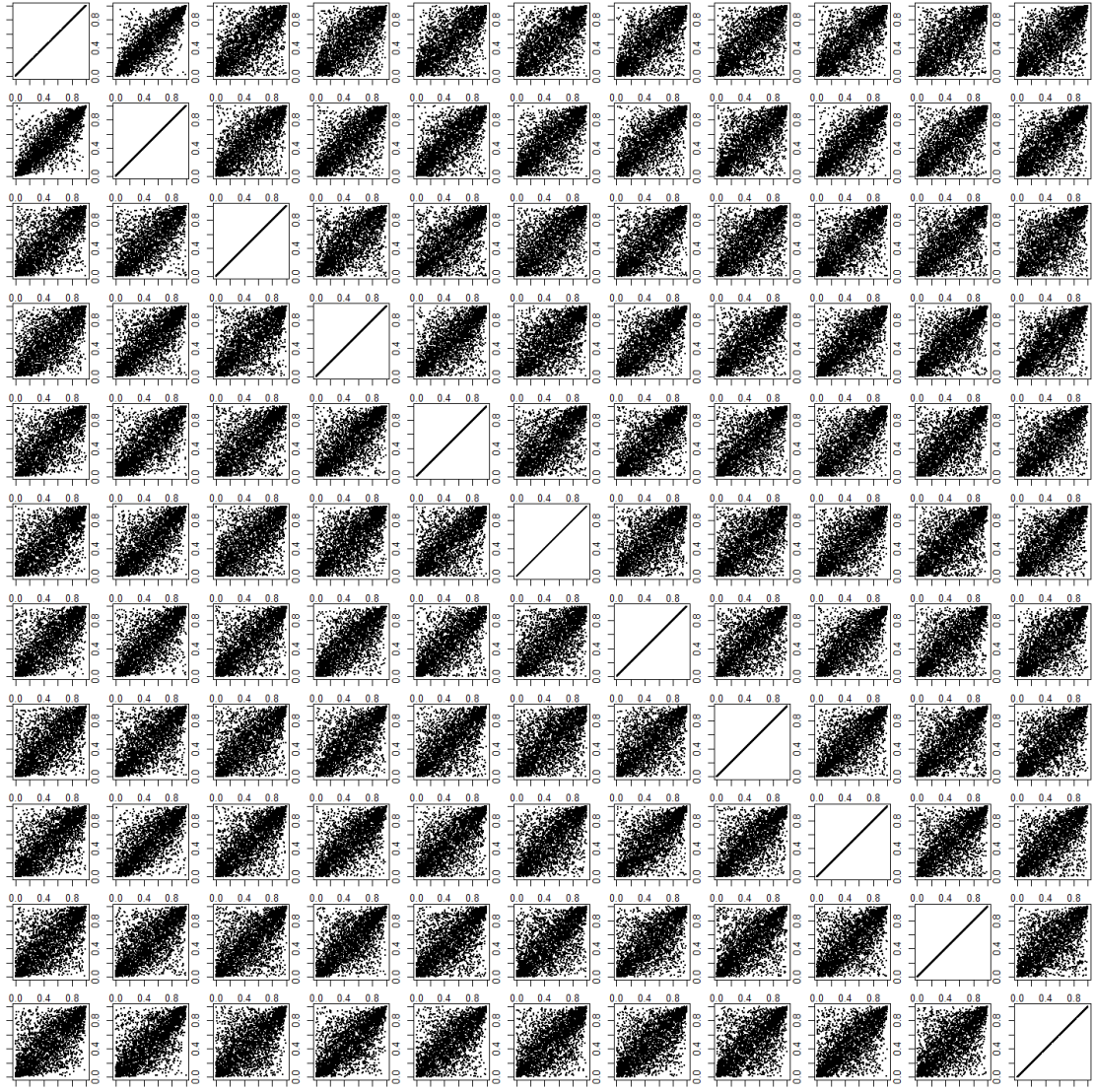
- Bitcoin (BTCUSD)
- Ethereum (ETCUSD)
- Solana (SOLUSD)
- Polkadot (DOTUSD)
- Dogecoin (DOGE.USD)
- Litecoin (LTCUSD)
- Cardano (ADAUSD)
- Fantom (FTMUSD)
- Polygon (MATIC.USD)
- Filecoin (FILUSD)
- EOS (EOSUSD)

Minden dimenzió 2920 rekordot tartalmaz, melyek az adott kriptovaluta 3 óránkénti logaritmikus-hozamait jelölik. A logaritmikus-hozam skálára váltásra azért van szükség, mert így kerülünk közel a stacionaritáshoz, ami kell ahhoz, hogy egy fix kopulával írassuk le az összefüggéseket. A hozam adatok mindig a valuta értékének változását jelölik annak előző megfigyelt értékéhez képest.



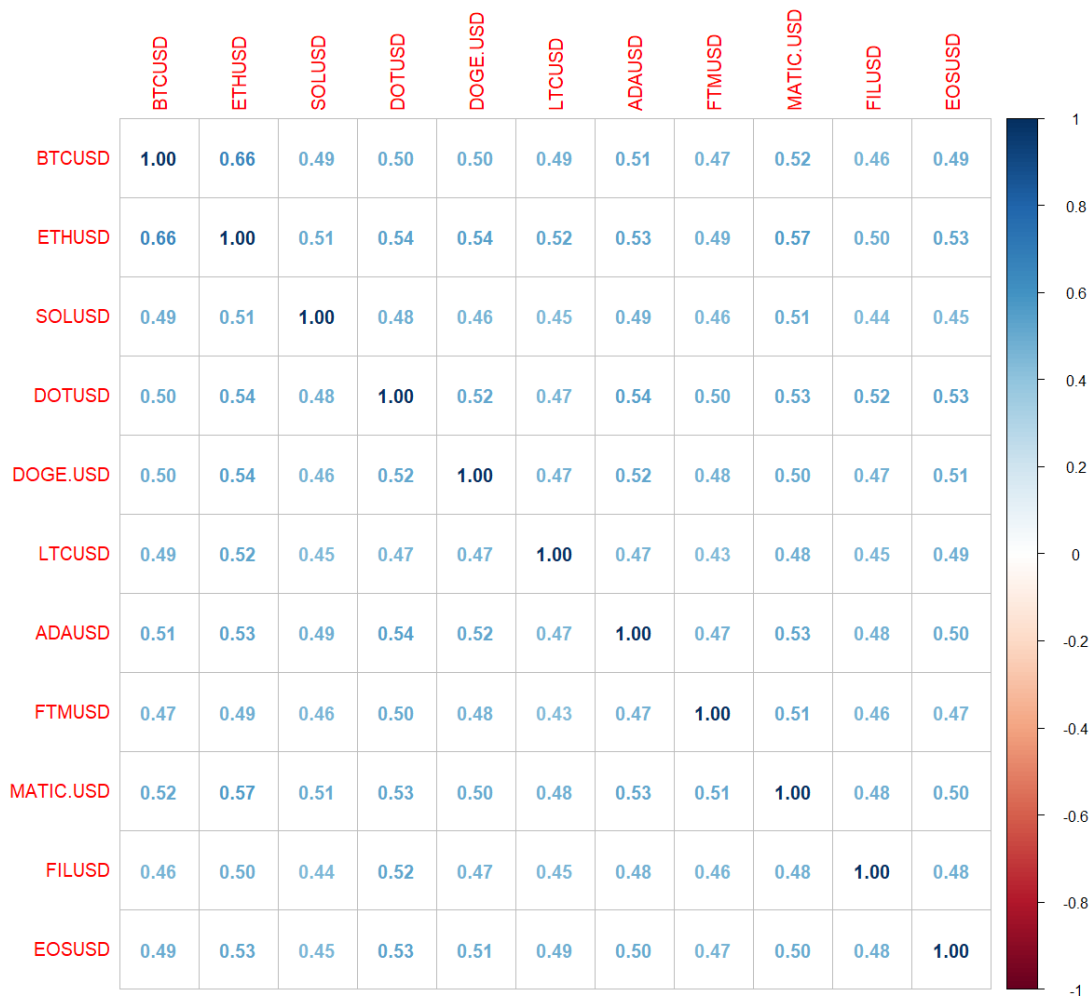
10. ábra. 3 kriptovaluta logaritmikus-hozam adatai

Ahhoz, hogy ezekre az adatokra bármiféle kopula-struktúrát tudjunk illeszteni, először is pseudo-megfigyelésekké kell alakítanunk az értékeket. Ezek után ábrázolhatjuk a megkapott struktúrát, mely segítségével következtetéseket vonhatunk le.



11. ábra. A pszeudomegfigyelések kétdimenziós peremei

Az ábra alapján láthatjuk, hogy a kétdimenziós peremek között többnyire szimmetrikus struktúrák találhatók. Minél inkább az átló mentén koncentrálódnak a megfigyeléseink, annál inkább valószínűsíthetünk magasabb korrelációs értéket a két kriptovaluta között. A Kendall-féle korrelációk ki is számolhatóak a mintából:

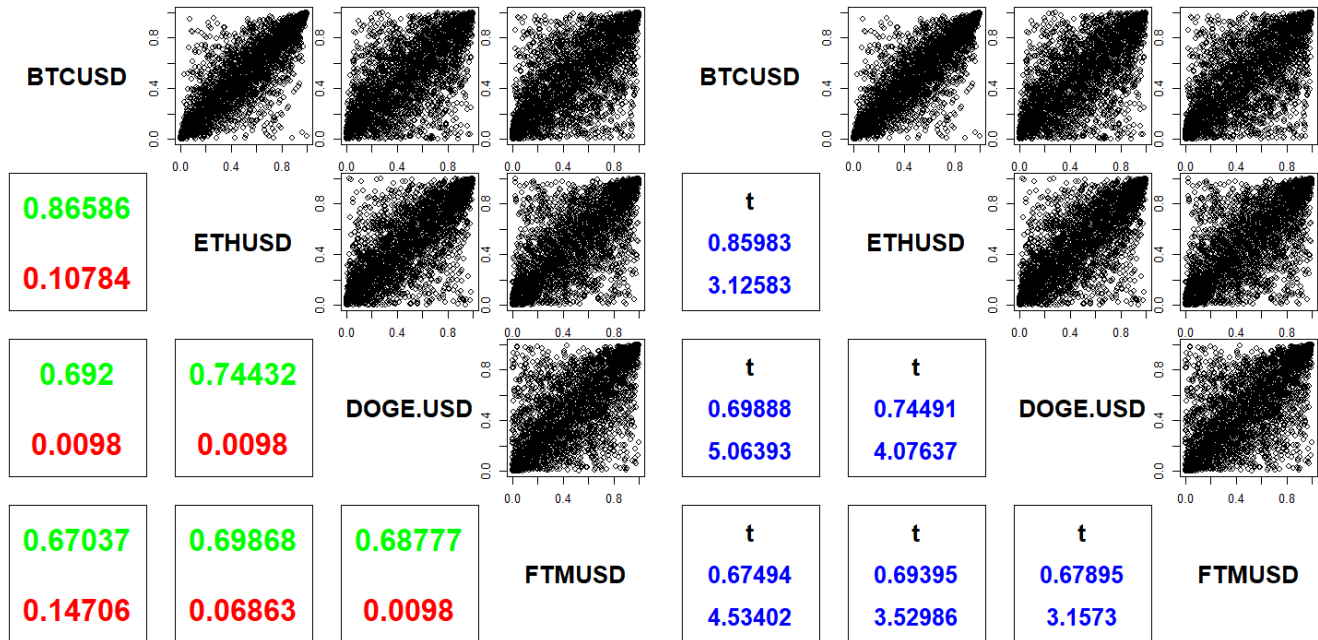


12. ábra. A megfigyelések Kendall-féle τ korrelációi

Mivel az eredeti adathalmaz mind a 11 dimenziójának az elemzése rendkívül körülményes, sőt egyes esetekben nem is lehetséges, az első lépésekben alacsonyabb dimenziókra szűkítjük azt, majd ezekből próbálunk eljutni a magasabb dimenziókhoz. Csökkentsük a mintát először 4 dimenzióra! Legyenek a vizsgált kriptovaluták a Bitcoin, az Ethereum, a Dogecoin és a Fantom. Első lépésként az összes kétdimenziós esetre megvizsgáljuk, hogy egyes kopulák mennyire illeszkednek a megfelelő adathalmazra. Ehhez a `copula` R csomag [8] nyújt segítséget, melyben található egy próba, amely egy adott kopulacsald esetén megpróbálja a legjobb paramétert megtalálni, majd eldönteni, hogy származhat-e a vizsgált minta az adott kopulából. A teszt működéséről az R csomag dokumentációjában találhatunk pontos leírást, a `gofCopula` függvény alatt. Négyfajta kopulacsaldra futtatunk tesztet, mégpedig a t , Clayton, Gumbel és Gauss kopulákra. Az eredményeket a lentebb található ábrán bal oldalt (13a) láthatjuk t -kopula esetén, ahol az összes kétdimenziós peremhez a zöld érték az illeszkedésvizsgálat által kapott legjobb paramétert, a piros pedig a p -értéket jelöli.

Ha a 0.01 szignifikanciaszintet vesszük, azt tapasztaljuk, hogy a t -kopula esetén vannak olyan kriptovaluta párosítások, melyekre a minta származhat a kopulából, pontosabban a Bitcoin-Ethereum, a Bitcoin-Fantom, illetve az Ethereum-Fantom. Az összes többi esetben a meghatározott szignifikanciaszintnél alacsonyabb p -értéket kapunk (a 0,0098 a legkisebb p -érték amit a teszt ennyi mintavételnél ki tud adni), így ezeknél a párosításoknál elutasíthatjuk azt, hogy a minta a vizsgált kopulából származik. Arra a kérdésre, hogy van-e olyan kétdimenziós kopula amely jobban illeszkedik az adatainkra mint a négy általunk vizsgált is választ kaphatunk. A `VineCopula` R csomag [5] tartalmaz egy módszert,

amely két dimenzió esetén képes rendkívül sok kopulacsalád közül kiválasztani azt, amelyik a leginkább illeszkedik az adatokra. Jelentősen megkönnyítené a dolgunkat, ha létezne ilyen algoritmus magasabb dimenziós kopulák esetén is, de egyelőre ez még nincs kidolgozva, így be kell érünk a kétdimenziós peremek vizsgálatával. Az eredményeket a jobb oldali ábrán (13b) láthatjuk, az egyes esetekre feketével jelölve olvashatjuk az ajánlott kopula nevét, alatta kékkel a paraméterét és t -kopula esetén a szabadsági fokát.



(a) t -kopula illeszkedésvizsgálata a kétdimenziós peremekre (b) Legjobban illeszkedő kopulák a kétdimenziós peremekre

Láthatjuk, hogy az algoritmus minden esetben a t -kopulát javasolta mint legjobb illeszkedés, gyakorlatilag ugyanazokkal a paraméterekkel, mint amiket a 13a. ábrán olvashatunk, azaz vannak olyan kétdimenziós peremek, melyekre hiába a legjobb opciónk a t kopula, az illeszkedésvizsgálat azt sem tudta elfogadni.

5.2. Különböző kopulák összehasonlítása

Ahhoz, hogy meg tudjuk határozni, hogy a dolgozatban említett kopulák közül melyik írja le legjobban az adathalmazunkat, be kell vezetnünk valamiféle mérőszámot, amelyet minden kopula esetén ki tudunk számolni és azt írja le, hogy mennyire jól illeszkedik az adott kopula az adatokra. Ezeket összehasonlítva meghatározható a modellek közötti erősrend. Módosítsuk ehhez a 4.1. fejezetben leírt mérőszámot, mely az eredeti, illetve a szimulált adatok tapasztalati kopuláját rajzolja egy grafikonra, majd tekinti a főátlótól való átlagos eltérést. Ahhoz, hogy a mérőszám a mintaelemszám növekedésével se tartson nullához, szorozzuk be annak négyzetgyökével, így megkapva az új értéket. A következő táblázatban megtalálhatjuk néhány, a dolgozatban vizsgált kopula esetén a legjobban illeszkedő modellt, illetve az előbb említett mérőszám értékét. A pontosabb eredmény érdekében a mintavételezést 100 ismétléssel végeztük el, majd ezek átlagával számoltunk.

Kopulacsalád	Legjobban illeszkedő modell	Mérőszám értéke
Gauss	$\rho_1 = 0.8, \rho_2 = 0.6843, \rho_3 = 0.6677, \rho_4 = 0.7239, \rho_5 = 0.6813, \rho_6 = 0.6577$	0.371
t	$\rho_1 = 0.8644, \rho_2 = 0.6988, \rho_3 = 0.6808, \rho_4 = 0.7446, \rho_5 = 0.7023, \rho_6 = 0.6890, \nu = 4.6717$	0.310
Egymásba ágyazott arkhimédészi	A 3.2.4. fejezetben leírt struktúra, Gumbel kopulából. $\theta_1 = 1.99, \theta_2 = 2.04, \theta_3 = 2.89$	0.766
Vine	A fastruktúra és a paraméterek értékei az R kódban [12] megtekinthetők, a kétdimenziós peremek t , BB1 [7] és Gumbel kopulákból származnak.	0.291

A mérőszámok összehasonlításával kiderül, hogy az adathalmazunkra a legerősebb modell a vine kopula volt, ezt szorosan a négydimenziós t kopula követi, utánuk pedig a Gauss kopula és az egymásba ágyazott arkhimédészi kopulastruktúra következik. Az, hogy a mintavételt 100 ismétléssel végeztük lehetőséget ad arra is, hogy kétmintás t -próba segítségével páronként megvizsgáljuk, hogy szignifikánsan jobb-e az egyik modell a másikonál. A próba p -értékeit a következő ábrán olvashatjuk:



14. ábra. Kétoldali t -próba p -értékei

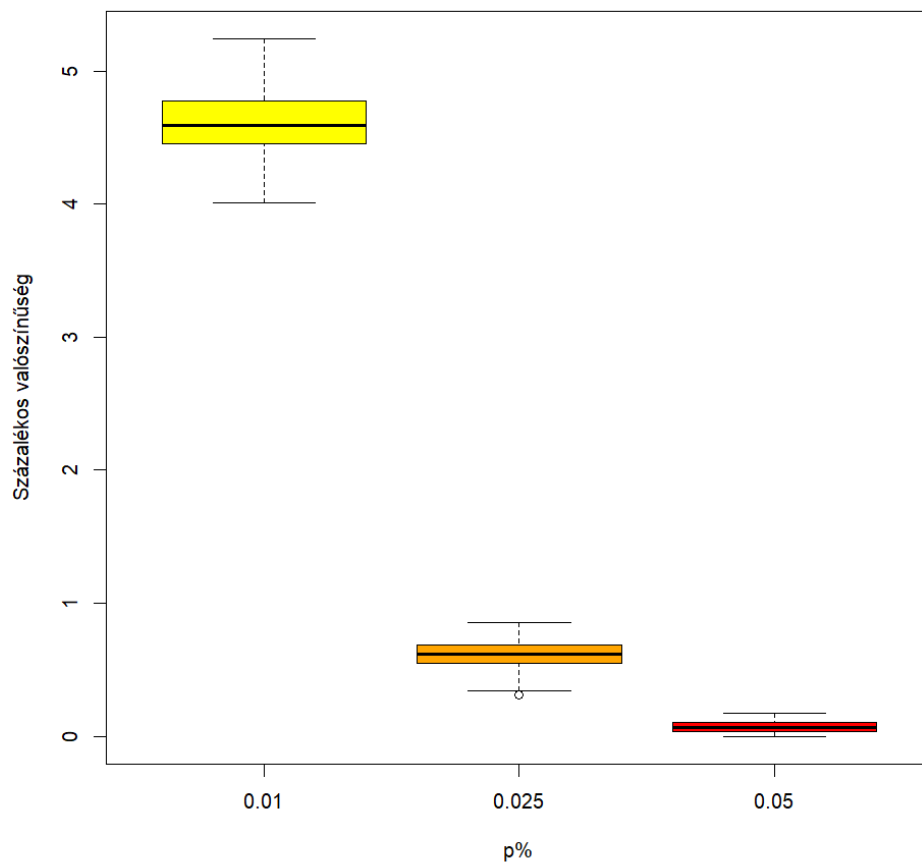
0.01 szignifikanciaszint mellett az összes teszt elutasította a nullhipotézist, azaz az eloszlások szignifikánsan eltérnek egymástól, így a fentebb felállított sorrend nagy valószínűséggel fennáll.

A teljesség kedvéért a fenti folyamatot elvégeztük a 11 dimenziós eredeti adathalmazunkra, de a kopulák paramétereinek mennyisége miatt már csak a mérőszám értékeit jelöltük, a legjobban illeszkedő kopulák adatai szintén az R kódban [12] találhatóak meg részletesen.

Kopulacsalád	Mérőszám értéke
Gauss	0.598
t	0.814
Egymásba ágyazott arkhimédészi	1.298
Vine	0.470

Látható, hogy ha a teljes adathalmazt nézzük, szintén a vine kopula lesz a legerősebb, viszont a 11 dimenziós Gauss kopula itt már pontosabban írja le azt, mint a t -kopula. Az egymásba ágyazott arkhimédészi kopulacsalád ebben az esetben is a leggyengébb modellnek minősül a kipróbáltak közül.

Sikerült így meghatároznunk azt a kopulamodellt, amely a legpontosabban írja le az adathalmazt. Szeretnénk ezt felhasználni példaképpen arra, hogy meghatározhassunk annak a valószínűségét, hogy a 11 kriptovalutánk közül legalább 9 értéke egy kívánt p százalékértéknél többet csökken, hiszen az ilyen jellegű mennyiségek nagyon fontosak a pénzügyi matematikában. Ehhez konvertáljuk vissza a szimulált adatokat az eredeti skálára. Amennyiben $C_1, \dots, C_n$ a szimulált d -dimenziós kopula, akkor az adott koordináta rendezésének megfelelően lesznek a rendezett $x_1, \dots, x_n$ eredeti megfigyelések hozzárendelve az adott dimenzióhoz. A következő ábrán láthatjuk a becsült valószínűségeket három p százalékérték és 100 mintavétel esetén:



15. ábra. Annak a valószínűségei, hogy legalább 9 kriptovaluta értéke több mint $p\%$ -ot csökken

Ellenőrzésképpen ugyanezeket az értékeket az eredeti mintából is kiszámítottuk, és jó egyezést kaptunk a szimulált értékekkel:

$p\%$	Százalékos valószínűség
0.01	4.520
0.025	0.548
0.05	0.034

Az illesztett kopulából való szimulálás azért is előnyös, mert így akár konfidenciaintervallumokat is meghatározhatunk a számunkra érdekes valószínűségekre.

6. Konklúzió

Az egymásba ágyazott arkhimédészi kopulacsalád egy rendkívül rugalmas struktúra, mely sokkal pontosabb modelleket tesz lehetővé, mint az egyparaméteres, többdimenziós arkhimédészi kopulák. Sajnos az általunk vizsgált adathalmazon kifejezetten gyengén teljesített, de aszimmetrikusabb adathalmazokon feltehetően lényegesen több felhasználása lehet.

Hivatkozások

- [1] Alexander J. McNeil (2008): *Sampling nested Archimedean copulas*, Journal of Statistical Computation and Simulation, 78:6, 567-581
- [2] David Ziener (2021): *Archimedean Copulas*,
https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/Seminar__Copulas_and_Applications_WS2021/David_Ziener_-_Notes__Archimedean_Copulas.pdf
- [3] Roger B. Nelsen (2008): *An Introduction to Copulas*, Springer New York, NY
- [4] Paul Embrechts, Filip Lindskog and Alexander McNeil (2001): *Modelling Dependence with Copulas and Applications to Risk Management*, Department of Mathematics, ETHZ
- [5] Claudia Czado (2019): *Analyzing Dependent Data with Vine Copulas, A Practical Guide With R*, Lecture Notes in Statistics, Springer Cham
- [6] Marshall, A.W. and Olkin, I. (1988): *Families of multivariate distributions*, Journal of American Statistical Association, 83, 834-841.
- [7] Joe, H., (1997): *Multivariate Models and Dependence Concepts*, Monogra. Stat. Appl. Probab. 73, London: Chapman and Hall.
- [8] Hofert M, Kojadinovic I, Maechler M, Yan J (2024): *copula: Multivariate Dependence with Copulas* R package version 1.1-4, <https://CRAN.R-project.org/package=copula>
- [9] Marius Hofert, Martin Mächler: *Nested Archimedean Copulas Meet R — The nacopula Package*, <https://cran.r-project.org/web/packages/copula/vignettes/nacopula-pkg.pdf>
- [10] Okhrin O, Ristig A (2014): *Hierarchical Archimedean Copulae: The HAC Package*. Journal of Statistical Software, 58(4), 1-20.
- [11] Yahoo Finance, <https://finance.yahoo.com/>
- [12] A dolgozatban használt R kódok a következő linken érhetőek el: <https://github.com/Cashoi/szakdolgozat>

NYILATKOZAT

Név: Kasoly Benedek József

ELTE Természettudományi Kar, szak: matematika

NEPTUN azonosító: GRXGFE

Szakdolgozat címe:

Egymásba ágyazott arkhimédészi kopulák

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2025.01.01.

Kasoly Benedek

a hallgató aláírása