EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Többdimenziós extrém értékek modellezése

Szakdolgozat

Készítette: Rózsa Katalin
alkalmazott matematikus szak

Témavezető: Dr. Zempléni András egyetemi docens
Valószínűségelméleti és Statisztikai Tanszék

2001

Tartalomjegyzék

Előszó **2**

1. Bevezetés **4**

1.1. Egydimenziós extrém érték eloszlások 4

1.2. Többdimenziós extrém érték eloszlások 8

2. Az összefüggőség vizsgálata **11**

2.1. Az együttes túlélési függvény 11

2.2. Az összefüggőségi függvény 17

3. Statisztikai elemzés **19**

3.1. Az összefüggőségi paraméterek és azok becslése 19

3.2. Extrém érték eloszlás modellek 24

3.2.1. Küszöbérték modell 24

3.2.2. Komponensenkénti blokk-maximumok modellezése . . . 27

4. Alkalmazások **29**

Irodalomjegyzék **34**

Előszó

A dolgozat többdimenziós extrém érték elmélettel foglalkozik, ami a matematikának még kevésbé kidolgozott, manapság is fejlődő területe. Fő tárgya a valamilyen értelemben extrém, kiemelkedően nagy vagy kicsi megfigyelések viselkedésének modellezése.

Az extrém értékek elemzésével először a hidrológusok foglalkoztak, azzal a céllal, hogy megállapítsák a folyóknak és a tengereknek azt a vízszintjét, amit T-évenként átlagosan egyszer túllépnek. Azóta más területeken is, pl. pénzügy, biztosítás, ökológia, meteorológia, fontos szerepe van az olyan extrémumok analízisének, mint a nagy biztosítasi összegek, jelentős áremelkedés vagy árcsökkenés a részvénypiacon, környezeti tényezők (ózon, por, szennyezőanyagok, stb.) mennyiségének extrém szintjei, nagy erejű szél különböző helyeken, hullámok magassága vihar idején, vagy nagy csapadékmennyiség. Főleg a környezettel kapcsolatos megfigyelések azok, amelyek többdimenziós jellegűek. Pl. kívancsiak lehetünk arra, hogy mi a valószínűsége annak, hogy erős szél lesz, sok csapadékkal, alacsony hőmérséklet mellett.

A dolgozat első részében röviden áttekintjük az egydimenziós extrém érték elmélettel kapcsolatos alapfogalmakat illetve tételeket. Látni fogjuk, hogy itt az extrém érték eloszlások egyetlen paraméteres eloszláscsaláddal jellemezhetők, ami a többdimenziós esetben sajnos nem igaz. Másik jelentős különbség a két elméletben az, hogy a többdimenziós extrém érték eloszlások modellezéséhez szükségünk van arra is, hogy a marginálisok közötti kapcsolatokat ismerjük. Pl. erős összefüggés állhat fenn, ha különböző folyamatokat ugyanazon meteorológiai körülmények között tanulmányozunk. Ezzel foglalkozunk a második részében. Itt a kétdimenziós esetre szorítkozunk, ugyanis magasabb dimenziókban egyre bonyolultabb a módszerek alkalmazása. Először bevezetjük az aszimptotikus összefüggőség fogalmát, majd ennek felhasználásával néhány modellt mutatunk be két valószínűségi változó közös túlélési függvényére. Ezek után ismertetjük az összefüggőségi függvényt és az ezzel szoros kapcsolatban álló két paramétert, amelyek segítségével az összefüggőség mértékét tudjuk meghatározni. Érdekes tulajdonságuk, hogy az egydimenziós elméletben kidolgozott módszerekkel becsülhetők.

A statisztikai elemzések során ezen kívül ismertetünk két módszert az extrémumok modellezésére, amelyek között az alapvető különbség az, hogy milyen adatokra illesztjük az eloszlást. Az egyik esetben a megfigyeléseket csoportokra osztjuk és ezek extrémumaival foglalkozunk, viszont ha már az eredeti megfigyelések is függetlenek, akkor pontosabb becslést illetve modellt kaphatunk Poisson pontfolyamat segítségével. És végül egy konkrét adatsor elemzésén mutatjuk be, hogyan lehet alkalmazni az itt leírtakat.

1. fejezet

Bevezetés

Mielőtt rátérnénk a többdimenziós eloszlások tárgyalására, mely a dolgozat fő témája, tekintsük át röviden az egydimenziós esetet. Mindez megtalálható a [4] könyvben. Az itt definiált fogalmakra a későbbiekben is szükségünk lesz.

1.1. Egydimenziós extrém érték eloszlások

Legyen $X_1, \ldots, X_n$ független, azonos eloszlású valószínűségi változó, nem elfajuló F eloszlásfüggvénnyel. Az extrém érték elmélet ezek maximumának:

$$M_n = \max(X_1, \ldots, X_n)$$

viselkedését vizsgálja. A megfelelően normált maximum határeloszlására vagyunk kíváncsiak, vagyis arra, hogy létezik-e egy H nem elfajuló valószínűségi változó, amelyre:

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H \tag{1.1}$$

valamely $c_n > 0$ és $d_n \in \mathbb{R}$ esetén (itt eloszlásbeli konvergenciáról beszélünk). A minimumokkal azért nem foglalkozunk, mert a

$$\min_{i \leqslant n} X_i = -\max_{i \leqslant n}(-X_i)$$

összefüggés miatt, ez visszavezethető az előző esetre. A határeloszlásokkal kapcsolatos tételekhez vezessük be a következő fogalmakat:

1. Definíció. *Egy nem elfajuló F eloszlást* ***max-stabilisnak*** *(max-stable) nevezünk, ha létezik olyan $c_n > 0$ és $d_n \in \mathbb{R}$, amelyre $M_n \stackrel{d}{=} c_n X + d_n$ minden $n \in \mathbb{N}$–re.*

2. Definíció. *Az X és Y valószínűségi változókat* ***azonos típusúaknak*** *nevezzük, ha létezik olyan $a \in \mathbb{R}$ és $b > 0$, amelyre*

$$X \stackrel{d}{=} bY + a.$$

Tehát a max-stabilis eloszlásokra teljesül, hogy minden n-re M_n és X azonos típusúak.
A maximumok határeloszlása egyértelműen meghatározott a típus erejéig. Vagyis ha megváltoztatjuk a normalizáló konstansokat, akkor a határeloszlás megváltozhat, de annak típusa nem. Pontosabban:

1. Tétel. *Legyen $F_n(x)$ eloszlásfüggvények sorozata, c_n, $d_n > 0$, e_n és $f_n > 0$ pedig olyan valós számsorozatok, amelyekre:*

$$\lim_{n\to\infty} F_n(c_n + d_n\, x) = G(x)\,, \qquad \lim_{n\to\infty} (F_n(e_n + f_n\, x) = H(x)$$

a határértékek minden folytonossági pontjára, ahol a $G(x)$ és a $H(x)$ nem elfajuló eloszlásfüggvények. Ekkor a

$$\lim_{n\to\infty} \frac{f_n}{d_n} = B \neq 0\,, \qquad \lim_{n\to\infty} \frac{e_n - c_n}{d_n} = A$$

határértékek végesek, és $H(x) = G(A + Bx)$.

A normalizált maximumok lehetséges határeloszlásait a következő tételek karakterizálják:

2. Tétel. *Egy eloszlás akkor és csak akkor áll elő független, azonos eloszlású valószínűségi változók megfelelően normált maximumának nem elfajuló határeloszlásaként, ha max-stabilis.*

3. Tétel. ***(Fisher-Tippet)*** *Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók sorozata. Ha léteznek normalizáló konstansok és egy (nem elfajuló) H, amelyre (1.1) fennáll, akkor H a következő eloszlástípusok egyikébe tartozik:*

$$\textit{Frèchet} : \Phi_\alpha(x) = \begin{cases} 0 & : x \leqslant 0 \\ \exp(-x^{-\alpha}) & : x > 0 \end{cases} \qquad \alpha > 0$$

$$\textit{Weilbull} : \Psi_\alpha(x) = \begin{cases} \exp(-(-x^{\alpha})) & : x \leqslant 0 \\ 1 & : x > 0 \end{cases} \qquad \alpha > 0$$

$$\textit{Gumbel} : \Lambda(x) = \exp(-(\exp x)) \qquad x \in \mathbb{R}$$

Ezek a **standard extrém érték eloszlások** illetve velük azonos típusúak az extrém érték eloszlások. Közülük a „standard" ($\alpha = 1$–hez tartozó) Frèchet eloszlást használjuk a dolgozat további fejezeteiben.
A statisztikai vizsgálatokat megkönnyíti, hogy ezeket egyetlen paraméteres eloszláscsaláddal tudjuk reprezentálni, vagyis a tesztek során nem kell előre eldönteni, hogy melyik típust használjuk.

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}) & : \xi \neq 0 \\ \exp(-\exp(-x)) & : \xi = 0 \end{cases} \qquad \text{, ahol } 1 + \xi x > 0\,.$$

Ha egy F eloszlásfüggvényre (1.1) fennáll, vagyis ha léteznek olyan $c_n > 0$ és $d_n \in \mathbb{R}$ konstansok, amelyre

$$P(c_n^{-1}(M_n - d_n) < x) \longrightarrow H_\xi(x) \qquad (n \to \infty) \qquad \forall x \in \mathbb{R}$$

akkor azt mondjuk, hogy F a H_ξ általánosított extémérték eloszlás **max-vonzási tartományába** (maximum domain of attraction) tartozik (jelölés: $F \in MDA(H_\xi)$). Ezekre az eloszlásokra az is igaz, hogy egy bizonyos magas küszöbérték felett általánosított Pareto eloszlással becsülhetők, ahol:

3. Definíció. *A ξ paraméterű **általánosított Pareto eloszlás**:*

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & : \xi \neq 0 \\ 1 - \exp(-x) & : \xi = 0 \end{cases} \qquad \textit{, ahol } 1 + \xi x > 0$$

az

$$\begin{cases} x \geqslant 0 & ,ha\ \xi \geqslant 0 \\ 0 \leqslant x \leqslant -1/\xi & ,ha\ \xi < 0 \end{cases} \quad , \textit{ ahol } 1+\xi x > 0$$

tartományon.

Fontos szerepe van az x/β, $(\beta > 0)$ átparaméterezés után kapott függvénycsaládnak:

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi\,\frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi,\beta),$$

ahol:

$$D(\xi,\beta) = \begin{cases} [0,\infty) & : ha\ \xi \geqslant 0, \\ [0,-\beta/\xi] & : ha\ \xi < 0. \end{cases}$$

Érvényes ugyanis a következő tétel:

4. Tétel. *Minden $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ akkor és csak akkor, ha*

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0,$$

valamilyen pozitív β függvényre (itt $x_F = \sup x : F(x) < 1$).

Vgyis, ha u elég nagy, akkor

$$\bar{F}_u(x) \approx \bar{G}_{\xi,\beta(u)}(u).$$

A ξ becslésére különböző módszerek állnak rendelkezésünkre, ezek közül a maximim-likelihood módszert ismertetjük.

Tegyük fel, hogy egy magas küszöbérték $Y_1, \ldots, Y_n$ túllépései $G_{\xi,\beta(u)}$ eloszlásúak. Ekkor a likelihood függvény:

$$L = \prod_{i=1}^{N} \frac{1}{\beta}\left(1 + \frac{\xi Y_i}{\beta}\right)^{-1/(\xi-1)}$$

A logaritmust véve a log-likelihood:

$$\lambda = -N\log\beta + \left(-\frac{1}{\xi-1}\right)\sum_{i=1}^{N}\log\left(1+\frac{\xi Y_i}{\beta}\right).$$

$\beta = \xi/\tau$ átparaméterezés után:

$$\lambda = -N\log(\xi/\tau) + \left(-\frac{1}{(\xi-1)}\right)\sum_{i=1}^{N}\log(1+\tau Y_i). \tag{1.2}$$

A (1.2)–t τ és ξ szerint deriválva:

$$\frac{\partial\lambda}{\partial\tau} = \frac{N}{\tau} + \left(-\frac{1}{\xi} - 1\right)\sum_{i=1}^{N}\frac{Y_i}{1+\tau Y_i},$$

$$\frac{\partial\lambda}{\partial\xi} = -\frac{N}{\xi} + \frac{1}{\xi^2}\sum_{i-1}^{N}\log(1+\tau Y_i).$$

A második egyenlet gyöke:

$$\xi = \xi(\tau) = \frac{1}{N}\sum_{i=1}^{N}\log(1+\tau Y_i).$$

Ha ezt behelyettesítjük az első egyenletbe, akkor már csak a τ az ismeretlen, amit numerikus módszerekkel kiszámolhatunk, majd ebből az eredeti paraméterelket is megkaphatjuk.

1.2. Többdimenziós extrém érték eloszlások

Legyen $\mathbf{X_1}, \mathbf{X_2}, \ldots$ k-dimenziós megfigyelések, F eloszlásfüggvénnyel. Az egydimenziós esettel ellentétben itt az extrémumok definíciója nem egyértelmű. Egyrészt extrémumnak tekinthető a maximum-vektor :
$\mathbf{M}_n = (M_{n,1}, \ldots, M_{n,k})$, ahol $M_{n,j} = \max(X_{1,j}, \ldots, X_{n,j})$, azonban előfordulhat, hogy ez nem is tartozik az eredeti megfigyelések közé. Emiatt sokszor a legalább egyik komponensükben extrém megfigyeléseket is felhasználják a statisztikai elemzések során. A korábbiakhoz hasonlóan egy $G(\mathbf{x})$ eloszlásfüggvényt többdimenziós extrém érték eloszlásnak nevezünk, ha előáll normalizált maximumok-vektorok határeloszlásaként, vagyis ha léteznek olyan $\mathbf{c}_n > 0$ és $\mathbf{d}_n$ vektorok, amelyekre $\mathbf{c}_n^{-1}(\mathbf{M}_n - \mathbf{d}_n) \xrightarrow{d} G$. Az egydimenziós esethez hasonlóan, most is érvényes:

5. Tétel. *G többdimenziós extrém érték eloszlás akkor, és csak akkor, ha minden $n \geqslant 1$–hez léteznek olyan $\mathbf{A}_n > 0$ és $\mathbf{B}_n$ vektorok, amelyekre*

$$G^n(\mathbf{x}) = G(\mathbf{A}_n + \mathbf{B}_n\mathbf{x}),$$

vagyis G max-stabilis.

Bizonyítás: Nyilvánvaló, hogy a max-stabilis eloszlások előállnak határeloszlásként, tehát nézzük a másik irányt! $G(\mathbf{x})$ legyen olyan, hogy $\lim_{n\to\infty} F^n(\mathbf{A}_n + \mathbf{B}_n\mathbf{x}) = G(\mathbf{x})$ valamilyen F eloszlásfüggvényre. Legyen $m > 1$ rögzített, ekkor

$$\lim_{n\to\infty} F^{nm}(\mathbf{A}_{nm} + \mathbf{B}_{nm}\mathbf{x}) = G(\mathbf{x})\,,$$

vagyis

$$\lim_{n\to\infty} F^{n}(\mathbf{A}_{nm} + \mathbf{B}_{nm}\mathbf{x}) = G^{1/m}(\mathbf{x})\,.$$

Az 1.Tétel többdimenziós megfelelője szerint:

$$\lim_{n\to\infty} \frac{A_{nm,j} - A_{n,j}}{B_{n}, j} = C_{n,j}$$

és

$$\lim_{n\to\infty} \frac{B_{nm,j}}{B_{n}, j} = D_{n,j}$$

létezik és véges. Továbbá $G^m(\mathbf{C}_m + \mathbf{D}_m\mathbf{x}) = G(\mathbf{x})$, tehát $G(\mathbf{x})$ max-stabilis.

■

Láttuk, hogy egydimenziós esetben egyszerűen tudtuk jellemezni az extrém érték eloszlásokat. Viszont most csak az mondható el, hogy ha F Frèchet eloszlású marginálisokkal rendelkezik és egy G többdimenziós extrém érték eloszlás vonzási tartományához tartozik, akkor:

$$P(M_{n,1}/n \leqslant x_1, \ldots, M_{n,k}/n \leqslant x_k) = F^n(nx_1, \ldots, nx_k) \to G(x_1, \ldots, x_k),$$

$n \to \infty$, ahol $G(x_1, \ldots, x_k) = \exp(-V(x_1, \ldots, x_k))$ és

$$V(x_1, \ldots, x_k) = \int_{S_k} \max_{j=1,\ldots,k} \left(\frac{w_j}{x_j}\right) \mathrm{d}H(w).$$

H egy véletlen, pozitív mérték az

$$S_k = \{(w_1, \dots, w_k) : \sum_{j=1}^{k} w_j = 1, w_j \geqslant 0, j = 1, \dots, k\}$$

egységszimplexen, ami meghatározza az összefüggőséget, és amelyre az egyetlen megkötés:

$$\int_{S_k} w_j \mathrm{d}H(w) = 1 \qquad j = 1, \dots, k.$$

Ezzel és a Poisson folyamat reprezentációval részletesen foglalkozunk a későbbiekben.

2. fejezet

Az összefüggőség vizsgálata

Ahhoz, hogy egy extémérték eloszlást modellezni tudjunk, független, azonos eloszlású megfigyelések alapján, egyrészt ismernünk kell a peremeloszlásokat, másrészt pedig a komponensek összefüggőségi struktúráját. Látni fogjuk, hogy különös jelentőségű az az eset, amikor a marginális változók extrém értékei függetlenek.

2.1. Az együttes túlélési függvény

A továbbiakban az egyszerűség kedvéért a kétdimenziós esetre szorítkozunk. A módszerek kiterjeszthetők magasabb dimenziós változókra is, de egyre nagyobb nehézségekbe ütközünk. Elsőszöris vezessünk be két alapvető fogalmat. Legyen (X, Y) kétdimenziós valószínűségi vektorváltozó, azonos eloszlású marginálisokkal, ekkor az X–et Y-t aszimptotikusan függetlennek nevezzük, ha

$$\lim_{z \to z^*} P(X > z | Y > z) = 0, \tag{2.1}$$

ahol

$$z^* = \sup\{z \in \mathbb{R} : P(X \leqslant z) < 1\}.$$

Ha a határértékként kapott valószínűség egy nullánál nagyobb konstans, akkor az X és Y aszimptotikusan összefüggők.

Nem tételezhetjük fel, hogy ha a marginálisok függetlenek, akkor teljes a függetlenség egy küszöbérték felett, hanem figyelembe kell venni, hogy talán fokozatosan válnak függetlenné. Az alábbi példa szemlélteti, hogy szükség van arra is, hogy az aszimptotikus függetlenséget tanulmányozzuk. Tekintsük a kétdimenziós normális eloszlást, ahol a korrelációs együttható: $\rho < 1$. Megmutatható, hogy a Fréchet eloszlású változókra a túlélési függvény

$$P(X > z, Y > z) \sim C_\rho z^{-2/(1+\rho)} (\log z)^{-\rho/(1+\rho)} \tag{2.2}$$

$z \to \infty$ esetén, ahol

$$C_\rho = (1+\rho)^{3/2}(1-\rho)^{-1/2}(4\pi)^{-\rho/(1+\rho)}$$

$\rho < 1$–re. Ezzel ellentétben a tejesen független Frèchet változókra

$$P(X > z, Y > z) = (1 - \exp(-1/z))^2 \sim z^{-2}$$

$z \to \infty$ esetén. Akármilyen nagyra választjuk a z küszöbértéket, a két érték nem lesz egyenlő.

Az is megmutatható, hogy

$$P(X > z, Y > z) \sim \begin{cases} z^{-1} & \text{teljes összefüggőség esetén,} \\ z^{-2} & \text{teljes függetlenség esetén.} \end{cases}$$

Olyan modellt keresünk, ami folytonosan köti össze a két határesetet. A (2.2) mintájára felírható a ketdimenziós túlélési függvény:

$$P(X > z, Y > z) \equiv \bar{F}(z, z) \sim \mathcal{L}(z) z^{-1/\eta}, \tag{2.3}$$

ha $z \to \infty$, ahol $\frac{1}{2} \leqslant \eta \leqslant 1$ konstans, és $\mathcal{L}(z)$ lassú változású függvény, azaz $\mathcal{L}(tz)/\mathcal{L}(z) \to 1$, ha $z \to \infty$ és $t > 0$ rögzített. Az η paraméter a farokeloszlás összefüggőségét jellemzi, az $\mathcal{L}$ pedig ennek erősségét. Látható, hogy $\eta = \frac{1}{2}$ és $\mathcal{L}(z) = 1$ teljes függetlenséget jelent, $\eta = 1$ és $\mathcal{L}(z) = 1$ pedig teljes összefüggőséget. (2.3) alapján

$$P(X > z | Y > z) \sim \mathcal{L}(z) z^{1-1/\eta},$$

tehát (2.1) szerint X és Y aszimptotikusan függetlenek, ha $\eta < 1$, és $\eta = 1$ és $\mathcal{L}(z) \nrightarrow 0$, $(z \to \infty)$ esetén aszimptotikusan összefüggnek.
A (2.3) modell kiterjesztése:

$$\bar{F}(x,y) = \mathcal{L}_1(x,y)x^{-c_1}y^{-c_2} + \mathcal{L}_2(x,y)x^{-(c_1+d_1)}y^{-(c_2+d_2)} + \dots \tag{2.4}$$

ahol $c_1 + c_2 = 1/\eta, d_1, d_2 \geqslant 0, \mathcal{L}_1(x,y)$ és $\mathcal{L}_2(x,y)$ nem azonosan 0 kétdimenziós lassúváltozású függvény g_j határfüggvénnyel, és $\mathcal{L}_2(tx,ty) = o(\mathcal{L}_1(tx,ty))$, $(t \to \infty)$, ha $d_1 = d_2 = 0$. Így

$$g_j(x,y) = \lim_{t\to\infty} \left\{ \frac{\mathcal{L}_j(tx,ty)}{\mathcal{L}_j(t,t)} \right\},$$

és $g_j(cx,cy) = g_j(x,y)$, minden $c > 0$ és $(x,y) \in \mathbb{R}^2_+$ esetén. Tehát

$$g_j(x,y) \equiv g_*^{(j)}(x/(x+y)) = g_*^{(j)}(w),$$

ahol $w = x/(x+y) \in (0,1)$ és g_j konstans az $y = x(1-w)/w$ alakú sugarakon. A modellek négy osztályát (A-D) mutatjuk be a következő példákon, ahol a B, C és D osztályban aszimptotikus függetlenség figyelhető meg. $F(.)$ a Frèchet eloszlásfüggvény , és legyen $w = x/(x+y)$.

A) **Aszimptotikus összefüggőség** , $\eta = 1$ és $\mathcal{L}_1(z,z) \nrightarrow 0$, ha $z \to \infty$

1. *Kétdimenziós extrém érték eloszlás túlélési függvénye*
Az eloszlásfüggvény

$$F(x,y) = \exp(-V(x,y))$$

Nézzük meg, ekkor hogyan számolható ki $\mathcal{L}_1(x,y)$ és $\mathcal{L}_2(x,y)$! Ha az exponenciális tagokat másodrendben közelítjük, akkor:

$$\bar{F}(x,y) = 1 - F(x) - F(y) + F(x,y) = 1 - e^{-1/x} - e^{-1/y} - e^{-V(x,y)} \approx$$
$$\approx 1 - \Big(1 - \frac{1}{x} + \frac{1}{2x^2}\Big) - \Big(1 - \frac{1}{y} + \frac{1}{2y^2}\Big) + \Big(1 - V(x,y) + \frac{V(x,y)^2}{2}\Big) =$$
$$= \frac{1}{x} + \frac{1}{y} - V(x,y) - \frac{1}{2}\Big(\frac{1}{x^2} + \frac{1}{y^2} - V(x,y)^2\Big)$$

Az elsőrendű tagokból kaphatjuk meg $\mathcal{L}_1(x,y)$–t, a kifejezés második részéből pedig $\mathcal{L}_2(x,y)$–t. Az átalakítás során felhasználjuk, hogy $w = x/(x+y)$.

$$\frac{1}{x} + \frac{1}{y} - V(x,y) = \frac{x+y}{x\,y} - V(x,y) = \frac{x+y}{x\,y}\Big(1 - V\Big(\frac{x+y}{y}, \frac{x+y}{x}\Big)\Big) =$$
$$= \big((w(1-w))^{-1/2}\,(1 - V((1-w)^{-1}, w^{-1}))\big)\; x^{-1/2}\, y^{-1/2}$$

és

$$\frac{1}{x^2} + \frac{1}{y^2} - V(x,y)^2 = \frac{x^2+y^2}{x^2\,y^2} - \frac{(x+y)^2}{x^2\,y^2}\,(V((1-w)^{-1}, w^{-1}))^2 =$$
$$= \left(\frac{w^2 + (1-w)^2 - (V((1-w)^{-1}, w^{-1}))^2}{w\,(1-w)}\right)\, x^{-1}\, y^{-1}$$

Tehát, azt kaptuk, hogy:

$$\mathcal{L}_1(x,y) = (w(1-w))^{-1/2}\,(1 - V((1-w)^{-1}, w^{-1}))$$
$$\mathcal{L}_2(x,y) = -\frac{w^2 + (1-w)^2 - (V((1-w)^{-1}, w^{-1}))^2}{2w(1-w)}$$

és $c_j = d_j = \frac{1}{2}$, $j = 1, 2$–re.

A függvények a többi esetben is hasonlóképpen számíthatók ki.

A (2.3) modellhez visszatérve, látható, hogy $\mathcal{L}(z) = \mathcal{L}_1(z,z) = 2 - V(1,1)$. Például a későbbiekben fontos szerepet játszó logisztikus modellnél, amelyre $V(x,y) = (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}$, az $\mathcal{L}(z) = 2 - 2^{\alpha}$.

2. *Clayton-féle eloszlás*
Az eloszlásfüggvény

$$F(x,y) = F(x) + F(y) - 1 + (\bar{F}(x)^{-1/\alpha} + \bar{F}(y)^{-1/\alpha} - 1)^{-\alpha}$$

Ekkor

$$\mathcal{L}_1(x,y) = (w(1-w))^{1/2}\,(w^{-1/\alpha} + (1-w)^{-1/\alpha})^{-\alpha}$$

$$\mathcal{L}_2(x,y) = \begin{cases} -\dfrac{(\mathcal{L}_1(x,y))^{(\alpha+1)/\alpha}\,(w^{(\alpha-1)/\alpha} + (1-w)^{(\alpha-1)/\alpha})}{2(w(1-w))^{\alpha-1)/2\alpha}} & \text{, ha } \alpha < 1 \\ \\ -\dfrac{(w(1-w))^{1/2}}{12} & \text{, ha } \alpha = 1 \\ \\ \alpha(\mathcal{L}_1(x,y))^{(\alpha+1)/\alpha)} & \text{, ha } \alpha > 1 \end{cases}$$

és $c_j = d_j = \frac{1}{2}$, ha $\alpha \in (0,1)$, $d_j = 1$, ha $\alpha = 1$, $d_j = (2\alpha)^{-1}$, ha $\alpha > 1$, $j = 1, 2$–re.

B) **Pozitív összefüggés**, $\frac{1}{2} < \eta < 1$, vagy $\eta = 1$ és $\mathcal{L}_1(x,y) \to 0$, ha $z \to \infty$

1. *Kétdimenziós normális eloszlás* $\rho > 0$ *korrelációs együtthatóval*
Elég nagy x és y esetén, amikre igaz, hogy $x^{\rho^2} < y < x^{\rho^{-2}}$, akkor

$$\mathcal{L}_1(x,y) = \frac{(xy)^{\rho(2(\log x \log y)^{1/2}/(1-\rho^2)\log(xy)-1)}}{(1-\rho^2)^{-3/2}} \quad \times$$
$$\times \quad \frac{((\log x)^{1-\rho c_{21}}(\log y)^{1-\rho c_{12}}(4\pi)^{\rho(2\rho-(c_{12}+c_{21}))})^{1/2(1-\rho^2)}}{(1+\rho^2)(\log x \log y)^{1/2} - \rho\log(xy)},$$

ahol $c_{12} \equiv (\log x/\log y)^{1/2}$ illetve $c_{21} \equiv (\log y/\log x)^{1/2}$

$$\mathcal{L}_2(x,y)/\mathcal{L}_1(x,y) \sim O(1/log(\min(x,y))),$$

$c_j = (1+\rho)^{-1}$, tehát $\eta = (1+\rho)/2$, és $d_j = 0, j = 1, 2$.

2. *Kétdimenziós extrém érték eloszlás*
Az eloszlásfüggvény

$$F(x,y) = F(x) + F(y) - 1 + \exp(-V((-\log \bar{F}(x))^{-1}, (-\log \bar{F}(y))^{-1}))$$

$$\mathcal{L}_1(x,y) = 1$$

$$\mathcal{L}_2(x,y) = \frac{(w(1-w))^{-1/2}\,((1-w)V_1(1,1) + w\,V_2(1,1))}{2}$$

és $c_j = -V_j(1,1)$, $d_j = \frac{1}{2}$, $V_1 = \frac{\partial V(x,y)}{\partial x}$ és $V_2 = \frac{\partial V(x,y)}{\partial y}$ mellett. Tehát $\eta^{-1} = c_1 + c_2 = -(V_1(1,1) + V_2(1,1)) = V(1,1)$.

C) **Közel függetlenség** , $\eta = \frac{1}{2}$

1. *Morgenstern eloszlás túlélés függvénye*
Az eloszlásfüggvény

$$F(x,y) = F(x)\,F(y)\,(1 + \alpha\bar{F}(x)\,\bar{F}(y))$$

$$\mathcal{L}_1(x,y) = 1 + \alpha$$

$$\mathcal{L}_2(x,y) = \frac{-(w(1-w))^{-1/2}\,(3\alpha + 1)}{2}$$

Az $\alpha = 0$ függetlenséget jelent.

2. *Gumbel eloszlás túlélés függvénye*
Az eloszlásfüggvény

$$F(x,y) = F(x)\,F(y)\exp(-\theta/x\,y)$$

$$\mathcal{L}_1(x,y) = 1 - \theta$$

$$\mathcal{L}_2(x,y) = (w(1-w))^{-1/2}\,(\theta - \tfrac{1}{2})$$

Az $\theta = 0$ függetlenséget jelent.

D) **Negatív összefüggés**, $0 < \eta < \frac{1}{2}$

1. *Kétdimenziós normális eloszlás* $\rho < 0$ Itt c_j, d_j és $\mathcal{L}_j$, $j = 1,2$–re ugyanaz, mint $\rho > 0$ esetén , így $\eta = (1+\rho)/2$.

2.2. Az összefüggőségi függvény

Ezek után vezessük be az összefüggőségi függvény fogalmát, amely segítségével leírhatjuk az egyes valószínűségi vektorváltozók komponensei között fennálló kapcsolatokat.
Egy $(X_1, \ldots, X_k)$ valószínűségi vektor esetén $F(x_1, \ldots, x_k) = P(X_1 \leqslant x_1, \ldots, X_k \leqslant x_k)$ az együttes eloszlásfüggvény, F_{X_i} (i = 1,..., k) peremeloszlásokkal. Az $\mathcal{A} = [0,1]^k$ tartományon értelmezett C függvény, amelyre

$$F(x_1, \ldots, x_k) = C(F_{X_1}(x_1), \ldots, F_{X_k}(x_k)),$$

az $F(x_1, \ldots, x_k)$ összefüggőségi függvénye. Ez tartalmaz minden információt az marginálisok közötti összefüggőségről és ezek együttes eloszlásáról. C úgy is felfogható, mint az X_i–k együttes eloszlásfüggvénye, miután azokat egyenletes eloszlású U_i változókká transzformáltuk, ahol: $(U_1, \ldots, U_k) = (F_{X_1}(X_1), \ldots, F_{X_k}(X_k))$ Az összefüggőségi függvény fontos tulajdonsága, hogy invariáns a marginálisok monoton növő, folytonos transzformációira. Ez azért lényeges számunkra, mert így az összefüggőségi struktúra megváltoztatása nélkül tudunk más peremeloszlásokkal is dolgozni. Például a statisztikai elemzések során a Frèchet eloszlás a legalkalmasabb.

1. Állítás. *Ha $(X_1, \ldots, X_k)$ valószínűségi vektorváltozó összefüggőségi függvénye C, és $(T_1, \ldots, T_k)$ monoton növekvő, folytonos függvények, akkor $(T_1(X_1), \ldots, T_k(X_k))$ összefüggőségi függvénye szintén C*

Bizonyítás: Legyen $(U_1, \ldots, U_k)$ eloszlásfüggvénye C (folytonos F_{X_i} marginálisok esetén $U_i = F_{X_i}(X_i)$). Ekkor:

$$\begin{aligned}
&C(F_{T_1(X_1)}(x_1), \ldots, F_{T_k(X_k)}(x_k)) = \\
&\quad = P(U_1 \leqslant F_{T_1(X_1)}(x_1), \ldots, U_k \leqslant F_{T_k(X_k)}(x_k)) \\
&\quad = P(F^{-1}_{T_1(X_1)}(U_1) \leqslant x_1, \ldots, F^{-1}_{T_k(X_k)}(U_k) \leqslant x_k) \\
&\quad = P(T_1 \circ F^{-1}_{(X_1)}(U_1) \leqslant x_1, \ldots, T_k \circ F^{-1}_{(X_k)}(U_k) \leqslant x_k) \\
&\quad = P(T_1(X_1) \leqslant x_1, \ldots, T_2(X_k) \leqslant x_k)
\end{aligned}$$

■

Megjegyzés: A T_i transzfomációk folytonossága általános $(X_1, \dots, X_k)$ valószínűségi vektorváltozók esetén szükséges, mivel ekkor igaz, hogy $F^{-1}_{T_i(X_i)} = T_i \circ F^{-1}_{(X_i)}$. Viszont, ha a margináls eloszlások folytonosak, akkor elegendő a monoton növekedés is.
Bizonyos esetekben könnyen kiszámolható a függvény. Nézzünk néhány példát kétdimenziós (X,Y) esetén:

Függetlenség. Ekkor $F(x,y) = F_X(x)F_Y(y)$, tehát $C(u,v) = uv$ az $\mathcal{A}$–n.

Teljes összefüggőség. Ekkor $Y = F_Y^{-1}(F_X(X))$ egy valószínűséggel , tehát $F(x,y) = min(F_X(x), F_Y(y))$ és $C(u,v) = min(u,v)$ az $\mathcal{A}$–n.

Kétdimenziós logisztikus extrém érték eloszlás. Ebben az esetben F_X és F_Y általánosított extrém érték eloszlás,

$$C(u,v) = \exp(-((-\log u)^{1\backslash\alpha} + (-\log v)^{1\backslash\alpha})^\alpha)$$

az $\mathcal{A}$–n. Az $\alpha \in [0,1]$ paraméter az összefüggőség mértékét határozza meg. $\alpha = 1$ fűggetlenséget jelent, és amint csökken, egyre erősebb az összefüggés.

Gauss-féle összefüggőségi modell. Ha (X,Y) kétdimenziós normális eloszlású ρ korrelációs együtthatóval, akkor F_X és F_Y normális eloszlásfüggvények és

$$C(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}(s^2 - 2\rho st + r^2)\right\} \mathrm{d}s\,\mathrm{d}t$$

az $\mathcal{A}$–n, ahol $\Phi(.)$ egydimenziós standard normális eloszlásfüggvény.

3. fejezet

Statisztikai elemzés

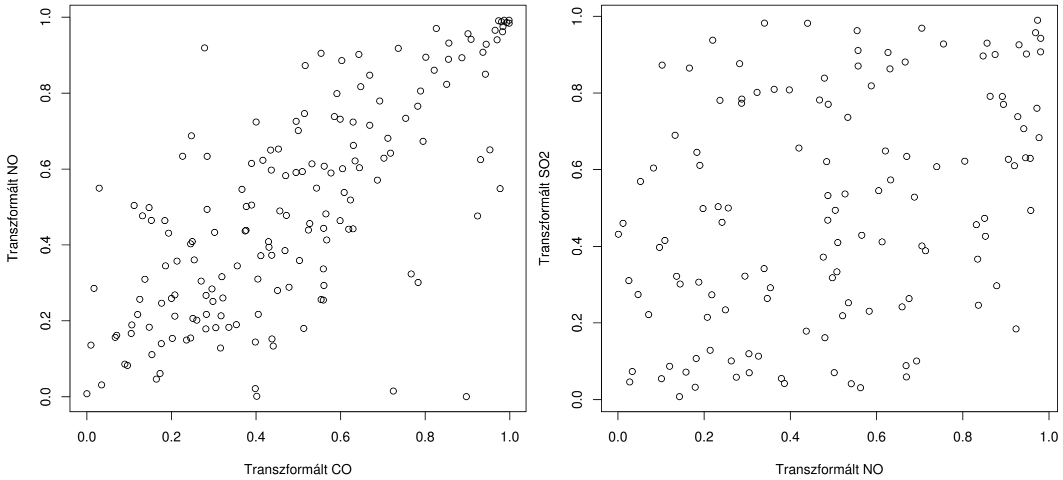
3.1. Az összefüggőségi paraméterek és azok becslése

A független, ismeretlen F eloszlású $(x_i, y_i), i = 1, \ldots, n$ megfigyelések esetén célszerű a marginalisokat egyenletes eloszlásúvá transzformálni, hogy az összefüggőségi függvény előzőkben leírt formáját alkalmazni tudjuk. Mivel F peremeloszlásai is ismeretlenek, használjuk az $\widehat{F}_X$ és $\widehat{F}_Y$ tapasztalati eloszlásfüggvényeket. Ekkor az (u_i, v_i) párok $i = 1, \ldots, n$, ahol $u_i = \widehat{F}_X(x_i)$ és $v_i = \widehat{F}_Y(y_i)$ függetlenek és megközelítően C eloszlásúak. Előzetes képet kaphatunk az extremális összefüggőségről, ha megvizsgáljuk a nagy u_i és v_i értékeket. A (3.1) ábrán erős összefüggés és a függetlenség közötti különbség jól megfigyelhető.
A következőkben bevezetünk két paramétert, amik különböző szemszögből mérik az extremális összefüggőséget.

A χ összefüggőségi mérték.

$$\chi = \lim_{u \to 1} P(V > u | U > u)$$

egy természetes mérték az extremális öszefüggőségre, ahol (U, V)-t a nem feltétlenül azonos eloszlású (X, Y) transzformációjával kapjuk. A számolást



CO és NO levegőben mért koncentrációinak heti maximumai, miután azokat egyenletes eloszlásúvá transzformáltuk. Itt összefüggőség figyelhető meg.

NO és SO2 levegőben mért koncentrációinak heti maximumai, miután azokat egyenletes eloszlásúvá transzformáltuk. Az ábra függgetlenségre utal.

3.1. ábra.

megkönnyíti, ha a jobboldalt a következő alakba írjuk át:

$$\begin{aligned} P(V > u|U > u) &= \frac{P(U > u, V > u)}{P(U > u)} \\ &= \frac{1 - 2u + C(u,u)}{1-u} \\ &= 2 - \frac{1 - C(u,u)}{1-u} \\ &\sim 2 - \frac{\log C(u,u)}{\log u} \end{aligned}$$

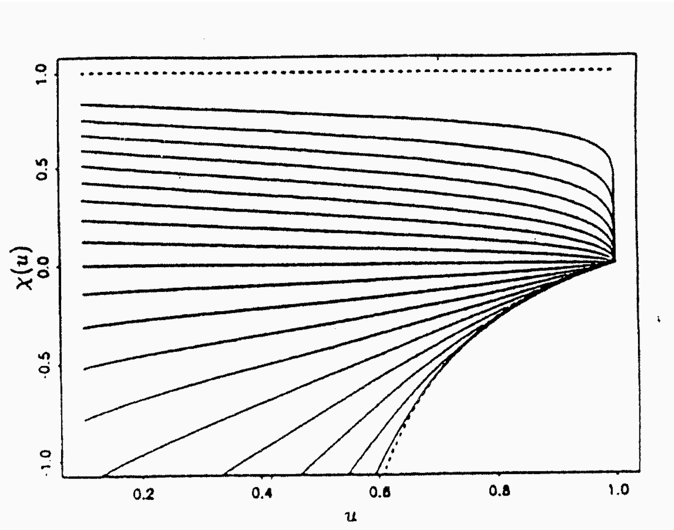
ha $u \to 1$. Legyen

$$\chi(u) = 2 - \frac{\log P(U < u, V < u)}{\log P(U < u)} \qquad 0 \leqslant u \leqslant 1, \tag{3.1}$$

ebből következik, hogy

$$\chi = \lim_{u \to 1} \chi(u).$$

A már említett modellekben nézzük meg a $\chi(u)$ értékét! Független változókra $\chi(u) = 0$, teljes összefüggőség esetén $\chi(u) = 1$, a kétváltozós logisztikus extrém érték eloszlásra $\chi(u) = 2 - 2^{\alpha}$. Mindhárom esetben $\chi(u)$ konstans u-ban. Minden kétdimenziós extrém érték eloszlasnál ugyanez a helyzet, tehát abból, hogy $\chi(u)$ tapasztalati becslése mennyire állandó, következtethetünk arra, hogy a kétdimenziós extémérték eloszlások osztályába tartozik-e.



A $\chi(u)$ összefüggőségi mérték a Gauss-féle összefüggőségi modell esetén. A görbék (lentről felfelé) a $\rho = -0.9, -0.8, \ldots, 0.9$ értékekhez tartoznak.

3.2. ábra.

Általában $\chi(u)$ u-nak bonyolultabb függvénye. Például a Gauss-féle összefüggőségi modell esetén is, ahol a ρ növelésével egyre erősebb az összefüggés. A (3.2) ábrán látható, hogy u közepes nagyságú értékeire $\chi(u)$ lineárisnak tekinthető, de az 1-hez közeli értékekre $\chi(u) \to 0$, ami aszimptotikus függetlenségre utal. $\rho > 0$ estén a konvergencia nagyon lassú, a végén viszont egyre meredekebb a függvény, tehát 1-hez közeli u-ra is lehet $\chi(u)$ 0-nál nagyobb. Mivel a tapasztalati becslést olyan megfigyelések alapán adjuk, amire $u < 1$, ezért előfordulhat, hogy $\chi(u)$ közel konstans és pozitív még aszimptotikusan független változók esetén is.

A $\bar{\chi}$ összefüggőségi mérték. A többdimenziós extrém értékek modellezésében nagy szerepet játszanak az aszimptotikusan független eloszlások is, amikre $\chi = 0$, vagyis ez nem mond semmit az összefüggőség mértékéről. Ezért van szükség még egy mértékre. A $P(X > x, Y > y)$ együttes túlélési függvényt $\bar{F}(x, y)$–nal jelölve:

$$\bar{F}(x, y) = 1 - F_X(x) - F_Y(y) + F(x, y) = \bar{C}(F_X(x), F_Y(y)), \qquad (3.2)$$

ahol

$$\bar{C}(u, v) = 1 - u - v + C(u, v).$$

Az előző számításokhoz hasonlóan, legyen:

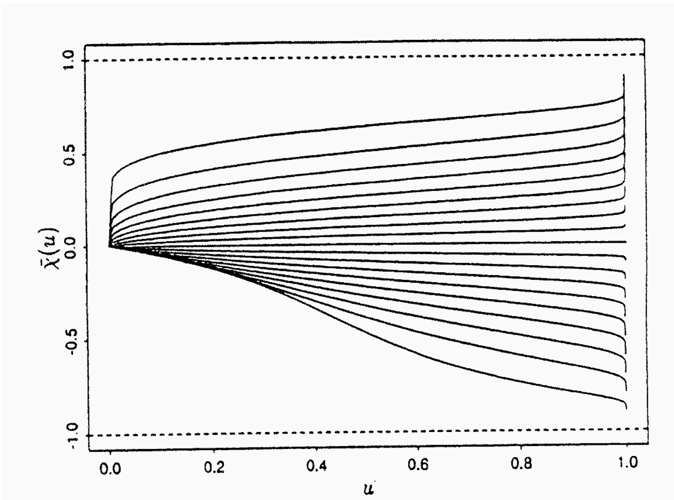
$$\bar{\chi}(u) = \frac{2\log P(U > u)}{\log P(U > u, V > u)} - 1 = \frac{2\log(1-u)}{\log \bar{C}(u,u)} - 1 \qquad 0 \leqslant u \leqslant 1, \quad (3.3)$$

és

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u).$$

Aszimptotikusan összefüggő változókra $\bar{\chi} = 1$, ez a teljes összefüggőség és a kétdimenziós logisztikus extrém érték eloszlás esetén is fennáll. Igazi jelentősége az aszimptotikusan független eloszlásoknál van, amikoris az összefüggőség mértékét határozza meg. Független változók esetén $\bar{C}(u,v) = (1-u)(1-v)$, tehát $\bar{\chi}(u) \equiv 0$ minden $u \in [0,1]$–ra, így $\bar{\chi} = 0$. A Gauss-féle összefüggőségi modellre

$$\bar{C}(u,v) = \int_{\Phi^{-1}(u)}^{\infty}\int_{\Phi^{-1}(v)}^{\infty} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\Big\{-\frac{1}{2(1-\rho^2)}(s^2 - 2\rho st + r^2)\Big\} \mathrm{d}s\, \mathrm{d}t.$$



A $\bar{\chi}(u)$ összefüggőségi mérték a Gauss-féle összefüggőségi modell esetén. A görbék (lentről felfelé) a $\rho = -0.9, -0.8, \ldots, 0.9$ értékekhez tartoznak.

3.3. ábra.

Az is megmutatható, hogy ekkor $\bar{\chi} = \rho$. Habár $\bar{\chi}$ 1-hez közeli u-kra gyorsan változik, megfigyelhető a (3.3) ábrán, hogy $u > 0.5$–re közelítőleg konstans, de 1–nél kisebb. Az előző ábrával összehasonlítva, látható, hogy az, hogy a változók aszimptotikus függetlenek-e, könnyebben eldönthető a

$\bar{\chi}(u)$ segítségével.

Összefoglalva, ha $\chi \in (0,1]$, az aszimptotikus összefüggőségre utal, $\bar{\chi} \in [-1,1)$ pedig aszimptotikus függetlenségre. A két mértéket együttesen kell figyelembe venni, tehát $(\chi > 0, \bar{\chi} = 1)$ aszimptotikus összefüggőséget, $(\chi = 0, \bar{\chi} < 1)$ aszimptotikus függetlenséget jelent.

Mivel az összefüggőségi mértékek nem függnek a peremeloszlásoktól, most célszerűbb a marginálisokat Frèchet eloszlásúvá transzformálni, vagyis X és Y eloszlásfüggvénye $F(z) = \exp(-1/z)$, $z > 0$–ra és $z \to \infty$ esetén $P(X > z) = P(Y > z) \sim z^{-1}$.

Ekkor

$$\chi = \lim_{z \to \infty} P(Y > z | X > z),$$

és

$$P(X > z, Y > z) \sim \mathcal{L}(z)(P(X > z))^{1/\eta} \tag{3.4}$$

nagy z-re. (3.2) és (3.4) alapján

$$\bar{C}(u,u) \sim \mathcal{L}((1-u)^{-1})(1-u)^{1/\eta} \qquad u \to 1$$

Ezért

$$\bar{\chi}(u) \sim \frac{2\log(1-u)}{\log \mathcal{L}((1-u)^{-1}) + \frac{1}{\eta}\log(1-u)} - 1 \to 2\eta - 1 \qquad u \to 1.$$

Tehát $\bar{\chi} = 2\eta - 1$. Továbbá, ha $\eta = 1$ és $\mathcal{L}(z) \to c$, $z \to \infty$ esetén valamilyen $0 < c \leqslant 1$–re, akkor $(\chi = c, \bar{\chi} = 1)$ és a változók aszimptotikusan összefüggőek c mértékben.

Legyen $T = \min(X, Y)$.

$$P(T > z) = P(X > z, Y > z) \sim \mathcal{L}(z) z^{-1/\eta} \qquad z \to \infty.$$

η a T egydimenziós változó alak-paramétere, ami az egydimenziós extémérték elméletben kidolgozott módszerek (pl. likelihood becslés) alkalmazásával becsülhető. Mivel $\bar{\chi} = 2\eta - 1$, ezért így az (X, Y) pár összefüggőségéről is információhoz jutunk. Ezen kívül a χ paraméter is becsülhető, mint a T paramétere, abban az esetben, ha $\eta = 1$ teljesül és $\mathcal{L}(z) = \chi$ elég nagy z-re. A módszer hiányossága, hogy aszimptotikusan független változók esetén nem ad helyes eredményt. A megoldás az, hogy $\bar{\chi}$–t és χ–t együttesen kell vizsgálni és χ becslését csak akkor fogadjuk el, ha $\bar{\chi}$ eltérése 1-től nem szignifikáns.

3.2. Extrém érték eloszlás modellek

A modelleknek lényegében két osztálya ismeretes:

- modellek a komponensenként vett blokk-maximumokra, ahol az adatok extrém értékekből állnak;
- küszöbérték modellek, ahol ismert az egész adatsor, és csak egy bizonyos magas küszöbérték fölött karakterizáljuk az eloszlást.

Az egydimenziós extrém értékek modellezésének klasszikus módszere, egy általánosított extrém érték eloszlás illesztése egy adatsor éves maximumaira. Azonban ennél jobb becslést kaphatunk azáltal, hogy egy megfelelően magas küszöbérték független túllépéseit modellezzük az általánosított Pareto eloszlás segítségével. Ezeket a módszereket terjesztjük ki a többdimenziós megfigyelések vizsgálatára is. Az alapvető különbség az, hogy most figyelembe kell venni az egyes komponensek közötti összefüggőséget is.
A hagyományos módszer, az egydimenziós esethez hasonlóan, a komponensenként vett maximumokból álló vektor határeloszlásával foglalkozik. A pontfolyamat reprezentáció viszont a legalább egyik komponensében extrém megfigyeléseket veszi figyelembe és ezáltal pontosabb eredményeket kaphatunk. Ezen kívül lehetővé teszi a marginálisok és az összefüggőségi struktúra együttes becslését, ami szintén javítja a pontosságot. Egyedüli hátránya az, hogy ehhez nagyszámú független megfigyelésre van szükségünk.

3.2.1. Küszöbérték modell

Ha egy teljes adatsor rendelkezésünkre áll, nem csak évenkénti maximumok, akkor az extremális viselkedést pontfolyamattal jellemezhetjük. A módszert tetszőleges k-dimenziós vektorváltozókra ismertetjük. Legyenek $\mathbf{X_1}, \mathbf{X_2}, \ldots$ független, azonos eloszlású $\mathbb{R}_+^k$–beli véletlen vektorok standard Frèchet peremeloszlásokkal. Legyen

$$P_n = \{n^{-1}\mathbf{X_i} : i = 1, \ldots, n\}$$

pontfolyamat sorozat $\mathbb{R}_+^k$–n. Itt a Frèchet változók max-stabilitás (1.1) miatt az n^{-1}–el való normálás esetén teljesül. Ekkor P_n eloszlásban konvergál egy nemhomogén P Poisson-folyamathoz $\mathbb{R}_+^k \setminus \{0\}$ -n. A következő koordinátatranszformációk után :

$$r_i = \sum_{j=1}^{k} \frac{X_{i,j}}{n}, \qquad w_{i,j} = \frac{X_{i,j}}{nr_i} = \frac{X_{i,j}}{\sum_{j=1}^{k} X_{i,j}},$$

(ahol $X_{i,j}$ az $\mathbf{X_i}$ j-edik komponense), a P intenzitásfüggvénye:

$$\mu(\mathrm{d}r \times \mathrm{d}w) = \frac{\mathrm{d}r}{r^2}\mathrm{d}H(w) \tag{3.5}$$

Itt H egy véges pozitív mérték az

$$S_k = \{(w_1, \ldots, w_k) : \sum_{j=1}^{k} w_j = 1, w_j \geqslant 0, j = 1, \ldots, k\}$$

egységszimplexen és az is teljesül, hogy:

$$\int_{S_k} w_j \mathrm{d}H(w) = 1 \qquad j = 1, \ldots, k.$$

Ha pl több helyszín feletti vihar hatásait akarjuk ilyen módon formalizálni, akkor felfogható úgy, hogy r_i az i. helyi vihar erősségének standardizált mértéke az egész térség felett, $w_{i,j}$ pedig a relatív erőssége a j. rész felett.

A folyamat legontosabb jellemzője az együttes farok-valószínűségek mértékének csökkenése. A (3.5) képletből következik, hogy ha $A \subset B$, ahol $B^C = \{(x_1, \ldots, x_k) : x_i < x_0, i = 1, \ldots, k\}$ elég nagy x_i-k esetén, akkor

$$P((\mathbf{X_1},\ldots,\mathbf{X_k})\in tA)=P((\mathbf{X_1}/t,\ldots,\mathbf{X_k}/t)\in A)\sim$$
$$\sim\frac{1}{t}P((\mathbf{X_1},\ldots,\mathbf{X_k})\in A)\qquad\forall t\geqslant 1\tag{3.6}$$

Ebben a pontfolyamat karakterizációban az aszimptotikus összefüggőség csak a $P((\mathbf{X_1},\ldots,\mathbf{X_k})\in A)$ tényezőre van hatással. Azon feltevés mellett, hogy a Poisson-folyamat helyes közelítés egy bizonyos magas küszöbérték felett, likelihood becsléssel meghatározhatjuk a paramétereit.

Egy másik módszer pedig az, hogy megbecsüljük a $P((\mathbf{X_1},\ldots,\mathbf{X_k})\in A)$ valószínűségeket és a (3.6) képlet alapján következtetünk a tA halmazhoz tartozó valószínűségekre, amelybe talán nem is esik megfigyelés. Elméletileg ez a megközelítés helyes aszimptotikusan független ill. összefüggő változókra is, de a függetlenség esetén nehézségekbe ütközhetünk, mert ilyenkor a H mérték elfajuló. Ekkor vagy a P_n viselkedése tér el jelentősen a P folyamattól, vagy az is előfordulhat, hogy a függetlenség beazonosítható, de az események valószínűségére rossz eredményeket kapunk a (3.6)–t felhasználva. Mindkét probléma kiküszöbölhető, ha kiterjesztjük ezt az eljárást olymódon, hogy figyelembe vesszük az összefüggőség mértékét az aszimptotikusan független eloszlásoknál is. Mivel az η összefüggőségi mértéket már bevezettük a 2-dimenziós vektorváltozóknál, szorítkozzunk most erre az esetre. Definiáljunk egy pontfolyamatot $\mathbb{R}_+^2$–en a következőképpen:

$$\tilde{P}_n=\{n^{-\eta}(\mathbf{X_i},\mathbf{Y_i}):i=1,\ldots,n\}$$

Ekkor $\mathbb{R}_+^2$ belsejében $\tilde{P}_n\to\tilde{P}(n\to\infty)$, ahol $\tilde{P}$ egy Poisson-folyamat

$$\tilde{\mu}(\mathrm{d}r\times\mathrm{d}w)=\frac{\mathrm{d}r}{r^2}\mathrm{d}\tilde{H}(w)$$

intenzitásfüggvénnyel. $\eta=1$ esetén visszakapjuk az előző előállítást. A (3.6)–hez hasonló tulajdonság itt is levezethető. Ha $A\subset B$, ahol

$$B=\{(x,y):x+y>r_0,w_0\leqslant\frac{x}{x+y}\leqslant 1-w_0\},$$

elég nagy r_o–ra és kicsi $w_0>0$–ra, akkor az intenzitásfüggvény alapján:

$$P((\mathbf{X},\mathbf{Y})\in tA)\sim\frac{1}{t^{1/\eta}}P((\mathbf{X},\mathbf{Y})\in A)\qquad t\geqslant 1$$

Tehát a valószínűségek csökkenésének mértékét meghatározza az η, vagy ami ezzel ekvivalens a $\bar{\chi}$. Erre a modellre is ugyanolyan tesztek végezhetők, mint az előző esetben, azzal a különbséggel, hogy itt η–t és $\tilde{H}$–t egyszerre becsülhetjük. Ennek következtében jobb becslést kaphatunk, hiszen az egész adatsort figyelembe vettük.

3.2.2. Komponensenkénti blokk-maximumok modellezése

Ezt a módszert akkor alkalmazzák, ha például csak az éves maximumok ismertek egy folymat esetében. A modell a Poisson-folyamat reprezentáció következménye. Legyen ugyanis $A = \mathbb{R}_+^k \setminus \{(0,x_1) \times \ldots \times (0,x_k)\}$. Ekkor $P(n^{-1}\mathbf{X_i} \notin A, i = 1,\ldots,n) \to \exp(\mu(A)), (n \to \infty)$, ahol

$$\begin{aligned}\mu(A) &= \int_A \frac{\mathrm{d}r}{r^2}\mathrm{d}H(w)\\ &= \int_{S_k}\int_{r\,=\,\min\limits_{1\leqslant j\leqslant k}(x_j/w_j)}^{\infty} \frac{\mathrm{d}r}{r^2}\mathrm{d}H(w)\\ &= \int_{S_k} \max_{j=1,\ldots,k}\left(\frac{w_j}{x_j}\right)\mathrm{d}H(w)\end{aligned}$$

Mivel $P(n^{-1}\mathbf{X_i} \notin A, i = 1,\ldots,n) \to \exp(\mu(A)) = P(n^{-1}M_{n,j} \leqslant x_j,$ $j = 1,\ldots,k)$, ezért a Frèchet eloszlású vektorváltozók komponensenként normalizált maximumainak határeloszlása mindig

$$G(\mathbf{x}) = \exp(-V(\mathbf{x}))$$

alakú, ahol

$$V(\mathbf{x}) = \int_{S_k} \max_{j=1,\ldots,k}\left(\frac{w_j}{x_j}\right)\mathrm{d}H(w)$$

és H (3.5)–t kielégítő mérték. Ez Pickand többdimenziós, Frèchet peremeloszlású extrém érték eloszlásokra vonatkozó reprezentációs tétele.
A részletesebb tárgyaláshoz most is térjünk vissza a kétdimenziós esetre.Ekkor

$$P(n^{-1}M_{n,1} \leqslant x, n^{-1}M_{n,2} \leqslant y) \to G(x,y) = exp(-(V(x,y))) \tag{3.7}$$

Vagyis V és H bármelyike meghatározza az összefüggőséget. Ha létezik $H(w)$–nek sűrűségfüggvénye, akkor

$$h\left(\frac{x}{x+y}\right) = -\frac{(x+y)^3}{2}\,\frac{\partial^2 V(x,y)}{\partial x\,\partial y},$$

tehát ilyenkor könnyen számolható egyik a másikból. Kihasználva, hogy V -1 -edrendű homogén függvény kell, hogy legyen, megállapíthatjuk, hogy, a függetlenség speciális esetét kivéve, minden kétdimenziós extrém érték eloszlás aszimptotikusan összefüggő, vagyis $\bar{\chi} = 1$ érvényes az egész függvénycsaládra. Továbbá

$$\chi = 2 - V(1,1)$$

vagy, ami ezzel ekvivalens:

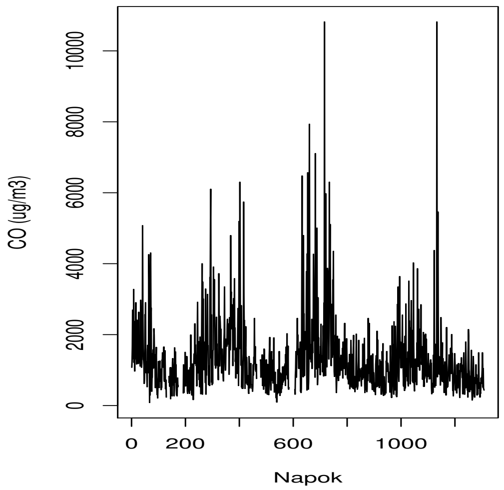
$$\chi = 2\int_0^1 \min(w, 1-w)\,\mathrm{d}H(w),$$

azaz a χ paraméter jellemzi V–t illetve H–t is.
A módszer alkalmazásakor feltesszük, hogy a maximum vektor eloszlása (3.7). Eredményül vagy azt kapjuk, hogy $\chi = \bar{\chi} = 0$, ami függetlenséget jelent, vagy aszimptotikus összefüggőség esetén $\bar{\chi} = 1$ és $\chi > 0$, ahol a χ nagyobb értéke erősebb összefüggésre utal. Hogy a kétdimenziós extrém érték eloszlás nem jó modell, arra abból következtethetünk, hogy $\bar{\chi}$ 1-nél szignifikánsan kisebb, vagy pedig nem χ állandó, hiszen ennek az eloszlás minden tagjára teljesülnie kell.

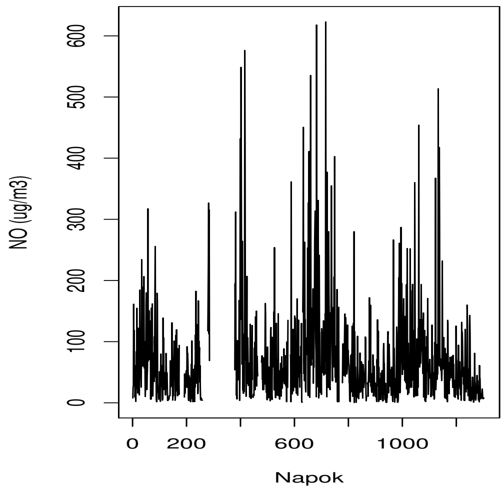
4. fejezet

Alkalmazások

A továbbiakban egy konkrét adatsor elemzésével foglalkozunk, ami különböző szennyezőanyagok (pl CO,CO2,NO,SO,SO2,por) koncentrációiból áll. Ezeket az értékeket félóránként mérték Szegeden 1997. januárja és 2000. augusztusa között. Közülük a szén-monoxidra és nitrogén-monoxidra vonatkozó adatokat választottuk ki és azt szeretnénk megállapítani, hogy ezekre milyen kétdimenziós extrém érték eloszlás illeszthető.



A CO koncentrációk napi maximumai



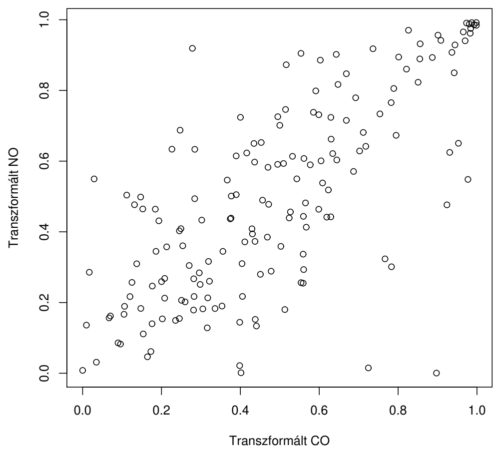
A NO koncentrációk napi maximumai

4.1. ábra.

Ez az elemzés eredetileg a városi levegő minőségével foglalkozó nemzetközi konferenciára készült. A szükséges programok megírásához az S-PLUS programcsomagot használtuk.

Az adatsorban a trend el hanyagolható volt, csak az éves ciklust távolítottuk el. Ez utóbbit a napi maximumokra (4.1) illesztettük, mert egyébként a nagy számú, általában alacsony koncentráció értékek nagy súllyal szerepeltek volna a számunkra fontos magasabb értékek mellett. Csak ezek után tértünk rá a heti maximumok vizsgálatára, amikről már feltehetjük, hogy független megfigyelések.

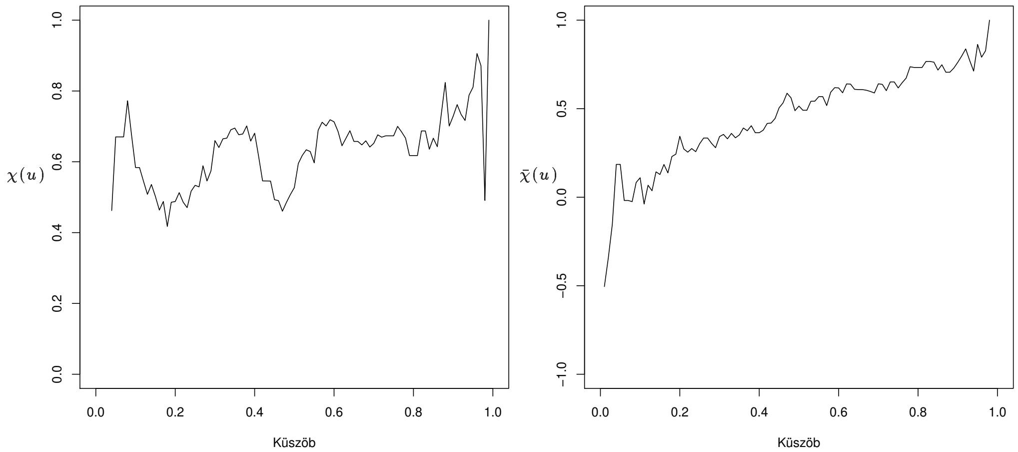
Ebben az esetben a komponensenkénti blokk-maximumokra kidolgozott módszert tudjuk alkalmazni. Azzal a feltételezéssel élve, hogy a marginálisok valamilyen extrém érték eloszlascsaládba tartoznak, az XTREMES nevű programcsomag segítségével becsülhetjük a paramétereit. Miután a tapasztalati eloszlásokat ismerjük, egyenletes eloszlásúvá transzformálhatjuk a marginálisokat. Ha a már átalakított adatokat ábrázoljuk (4.2), előzetes képet kaphatunk az összefüggőségről. A pontok az átló köré csoportosulnak, vagyis ritka



CO és NO levegőben mért koncentrációinak heti maximumai, miután azokat egyenletes eloszlásúvá transzformáltuk.

4.2. ábra.

az az eset, hogy pl nagy CO koncentráció esetén az NO koncetráció értéke kicsi. Tehát ebből arra következtethetünk, hogy a két érték nem független egymástól. Ezek után ábrázoljuk a $\chi(u)$ és $\bar{\chi}(u)$ függvényeket, amelyeket a (3.1) és (3.3) képleteket felhasználva kapunk úgy, hogy a valószínűségeket a relatív gyakoriásggokkal becsültük. Látható (4.3), hogy $\bar{\chi}(u) \to 1$, ami szintén a marginálisok összefüggőségére utal. Ennek mértéke a χ-vel jellemezhető, ami erős összefüggést mutat.



A χ összefüggőségi paraméter becslése, u küszöbértékek függvényében

A $\bar{\chi}$ összefüggőségi paraméter becslése, u küszöbértékek függvényében

4.3. ábra.

A kétdimenziós extrém értékek modellezésére széleskörűen elterjedt a logisztikus modell használata, mi is ezt illesztettük az adatokra. Ebben az esetben

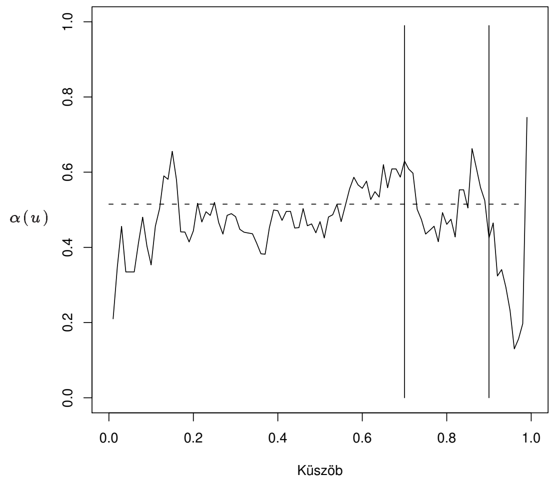
$$V(x, y) = (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha},$$

így

$$\chi = 2 - V(1, 1) = 2 - 2^{\alpha}$$

miatt egyúttal α–ra is kaptunk becslést (4.4) a küszöbérték függvényében. Mivel minket az extrém értékek összefüggősége érdekel, úgy döntöttünk, hogy a 0,7 és 0,9 közötti küszöböket vesszük figyelembe. A 0,9-nél nagyobb küszöböket azért zártuk ki, mert ott a kis megfigyelésszám miatt nagy a szórás. Első becslésként az ebben az intervallumban kapott értékek átlagát fogadtuk el (α=0,5155).

Két másik (egydimenziós) módszerrel is becsültük α–t és ezekkel is hasonló eredményeket kaptunk (ld. a táblázatot). A visszatérési idők (vagyis, hogy egy extrém esemény mikor következik be legközelebb) becslései kevesebb, mint 10 %-kal térnek el egymástól, s a legnagyobb értéket a többdimenziós



Az α becslése a küszöbértékek függvényében. A két függőleges vonal közé eső értékeket átlagoltuk. Ennek eredménye a vízszintes szaggatott vonallal jelzett érték ($\alpha = 0,5155$).

4.4. ábra.

módszer adja. Ez valószínűleg annak tudható be, hogy itt csak a nagyobb értékeket vettük figyelembe, ami a mi esetünkben helyénvalónak tűnik, mivel az extrém érték modell tökéletes illeszkedése nem remélhető (amit az ML módszernél feltételezünk).

	(CO,NO) μg/m^3	többdimenziós módszer	maximum likelihood	momentum típusú
α		0.5515	0.4608	0.4951
visszatérési idő, év	(6000,500)	3.934	3.603	3.802
visszatérési idő, év	(8000,600)	9.333	9.809	10.128
visszatérési idő, év	(14000,900)	105.0613	99.46	102.74

KÖSZÖNETNYILVÁNÍTÁS

Köszönetet mondok elsősorban témavezető tanáromnak,
Dr. Zempléni Andrásnak, továbbá Csiszár Villőnek
a dolgozat elkészítéséhez nyújtott segítségéért.

Irodalomjegyzék

[1] Coles, S.G., Heffernan,J., Tawn, J.A.: Dependence Measures for Extreme Value Analyses, *Extremes* 2:4, 339-365,(1999).

[2] Coles, S.G., Tawn, J.A.: Modelling Extreme Multivariate Events, *J.R.Statist.Soc.*B **53**,(1991).

[3] Joe,H., Smith, R.L., Weissmann, I.: Bivariate Threshold Methods for Extremes, *J.R.Statist.Soc.*B **54**, 171-183, (1992).

[4] Embrecht, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events

[5] de Haan, L.: Extremes in Higher Dimensions: The Modell and Some Statistics, In *Proc.45th Sess.Int.Statist.Inst.*, paper 26.3 International Statistical Institute, The Hague, (1985).

[6] Ledford, A.W., Tawn, J.A.: Modelling Dependence within Joint Tail Regions, *J.R.Statist.Soc.*B **59**, 475-499, (1997)

[7] Ledford, A.W., Tawn, J.A.: Statistics for Near Independence in Multivariate Extreme Values, *Biometrika*, **83**, 169-187, (1996)

[8] Makra, L., Horváth, Sz., Zempléni, A., Csiszár, V., Rózsa, K., Motika, G.: Air Quality Trends in Southern Hungary, *The Third International Conference on Urban Air Quality and Fifth Saturn Workshop*, (2001)

[9] Makra, L., Horváth, Sz., Zempléni, A., Csiszár, V., Rózsa, K., Motika, G.: Levegőminőségi trendek Magyarországon, különös tekintettel a dél-alföldi régióra, *Légkör*, (2001)