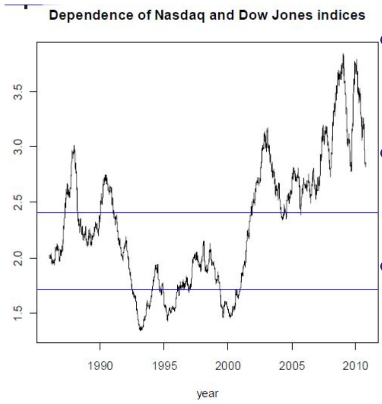
8. előadás, 2022. március 31.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Áringadozások előadás

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

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

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

- Cramér-von Mises típusú statisztika:

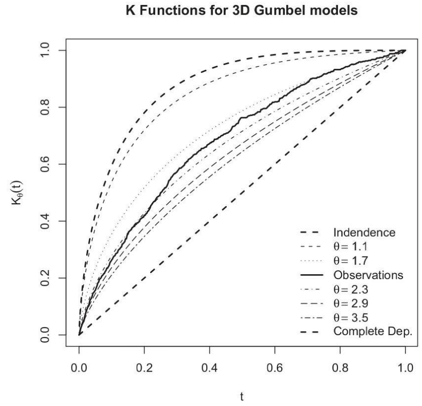
$$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t) dt$$

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

Többdimenziós modellek illesztése

- Tanultunk
  - Többdimenziós stabilis eloszlásokról
  - Többdimenziós extrém-érték eloszlásokról
  - Többdimenziós Pareto eloszlásokról
  - Kopulákról
- Tehát úgy tűnik, mintha csak választani kellene a modellek közül
- De bootstrap szimulációkra van szükség a kritikus értékek meghatározásához az illeszkedésvizsgálati teszteknél
- És a modelleket minden bootstrap mintára illeszteni kell, ami nagyon lassú magas dimenzióban
- Látni fogunk megoldás-ötleteket

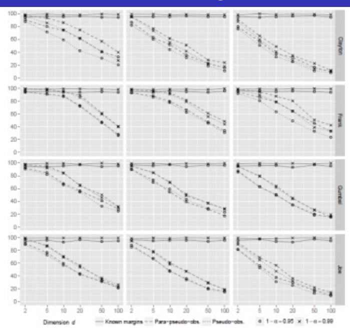
Modellek

- 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

Gyakorlati tapasztalatok magas dimenzióban

- A becslések többnyire a pszeudo-likelihoodon alapulnak (ekkor a peremeket nemparaméteresen becsüljük)
- Ezek sem nem függetlenek sem nem tekinthetőek egyenletes eloszlású mintának
- A becslések hibája (egyparaméteres Arkhimédeszi kopulákra) csökken, ha a dimenzió nő (legalábbis az ismert peremeloszlású esetben, máskor inkább konstans)
- A következő oldalon a likelihood-hányadoson alapuló konfidencia intervallumokat vizsgáljuk

Konfidencia intervallumok vizsgálata



- A kopula paraméter lefedési valószínűsége a dimenzió függvényében.
- Ismert peremeloszlásokra rendben van, de ismeretlen peremek esetén magas dimenzióra nagyon lecsökken

Még magasabbra

- Az eddig látott módszerek a gyakorlatban 2-4 dimenzióban alkalmazhatóak
- Magasabb dimenzióban a fő probléma a megbízhatóelemzéshez szükséges adatmennyiség hiánya (exponenciálisan kellene nőnie a dimenzióval)
- Egyszerűsítésre van szükség:
  - Ritkasági feltételek (Lasso és verziói), de ez nem reális a gyakorlati alkalmazásoknál (pl. kovariancia mátrix-becslés)
  - Dimenzió redukció

LASSO

- LASSO=least absolute shrinkage and selection operator
- L_1-regularizáció: az együttható vektor L_1 normájára vonatkozó feltétel mellett optimalizálunk
- Tipikus megvalósítás: hozzáadjuk a $\lambda \sum_{i=1}^{N} |\vartheta_i|$ függvényt a minimalizálandó célfüggvényhez (például a legkisebb négyzetek módszerénél)
- λ határozza meg a regularizáció erősségét - nem triviális a megválasztása
- Alkalmazás: portfólió optimalizálás (tipikusan kevésbé változékony súlyokat kapunk)

Becslési módszer magas dimenzióban: páronkénti likelihood

- Definíció:
$$\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} f_2(x_i, x_j; \vartheta)$$
- Tehát csak a páronkénti összefüggés szerepel a modellben
Egyszerűbb számolni (lényeges, ha igazán magas a dimenzió)
Példa: térbeli modell (pontfolyamat)
- $$\hat{\vartheta} = \operatorname{argmax} \sum_{t=1}^{T-\max k} \sum_{h \in K_t} \sum_{s_1=1}^{s} \sum_{s_2=1}^{s} \log f_2(z_{s_1,t}, z_{s_2,t+h}; \vartheta)$$
ahol z a megfigyelt érték, s a helyek, t az idő, K pedig adott indexhalmaz (például 0,1,2,4,8,..)

Kopulák magas dimenzióban: vine kopulák

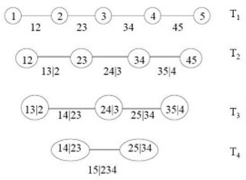
- Sokdimenzióban is használható struktúrák
- Kétdimenziós kopulákon alapulnak, a további struktúrát gráf határozza meg
- A rendelkezésre álló mintaelemszám/struktúra függvényében rugalmasan konfigurálhatóak
- Kopula sűrűségfüggvény: $c_{12}(x, y) = \frac{\partial^2 C(x,y)}{\partial x \partial y}$
- Ebből az eredeti eloszlás sűrűségfüggvénye: $c_{12}(F_1(x), F_2(y)) f_1(x) f_2(y)$
- Feltételes sűrűségfüggvény: $f(x|y) = c_{12}(F_1(x), F_2(y)) f_1(x)$

A pár-kopulákon alapuló konstrukció

- 3 dimenzióban:
$$\begin{aligned} f(x_1, x_2, x_3) &= f_{1|23}(x_1|x_2, x_3) f_{2|3}(x_2|x_3) f_3(x_3) \\ &= c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) c_{13}(F(x_1), F(x_3)) \\ &\times c_{23}(F(x_2), F(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned}$$
- A felbontás nem egyértelmű;
- 3 dimenzióban 3 felbontás van;
- De 5 dimenzióban már 240;
- Egyszerűsítés: elhagyjuk a feltételes kopuláktól való függést a feltételben:
$$\begin{aligned} f(x_1, x_2, x_3) &= f_{1|23}(x_1|x_2, x_3) f_{2|3}(x_2|x_3) f_3(x_3) \\ &= c_{12|3}(F_1(x_1), F_2(x_2)) c_{13}(F(x_1), F(x_3)) \\ &\times c_{23}(F(x_2), F(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned}$$

Reguláris vine kopulák

- $d-1$ fagráffal adhatók meg
- T_1 egy fa az $1, \dots, d$ csúcsokon
- A T_j fának $d+1-j$ csúcsa és $d-j$ éle van
- A T_j élei csúcsok lesznek a T_{j+1}-ben
- A T_{j+1} két csúcsa között csak akkor haladhat él, ha a T_j megfelelő éleinek van közös csúcsa



Vine kopulák típusai és gyakorlati alkalmazásuk

- C-vine: a gráfok csillag-alakúak
- D-vine: a gráfok utak
- Első lépés a kopula családok kiválasztása (illeszkedésvizsgálattal)
- Paraméterbecslés a gyakorlatban például a Kendall-féle τ alapján, vagy ML módszerrel: a legfontosabb párokat külön-külön becsüljük, a többit pedig együttesen (univerzálisan, ugyanazzal a kopulával - ez az úgynevezett egyszerűsítés)

Statisztikai módszerek

- Paraméterbecslés: iteratívan a gráf szintjeire:
  - Először az első szint kopuláira
  - Ezután ugyanezt elvégezhetjük a következő szintre (az adatokat transzformálva, azaz kiszámítva a feltételes eloszlásokat)
  - Az iteráció addig megy, amíg nem tudjuk a további szinteket egyszerűsíteni
- Milyenek legyenek a pár-kopulák? Tesztekkel vizsgálható az illeszkedés
- Adott vine-kopulából a kétdimenziós feltételes eloszlások segítségével lehet mintát szimulálni

Gyakorlati alkalmazások

- 16 dimenziós adatsorra kivitelezhető volt a teljes modell illesztése
- Az első lépésben azt a feszítőfát keressük, amelyre az éleken a Kendall-féle τ értékek összege maximális
- Levágás: egy bizonyos szint felett minden pár-kopulát függetlennek tételezünk fel
- Egyszerűsítés: egy bizonyos szint felett minden pár-kopulát azonosnak tekintünk

Megjegyzések

- Ha a kétdimenziós kopulák t-kopulák, akkor a vine-kopula a teljes d-dimenziós t-kopula részmodellje
- Az egymásba ágyazott modellek között a loglikelihood értékeken alapulhat a választás
- Ha a modellek nem egymásba ágyazottak, akkor alkalmazható az úgynevezett Vuong teszt statisztika, ami szintén a loglikelihood függvényen alapul és információelméleti háttere van (R csomag: CDVine)

További illeszkedésvizsgálati lehetőségek

- Információs mátrix-hányados tesztek
- White féle téves specifikációs teszt
- A korábban látott tesztek (K-függvényes, Rosenblatt transzformációs) általánosíthatóak – kritikus érték számításhoz a súlyozott bootstrap módszer kell; a Cramér-von Mises típusúak itt is az erősebbek

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-Hofer: Statistical inference for copulas in high dimension: a simulation study (2013)
- Varin-Reid-Firth: An overview of composite likelihood methods (2011)
- J. Dißmann, E. C. Brechmann, C. Czado, D. Kurowicka: Selecting and estimating regular vine copulae ... 2012.
- Schepsmeier: Efficient goodness-of-fit tests in multi-dimensional vine copula models (2013).
- Kojadinovic,I., Yan,J. and Holmes,M.: Fast large-sample goodness-of-fit tests for copulas. 2011.