7. előadás, 2022. március 24.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Áringadozások előadás

Kopulák

- Az összefüggőségi struktúra univerzális megjelenítői
- Többdimenziós eloszlás egyenletes marginálisokkal, (Hoeffding, 1940) - az 1990-es években újra felfedezték és azóta széles körben alkalmazzák is.
- Tetszőleges d-dimenziós, folytonos F eloszlásfüggvényhez egyértelműen megadható olyan C_F kopula, melyre

$$F(x_1,...,x_d) = C_F(F_1(x_1),...,F_d(x_d)).$$

- Ha F folytonos, akkor egyértelmű a megoldás:

$$C_F(u_1,...,u_d) = F(F_1^{-1}(u_1),...,F_d^{-1}(u_d)).$$

Példák

- Független eset: $C(x,y) = xy$.
- Teljes összefüggőség (Frechet) $C(x,y) = min(x,y)$, $C(x,y) = max\{(x+y-1),0\}$
- Gauss-kopula

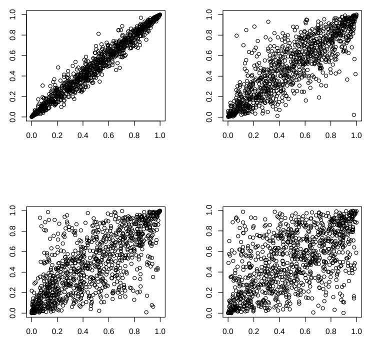
$$C_R(u) = \Phi_{R,d}\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right)$$

ahol $\Phi_{R,d}$ az R korrelációs mátrixú, d-dimenziós normális eloszlás eloszlásfüggvénye.
- Rugalmasabb modell: t-kopula

$$C_{R,\nu}(u) = t_{R,\nu,d}\left(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d)\right)$$

ahol $t_{R,\nu,d}$ az R korrelációs mátrixú, ν szabadságfokú d-dimenziós t-eloszlás eloszlásfüggvénye.

Példák 2 dimenzióban különböző összefüggőségekkel



ábra: Szimulált kopulák

- Az azonosításukhoz nagy mintaelemszám kell (főleg 2-nél magasabb dimenzióban)
- Nagyon gyenge és nagyon erős összefüggőség esetén a kopula típusa nem számít

Sűrűségfüggvények

- Hasznos gyakorlati eszközök a kopula tulajdonságainak vizualizációjánál
- A Gauss kopulára:

$$c_R(u) = \frac{\varphi_{R,d}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)}{\prod_{i=1}^{d} \varphi\left(\Phi^{-1}(u_i)\right)}$$

- Hasonlóan számolható a t-kopulára is

Elliptikus eloszlások, kopuláik

- Sűrűségfüggvényük kontúrjai ellipszisek
- Példa: Gauss, t
- Azonos típusú elliptikus eloszlások konvolúciója ismét ugyanolyan típusú elliptikus eloszlás
- Az elliptikus kopulákra teljesül a radiális szimmetria: $C(u, v) = u + v - 1 + C(1 - u, 1 - v)$
- Éppen ez az, ami tipikusan nem áll fenn a portfóliók hozamára (a kiugró veszteségek tipikusan nagyobbak a kiugró nyereségeknél)

Elliptikusság tesztelése

- Standardizálás után gömbszimmetrikus kell, hogy legyen az eredmény
- $R = ||Y||$ és $S = Y/||Y||$ függetlenek, S egyenletes eloszlású
- Például χ^2 próba alkalmazható

Arkhimédeszi kopulák

- A kopula generátor függvénnyel adhatók meg: $\varphi(u) : [0, 1] \to [0, \infty]$, folytonos, konvex (2D-ben, általánosan: teljesen monoton) és szigorúan monoton csökkenő, $\varphi(1) = 0$.
- Ebből a d-dimenziós Arkhimédeszi kopula

$$C_\varphi(\underline{u}) = \varphi^{-1}\left(\sum_{i=1}^{d} \varphi(u_i)\right).$$

- Egyszerű a konstrukciójuk, de van hátrányuk is: csak egy (vagy néhány) paraméterük van.
- Az összes $s < d$ dimenziós peremeloszlásuk azonos

Példák Arkhimédeszi kopulára

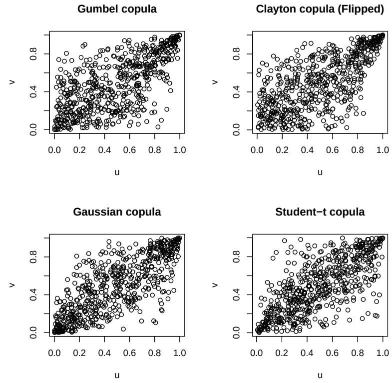
- A Gumbel kopula (logisztikus modell) generátora: $\varphi_\theta(u) = [-ln(u)]^\theta$, ahol $\theta \in [1, +\infty)$. Tehát a d-dimenziós Gumbel-kopula

$$C_{Gumbel}(\underline{u}) = e^{-(\sum_{i=1}^{d} -\log(u_i)^{\theta})^{\frac{1}{\theta}}}.$$

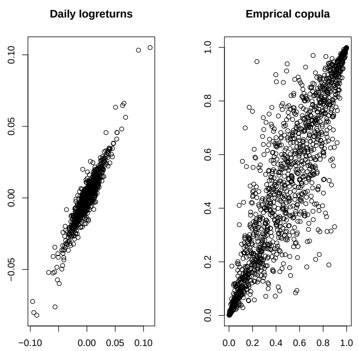
- Egy **C** kopula extrém-érték kopula, ha i $\mathbf{C}(u_1^t, ..., u_d^t) = \mathbf{C}^t(u_1, ..., u_d)$ minden $t > 0$. Ez megfelel a többdimenziós extrém-érték eloszlásoknak. Ezek közül a Gumbel kopula az egyetlen Arkhimédeszi kopula.
- A Clayton kopula generátora $\varphi_\theta(u) = u^{-\theta} - 1$, ahol $\theta > 0$. Tehát a d-dimenziós Clayton kopula:

$$C_{Clayton}(\underline{u}) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta}}.$$

Kopulák összehasonlítása



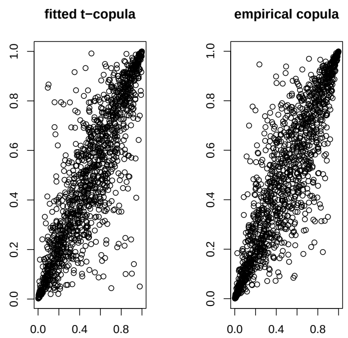
A log-hozamok transzformációja



- A kiugró értékek az eredeti ábrán láthatóak, az összefüggőség a kopulán

ábra: A Nasdaq-Dow Jones pár, 2005-2010 és az empirikus kopulája

t-kopula illesztése



- Szabadságfok: 4
- Ez tűnik a legjobbnak
- De hogyan teszteljük az illeszkedést? Rövidesen visszatérünk a kérdésre

ábra: A Nasdaq-Dow Jones párra illesztett t-kopula, 2005-2010

Extremális összefüggőség

- $$\chi = \lim_{u \to 1} P\{X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)\},$$
- Kvantilisfüggő változat:

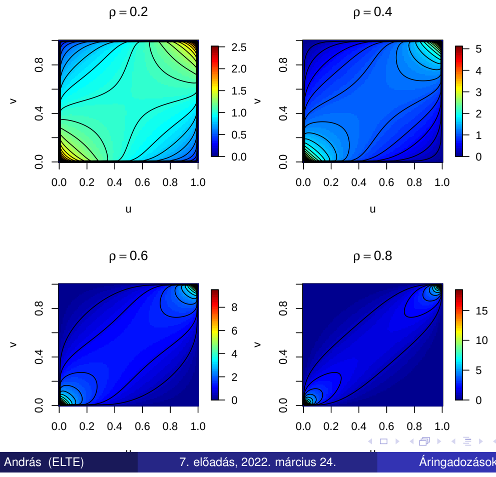
$$\chi(u) = 2 - \left(\frac{\log P\{X_1 < F_1^{-1}(u), X_2 < F_2^{-1}(u)\}}{\log P\{X_1 < F_1^{-1}(u)\}} \right), \quad 0 \leq u \leq 1.$$

- Nemelfajuló Gauss kopulára $\chi = 0$ (aszimptotikus függetlenség)
- A t-kopula aszimptotikusan összefüggő a negatív korreláció esetén is:

$$\chi = 2\bar{t}_{\nu+1} \left(\frac{\sqrt{\nu+1}\sqrt{1-R_{12}}}{\sqrt{1+R_{12}}} \right)$$

- A Gumbel kopulára $\chi = 2 - 2^{1/\theta}$
- Ezeket becslésre is lehet használni

Gauss kopulák sűrűségfüggvénye



Kopulák összefüggőségi indexei

- Lineáris korreláció: $R(X, Y) = \frac{E(X-EX)(Y-EY)}{D(X)D(Y)}$
- hátrányai:
  - érzékeny a kiugró értékekre
  - változik, ha transzformáljuk a marginálisokat
- Alternatívák: Kendall-τ:

$$\tau(X, Y) = P\left[(X - \tilde{X})(Y - \tilde{Y}) > 0\right] - P\left[(X - \tilde{X})(Y - \tilde{Y}) < 0\right].$$

Spearman-ρ:

$$\rho(X, Y) = 3\left(P\left[(X - \tilde{X})(Y - Y') > 0\right] - P\left[(X - \tilde{X})(Y - Y') < 0\right]\right).$$

ahol (X, Y), $(\tilde{X}, \tilde{Y})$, (X', Y') független, azonos eloszlásúak.

Tulajdonságok

- Ezek úgynevezett rangkorrelációk (csak az értékek sorrendje érdekes)
- Nem érzékenyek a kiugró értékekre
- Kiszámításuk a kopulával

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1$$

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] \, dudv.$$

További tulajdonságok

- Mindkettő invariáns a monoton transzformációkra. Legyen $\kappa = \rho$ vagy $\kappa = \tau$. Ekkor
  - $-1 \leq \kappa \leq 1$; $\kappa_{X,X} = 1$, $\kappa_{X,-X} = -1$.
  - Ha X és Y független, akkor $\kappa_{X,Y} = 0$.
  - $\kappa_{X,-Y} = \kappa_{-X,Y} = -\kappa_{X,Y}$.
- Az egyes kopulákra adódó összefüggőségi mérőszámok függnek a paramétertől, így becslésükből egyúttal a kopula becslése is megkapható. Például a Gumbel kopulára $\tau = 1 - 1/\beta$.

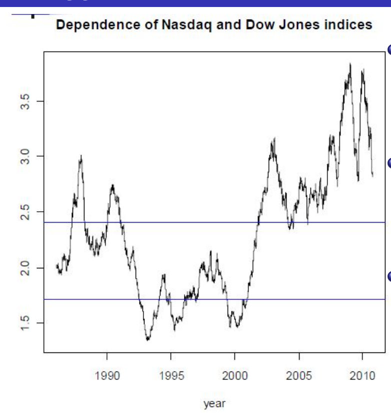
Alkalmazások

- A Gauss kopulára a páronkénti korrelációkra

$$R_{ij} = \sin\left(\pi\tau(X_i, X_j)/2\right)$$

- Lényeges a választás a különböző kopula-típusok között (pl. a farok-összefüggőség segítségével, illetve elméleti meggondolások alapján).
- Tapasztalati tény, hogy pl. a pénzügyi portfólióknál gyakran minden egyes elem extrém értékű (tőzsdekrach) - azaz itt várhatóan fellép a farok-összefüggőség.
- A különböző modellekből nagyon nagy különbségek adódhatnak a valószínűségbecslésre.

Időfüggés



ábra: A becsült paraméterek idősora

- Az ábra 251 napos időablakok (1 év) alapján becsült összefüggőségi paraméter értékét mutatja
- A kék vonalak az első év paraméterének 0.003 és 0.997-bootstrap kvantilisét mutatják
- A pénzügyi válság növelte az összefüggőséget

Illeszkedésvizsgálat

- A számításigény csökkentéséhez a dimenziószámot is csökkenteni kell. A K-függvény:

$$K(\vartheta, t) = P(F(\underline{X}) < t) = P\left(C_\vartheta(F_1(X_1), \ldots, F_d(X_d)) < t\right)$$

- Arkhimédeszi kopulákra a kiszámítása

$$K(\vartheta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \left[\varphi_\vartheta(t)^i\right] f_i(\vartheta, t)$$

ahol

$$f_i(\vartheta, t) = \frac{d^i}{dx^i} \varphi_\vartheta(x)|_{x=\varphi_\vartheta(t)}.$$

- Ha nincs zárt alakja, szimulálni akkor is lehet

A K függvényen alapuló teszt

- Empirikus verzió:

$$K_n(t) = \frac{1}{n}\sum_{j=1}^{n} \chi(E_j < t) \quad t \in [0,1]$$

ahol

$$E_j = \frac{1}{n}\sum_{i=1}^{n} \chi\left(U_{j,1} < U_{i,1}, \dots, U_{j,d} < U_{i,d}\right)$$

Kendall folyamat $\kappa_n(t) = \sqrt{n}\left(K(\vartheta_n, t) - K_n(t)\right)$.

- Cramér-von Mises típusú statisztika:

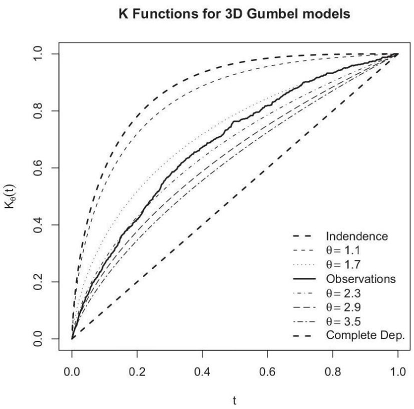
$$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t)dt$$

ahol Φ a súlyfüggvény

A teszt

- Formális tesztet is kaphatunk az S_n statisztikából (ha nagy, elutasítjuk az illeszkedést).
- Az aszimptotikus eloszlását csak ismert kopula esetén lehet kiszámítani.
- Azokban a realisztikus esetekben, ahol C-t becsüljük, szimulációval kaphatjuk meg a kritikus értékeket

Kopulák összehasonlítása



Rosenblatt-transzformáció

- Egy másik módszer: Breyman-teszt (Breymann et al, Berg & Bakken) a Rosenblatt transzformáción alapul $\mathcal{R} : (0,1)^d \times (0,1)^d$ $\mathcal{R}(\underline{u}) = (e_1, ..., e_d)$, ahol $e_1 = u_1$ és $i \geq 2$-re

$$e_i = \frac{\partial^{i-1}\mathbf{C}(u_1, ..., u_i, 1, 1, ...1)}{\partial u_1 ... \partial u_{i-1}} \Big/ \frac{\partial^{i-1}\mathbf{C}(u_1, ..., u_{i-1}, 1, 1, ...1)}{\partial u_1 ... \partial u_{i-1}}.$$

- Tulajdonsága: U eloszlása pontosan akkor a C kopula, ha $R(U)$ a független kopula.

Breymann-teszt: függetlenségvizsgálat

- $Y_B = \sum_{i=1}^{d} \Phi^{-1}(E_i)^2$ éppen chi-négyzet eloszlású, d szabadságfokkal.
- Ha ezt a saját eloszlásfüggvényébe helyettesítjük, egyenletes eloszlást kapunk. Ezt tesztelhetjük például az Anderson-Darling próbával.
- Berg és Bakken továbbfejlesztette a módszert, konzisztenssé téve azt.

Többdimenziós modellek tesztelése

- A tesztstatisztikák eloszlása nem ismert, ezért bootstrap szimuláció alapján határozhatók meg a kritikus értékek
- De: minden bootstrap mintára is illeszteni kell a modellt, ami magas dimenzióban igen lassú - ezért ez gyakorlatilag kivitelezhetetlen
- Az empirikus kopula és az illesztett paraméteres modell eltérése a természetes statisztika. Ennek határeloszlása

$$\sqrt{n}(C_n - C_{\vartheta_n}) = \sqrt{n}(C_n - C_\vartheta + C_\vartheta - C_{\vartheta_n}) \to \mathbf{C}_\vartheta - \Theta C_\vartheta$$

- A súlyozott bootstrap mintára vonatkozó határeloszlás tétel révén ez közelíthető anélkül, hogy mindig becsülni kellene a paramétert.

A teszt lépései

- C_n kiszámítása és a ϑ megfelelő tulajdonságú becslésének meghatározása
- A Cramer- von Mises statisztika kiszámítása:

$$\int_{[0,1]} \left(C_n(u,v) - C_{\vartheta_n(u,v)}\right)^2 dC_n(u,v) = \sum_{i=1}^{n} \left(C_n(U_{i,n}, V_{i,n}) - C_{\vartheta_n}(U_{i,n}, V_{i,n})\right)^2$$

- A súlyozott boostrap statisztikák kiszámítása
- Ebből a kritikus érték (ill. a p-érték) becsülhető
- Az eljárás gyorsabb, mint a paraméteres bootstrap

Gyakorlati tapasztalatok

- A módszer ereje függ a paraméter becslés módjától: a maximum-pszeudo likelihood általában jó eredményt ad
- 3-5 dimenzióban meglehetősen gyorsan kiszámolható
- A copula csomagban már benne van

Modellek

- A struktúrák nemparametrikusak, de a becslésekhez modellekre van szükség
- Paramétebecslés: pl. maximum likelihood módszerrel
- A paraméterszám (többdimenzióban):
  - ha túl alacsony (pl. 1 paraméterünk van egy tipikus Arkhimédeszi kopulánál), általában nem túl jó az illeszkedés
  - Ha túl sok (pl. minden párra külön paraméter a Gauss kopulánál) akkor a becslések nem lesznek megbízhatóak

Hivatkozások

- Berg, D. (2009) Copula Goodness-of-fit testing: An overview and power comparison.
- Berg, D. and Bakken, H. (2006) Copula Goodness-of-fit Tests: A Comparative Study.
- Embrechts, P., Lindskog, P. F. and McNeil, A.: Modelling Dependence with Copulas and Applications to Risk Management, 2001.
- Nelsen, R.B. (2006) An Introduction to Copulas. 2nd ed. John Wiley & Sons.
- Kojadinovic,I., Yan,J. and Holmes,M.: Fast large-sample goodness-of-fit tests for copulas. 2011.