4. előadás, 2022. március 3.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Áringadozások előadás

Többdimenziós extrém-érték eloszlások

- Tipikusan a koordinátánkénti maximum definiálja
- Megj.: ez nem biztos, hogy egyidejűleg fordul elő!
- A peremeloszlást tetszőlegesen megválaszthatjuk, a hagyományos a Frechet(1) - elérhető tetszőleges (ismert) F_j perem esetén: $Y_j = -1/\log(F_j(X_j))$

Többdimenziós extrém-érték eloszlások/2

- Legyenek $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ független, azonos eloszlású d-dimenziós valószínűségi változók.
- Ha vannak $\underline{a}_n, \underline{b}_n$ normáló vektorok, hogy

$$[\max(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n) - \underline{a}_n]/\underline{b}_n$$

  nemelfajuló határeloszláshoz közelít, akkor ez a határeloszlás szükségképpen d-dimenziós max-stabilis vagy úgynevezett extrém-érték eloszlás (MGEV).
- Max-stabilitás: minden n-hez van $\underline{a}, \underline{b}$, hogy $F^n(\underline{x}) = F(\underline{a}\underline{x} + \underline{b})$.

Reprezentáció (de Haan, 1984)

- Legyen $\|\underline{x}\| = |x_1| + \cdots + |x_d|$
- S_d a d-dimenziós egységszimplex: $\{\underline{x} \geq 0 : \|\underline{x}\| = 1\}$
- Létezik egy véges H mérték S_d-n, amire minden $j = 1, \ldots, d$-re
- H elnevezése: spektrálmérték
- Ezzel $G(\underline{x}) = \exp\{-V(\underline{x})\}$, ahol

$$V(\underline{x}) = \int_{S_d} \max_{1 \leq j \leq d} \frac{\omega_j}{x_j} dH(\underline{\omega})$$

Tulajdonságok

- Az MGEV eloszlások pozitív kvandránsösszefüggiek:

$$G(\underline{x}) \geq G_1(x_1)G_2(x_2)\dots G_d(x_d).$$

- Függetlenség esete:

$$G(\underline{x}) = G_1(x_1)G_2(x_2)\dots G_d(x_d).$$

  Ennek a G-nek a vonzási tartományába eső F eloszlásokat aszimptotikusan függetlennek nevezzük

- G spektrálmértéke a szimplex csúcsaiba helyez egységnyi mértékeket

Aszimptotikusan független eloszlások

- Tétel (Sibuya) Az F kétdimenziós vektor pontosan akkor aszimptotikusan független, ha $P(X_1 > q_1(u)|X_2 > q_2(u)) \to 0$, ha $u \to 1$ ($q_i(u)$ az i-edik marginális eloszlás u-kvantilise)
- Következmény: a többdimenziós normális eloszlás aszimptotikusan független, ha a páronkénti korrelációkra $\rho < 1$.

A kétdimenziós eset (Pickands)

A G pontosan akkor 2 dimenziós extrém-érték eloszlás Frechet(1) peremeloszlással, ha létezik $A : [0; 1] \to [0; 1]$ függvény, hogy

- $\max(t, 1 - t) \leq A(t) \leq 1$ minden $0 < t < 1$ -re
- $A(t)$ konvex és

$$G(x) = \exp - \left\{ \left(\frac{1}{x_1} + \frac{1}{x_2} \right) A \left(\frac{x_2}{x_1 + x_2} \right) \right\}.$$

- Elnevezés: A a (Pickands-féle) összefüggőségi függvény
- Megjegyzés: d dimenzióban a feltétel csak szükséges, de nem elégséges

Teljes összefüggőség

- G teljesen összefüggő, ha $G(x) = \min(G_1(x_1), G_2(x_2), \ldots, G_d(x_d))$.
- Ekkor H az S_d középpontjára: $(1/d, \ldots, 1/d)$ koncentrálódik
- $V(x) = \max(1/x_1, \ldots, 1/x_d)$ (minden $x > 0$-ra)
- $A(t) = \max(t, 1-t)$

Ez csak elméleti szempontból érdekes

Paraméteres családok

- Gumbel (logisztikus): $0 < \alpha \leq 1$ az összefüggőség erősségét méri:

$$G(\underline{x}) = \exp\left\{-\left(\sum_{i=1}^{d} x_i^{-1/\alpha}\right)^{\alpha}\right\}.$$

  - $\alpha = 1$: függetlenség
  - $\alpha \to 0$: teljes összefüggőség
- Aszimmetrikus logisztikus
- Negatív aszimmetrikus logisztikus
- ...

Becslések

- Nemparaméteres: a spektrálmérték, illetve az összefüggőségi függvény becslése az empirikus eloszlásból
  - Nem kell modell-feltevés
  - Nehezebbek a módszerek
- Paraméteres: paraméteres modellek illesztése, tipikusan maximum likelihood módszerrel
  - Csak akkor megbízható, ha jó az illeszkedés
  - Rutinszerű becslések, konfidencia intervallumok

Nemparaméteres becslés, $d = 2$

- Legyen $\underline{X}_i$ GEV eloszlású minta ($i = 1, \ldots, n$) G elo.fvel
- $Y_{i,j} = -\log G_j(X_{i,j})$ ($j = 1, 2$)
- $Z_i = \min(Y_{i,1}/t, Y_{i,2}/(1-t))$ exponenciális eloszlású $1/A(t)$ várható értékkel
- $A(t)$ becslése: $n/(Z_1 + \cdots + Z_n)$
- De ez nem feltétlenül konvex. Konvexitás pl. a legnagyobb konvex minoránssal érheti el (Hall - Tajvidi, 2000)
- Ekkor még a deriválhatóság kérdéses, tovább kell simítani

Másik módszer

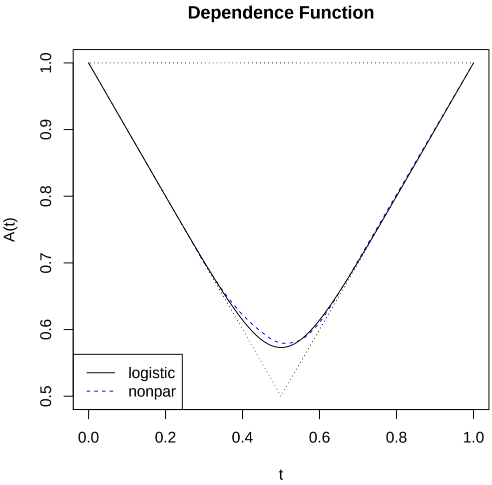
$$E\left[\log\left(\max(tY_{i,1},(1-t)Y_{i,2})\right)\right] = A(t) + \int_0^\infty log(u)du$$

- Amiből az $A(0) = A(1) = 1$ feltételt is kielégítő becslés:

$$\log(\hat{A}_c(t)) = \log(\hat{A}(t)) - t\log(\hat{A}(1)) - (1-t)\log(\hat{A}(0))$$

(Capéraá, Fougeres, Genest, 1997)

Példa: NASDAQ vs. DJIA



ábra: A különböző becslések összehasonlítása

Eredmények

- A GEV eloszlások jól illeszkedtek, ezt használtuk a marginálisok transzformációjánál.
- ML becslés működik a szimmetrikus logisztikus modell paramétereire.
- Ebből pl. kvantilis becslések is számolhatóak, de ez itt most egy görbét jelent!

Küszöb fölötti csúcsok módszere (POT)

- Azok az események extrémek, amelyek meghaladnak egy rögzített, magas küszöböt . Előnyei:
  - Több adatot lehet használni
  - A becsléseket nem befolyásolják a kicsi blokk-maximumok
- Hátrányai:
  - Függ a küszöb megválasztásától
  - Az idősorok esetén szükséges declusterezés (annak eldöntése, hogy mely maximumok származnak egy eseményből) nem mindig egyértelmű.

Elméleti alapok

- Legyenek $X_1, X_2, \ldots, X_n$ független, azonos eloszlású val. változók. Ha ennek a sorozatnak a normalizált maximuma konvergál egy extrém-érték eloszláshoz (μ, σ, ξ) paraméterekkel), akkor

$$P(X - u < y | X > u) \sim 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

  ha $y > 0$ és $1 + \frac{\xi y}{\tilde{\sigma}} > 0$, ahol $\tilde{\sigma} = \sigma + \xi(u - \mu)$. (Általánosított Pareto eloszlás, GPD), (ha $\xi \neq 0$, különben az exponenciális eloszlás).
- Az aszimptotika n és u végtelenhez tartása mellett érvényes.

Visszatérési szintek

- A p-kvantilis a POT modellből:

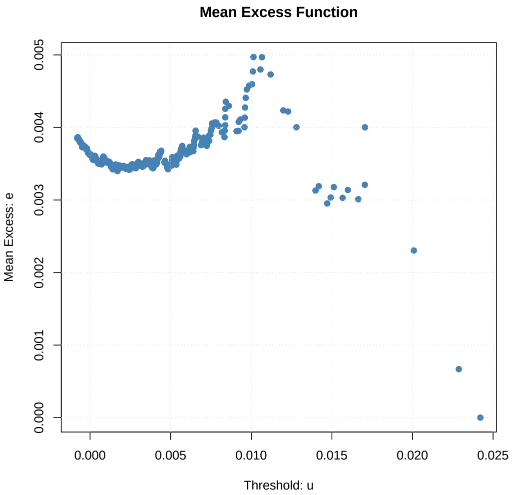
$$z_p = \begin{cases} u + \frac{\tilde{\sigma}}{\xi}\left[\left(\frac{p}{\eta}\right)^{-\xi} - 1\right], & \xi \neq 0 \\ \mu - \tilde{\sigma}\log(y_p), & \xi = 0 \end{cases}$$

  ahol $\eta = P(X > u)$, $\hat{\eta} = \frac{n_u}{n}$, ahol n_u az u-t meghaladó megfigyelések száma a mintában.
- Ez az az érték, amelyet átlagosan $1/p$ megfigyelésenként egyszer halad meg az adatsor.
- Ha n_y az évente észlelt szint feletti maximumok átlagos száma, akkor T évente visszatérő az $\frac{1}{T * n_y}$ kvantilis.
- Ha $\xi < 0$, akkor az eloszlás felső végpontja $u - \tilde{\sigma}/\xi$.

Küszöb választás

- Átlagos meghaladás ábrája: tetszőleges u küszöbre ábrázoljuk az $X - u$ átlagát (azokra a megfigyelésekre, amelyekre $X > u$) u függvényében. Ha a Pareto modell igaz, ez a görbe közel lineáris. A megmagyarázása nehéz lehet a megfigyelések maximumához közel megfigyelhető nagy ingadozása miatt.
- Alternatíva: tekintsük a paraméterbecslések értékeit különböző küszöbök esetén.

Példa: S&P 500, napi veszteség



S&P 500, napi veszteség

- Adatok 2002-2005 között
- Napi veszteség %-ban
- Küszöb: 1%

- MLE: $\xi = -0,31, \sigma = 0,62$
- 10 éves becsült visszatérési szint: 2,8
- A legnagyobb megfigyelt veszteség 2008.10.13: 9,05%