Többdimenziós küszöb-meghaladás modellek és alkalmazásuk szélsebesség adatsorokra

Doktori értekezés tézisei

Pál Rakonczai

Témavezető:
Zempléni András, egyetemi docens, kandidátus

MATEMATIKA DOKTORI ISKOLA
Vezető: Laczkovich Miklós
ALKALMAZOTT MATEMATIKA DOKTORI PROGRAM
Vezető: Michaletzky György

Eötvös Loránd Tudományegyetem
Természettudományi Kar
Budapest, 2012

1. Bevezetés

Témakör. A doktori értekezésben olyan módszereket vizsgálok és fejlesztek, amelyek alkalmasak többdimenziós adatsorok extrém magas értékeinek az együttes modellezésére. Az értekezés tartalmazza a többdimeziós extrék-érték elmélet (MEVT) tömör áttekintését, és emellett az elmúlt 5 év alatt végzett kutatómunkám eredményeinek részletes bemutatását.

Motiváció. A témafelvetés egy alkalmazott statisztikai projekthez kapcsolódik, amely során különböző mérőállomásokon egyidőben megjelenő magas szélsebesség adatok - havi maximumok, vagy magas küszöböket meghaladó értékek - modellezése volt az egyik fő kutatási cél. Természetesen a bemutatott módszerek ennél sokkal általánosabb körben is alkalmazhatók, lásd Krusper (2011), ahol pénzügyi és biztosításmatematikai alkalmazások vagy Zempléni és Rakonczai (2011), ahol hidrológiai alkalmazások kerülnek bemutatásra.

Hatás. Az eredményeim a statisztikai alkalmazások területén hasznosíthatók leginkább, főleg azon problémáknál, amelyekben az extrém-értékek modellezése fontos feladat, de az alkalmazások mellett elméleti eredményeket is felsorakoztatok a modellek statisztikai tulajdonságairól illetve alkalmazhatóságuk feltételeiről.

Szoftver. Munkám során kiemelt figyelmet szenteltem egy új `R` szoftver csomag kifejlesztésének, melynek az első - kétdimenziós modellezésre alkalmas - verzióját, `mgpd` néven 2011-ben publikáltam. A szoftver szabadon letölthető a `Comprehensive R Archive Network` honlapjáról, és az általános `Distributions` csomag részeként is megtalálható. Mára széleskörben elterjedt a programcsomag felhasználása és a szoftver karbantartójaként a további fejlesztéseket a felhasználók igényeit követve végzem. Az értekezés gyakorlati alkalmazásai is ezzel az eszközzel készültek, így a bemutatásra kerülő eredmények az idézett kódrészletek futtatásával bárki számára reprodukálhatók.

Vázlat. A tézis összefoglaló követi magának a doktori értekezésnek a szerkezetét. A 2. szakaszban egy rövid bevezetőt követően a későbbi modellezés alapjául szolgáló valószínűségelméleti eredményeimet vázolom. A 3. szakaszban bemutatok egy új, univerzálisan használható konstrukciós módszert, amely segítségével aszimmetrikus összefüggési struktúrák családjai állíthatók elő, az eljárást két konkrét példával illusztrálom. A szakasz végén egy új statisztikai eszközt is bemutatok, melyet az autokorrelációhoz való hasonlósága révén autokopulának nevezek. A 4. szakaszban illeszkedésvizsgálati és szimulációs módszereket tárgyalok, majd szimulációk segítségével a bemutatott új módszerek és modellek plauzibilitását is bizonyítom. Számos 2 dimenziós (2D) és 3 dimenziós (3D) alkalmazást mutatok be szélsebesség adatokon a 5. szakaszban, végül további megvalósításra váró ötleteket vázolok a 6. szakaszban.

2. Extrém-érték elmélet

A szükséges jelölések bevezetését követően a többdimenziós általánosított Pareto eloszlás (MGPD) egy érdekes invariancia tulajdonságát bizonyítom. Majd megmutatom, hogy az MGPD jelen definícióját feltételezve a lehetséges paraméteres összefüggésmodell-osztályok közül csak bizonyosak vezetnek abszolút folytonos eloszláshoz és karakterizálom az abszolút folytonosságot egy könnyen ellenőrizhető modell-tulajdonsággal. A véges dimenziós modelleken túl a végtelen dimenziós általánosítást is tárgyalom. Ennek kapcsán megmutatom, hogy mely paraméteres részosztályok teljesítik a fenti alkalmazhatósági feltételt és melyek nem.

2.1. Előzmények

Jelölje d-dimenziós megfigyelések koordinátánkénti maximumát $\mathbf{M}_n = \left(\bigvee_{i=1}^n X_{i,1}, ..., \bigvee_{i=1}^n X_{i,d}\right)$. Ha létezik normalizáló vektorok $\mathbf{a}_n$ és $\mathbf{b}_n > \mathbf{0}$ sorozata, amelyre

$$P\Big(\frac{\mathbf{M}_n - \mathbf{a}_n}{\mathbf{b}_n} \leq \mathbf{z}\Big) = \!\rightarrow \mathbf{G}(\mathbf{z}), \tag{1}$$

ahol a G_i marginális eloszlások nemelfajuló egydimenziós eloszlások, akkor a $\mathbf{G}$ határeloszást többdimenziós extrém-érték eloszlásnak nevezzük (később: MEVD). Az MEVD karakterizációja során gyakran ún. standard Fréchet marginálisokat feltételezünk, jelölje ezt $\mathbf{G}_\star$. További egyszerűsítésként vizsgáljuk csupán a kétdimenziós esetet, jelölje az exponensmértéket $\mu_\star$, az exponensmérték-függvényt $V_\star$, a spektrálmértéket W és a (Pickands-féle) összefüggésfüggvényt A. A későbbiekben ezekre összefüggési struktúraként hivatkozunk. A fenti jelölésekkel az alábbi reprezentációk lehetségesek (lásd de Haan és Resnick, 1977 vagy Pickands, 1981)

$$\begin{aligned} & -\log G_\star(y_1, y_2) = \mu_\star([\mathbf{0}, \infty) \setminus [\mathbf{0}, \mathbf{y}]) = V_\star(y_1, y_2) \qquad (2) \\ = & \int_{S_2} \bigwedge_{j=1}^{2} \{\omega_j \log G_j(y_j)\} W(d\omega) = \left(\frac{1}{y_1} + \frac{1}{y_2}\right) A\left(\frac{y_1}{y_1 + y_2}\right). \end{aligned}$$

Az A összefüggésfüggvény egy konvex függvény, amelyre $(1-t) \vee t \leq A(t) \leq 1$ for $t \in [0,1]$, amíg a $\mu_\star$ exponensmértékhez tartozó W spektrálmérték teljesíti az $\int_{S_2} \omega_j W(d\omega) = 1$, egyenletrendszert, ahol $j = 1,2$, és S_2 a kétdimenziós egységszimplex.

2.2. Az MGPD és a sűrűségfüggvény létezésének feltétele

1. Definíció (Rootzén és Tajvidi, 2006). *Egy H eloszlásfüggvényt többdimenziós általánosított Pareto eloszlásúnak (MGPD) hívunk, ha az a következő alakú*

$$H(\mathbf{x}) = \frac{-1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}, \tag{3}$$

ahol G nemelfajuló marginálisú MEVD és $0 < G(\mathbf{0}) < 1$.

Az MGPD kétdimenziós változatát BGPD-nek, a háromdimenziós változatát TGPD-nek nevezzük a későbbiekben. Legyen $\mathbf{X}$ egy d-dimenziós valószínűségi vektorváltozó F eloszlásfüggvénnyel, valamint $\{\mathbf{u}(t) : t \in [1, \infty)\}$ egy d-dimenziós görbe, amelyre $\mathbf{u}(1) = \mathbf{0}$. Legyen továbbá $\sigma(\mathbf{u}) = \sigma(\mathbf{u}(t))$ pozitív függvény $\mathbb{R}^d$-beli értékekkel. Jelölje az $\mathbf{u}$ szint melletti normalizált szintmeghaladásokat

$$\mathbf{X_u} = \frac{\mathbf{X} - \mathbf{u}}{\sigma(\mathbf{u})}.$$

A kapcsolat az MEVD és az MGPD között a következő.

1. Tétel (Rootzén és Tajvidi, 2006). *Tfh. G egy d-dimenziós MEVD, amelyre $0 < G(\mathbf{0}) < 1$. Ha F a G vonzási tartományába tartozik, akkor létezik egy növekedő, folytonos $\mathbf{u}$ görbe, amelyre $F(\mathbf{u}(t)) \to 1$ ha $t \to \infty$, és egy $\sigma(\mathbf{u}) > \mathbf{0}$ függvény, amelyre*

$$P(\mathbf{X_u} \leq \mathbf{x} | \mathbf{X_u} \nleq \mathbf{0}) \to \frac{-1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \tag{4}$$

ha $t \to \infty$.

Egy további érdekes tulajdonsága az MGPD-nek az alábbi.

2. Tétel (Rakonczai és Turkman, 2012). *Bármely* (X_1, X_2) *BGPD valószínűségi változóhoz tartozik egy* $(0{,}0)$ *pontból induló, folytonos, növekedő* ϱ *görbe, amelyre a* $H(x_1, x_2)$ *eloszlásfüggvény értéke a* ϱ *görbe pontjain invariáns az összefüggőségi struktúra (2. egyenlet) megváltoztatására.*

Hasonló tulajdonságot bizonyítok általánosan, magasabb dimenziós esetekre is. A következő észrevétel hasznos abszolút folytonos MGPD modellek konstruálásakor.

3. Tétel (Rakonczai és Zempléni, 2011). *Legyen* H *egy MGPD eloszlásfüggvény, amely egy abszolút folytonos* G *MEVD eloszlásfüggvény segítségével áll elő, továbbá jelölje ennek spektrálmértékét* W. H *abszolút folytonos akkor és csak akkor, ha* $W(int(\mathcal{S}_d)) = d$, *vagyis ha* W *a szimplex belsejére koncentrálódik. A kétváltozós esetben mindez ekvivalens a* $W(\{0\}) = W(\{1\}) = 0$ *vagy*

$$-A'(0) = A'(1) = 1, \tag{5}$$

esetekkel, ahol A' *az összefüggésfüggvény deriváltja.*

2.3. Extém-érték elmélet végtelen dimenzióban

A max-stabilis folyamatokat (de Haan, 1984 és Vatan, 1985) tekinthetjük a véges dimenziós extrémérték eloszlások végtelen dimenziós általánosításának.

2. Definíció (de Haan, 1984). *Legyen* T *egy index halmaz, valamint* $\{Y_i(t)\}_{t\in T}$, $i = 1, \dots, n$ *egy folytonos sztochasztikus folyamat* n *független replikációja. Tfh. léteznek folytonos függvényeknek* $a_n(t) \in \mathbb{R}$ *és* $b_n(t) > 0$ *sorozatai amelyekre*

$$Z(t) = \lim_{n\to\infty} \frac{\bigvee_{i=1}^{n} Y_i(t) - a_n(t)}{b_n(t)}\ ,\ t \in T.$$

Ha a fenti határ létezik, a $Z(t)$ *határfolyamat neve max-stabilis folyamat.*

A következő tételek folyamatok szintmeghaladásainak modellezésekor hasznosak.

4. Tétel (Rakonczai és Turkman, 2012). *Az a BGPD model, melyet az alábbi Smith-féle exponens mérték segítségével állítunk elő*

$$\begin{aligned} V_\star(z_1, z_2) &= \frac{1}{z_1}\Phi\left(\frac{m(h)}{2} + \frac{1}{m(h)}\log\left(\frac{z_2}{z_1}\right)\right) \\ &+ \frac{1}{z_2}\Phi\left(\frac{m(h)}{2} + \frac{1}{m(h)}\log\left(\frac{z_1}{z_2}\right)\right), \end{aligned} \tag{6}$$

abszolút folytonos.

5. Tétel (Rakonczai és Turkman, 2012). *Az a BGPD model, melyet az alábbi Schlather-féle exponensmérték segítségével állítunk elő*

$$V_\star(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \left[1 - 2(\varrho(h) + 1)\frac{z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right) \tag{7}$$

nem abszolút folytonos.

3. Paraméteres családok és kiterjesztéseik

Az abszolút folytonos MGPD modellek limitált voltának lehetséges feloldásaként egy új konstrukciós eljárást mutatok be (lásd Rakonczai és Zempléni, 2011), amely lehetővé teszi, hogy további összefüggésmodelleket építsünk fel már létezőekre alapozva. Az eljárás segítségével mutatok néhány példát új paraméteres családra, melyek könnyen alkalmazhatók a kétdimenziós esetben (lásd Rakonczai, 2012) de akár magasabb dimenziószám mellett is. Végül, de nem utolsó sorban kiterjesztem a kopulák használatát idősorok egymás utáni (vagy távolabbi) értékeinek összefüggési struktúrájának leírására.

3.1. Új aszimmetrikus modellek 2D-ben

Emlékeztetőül, ha $-A'(0) = A'(1) = 1$ teljesül, akkor az A összefüggésfüggvény által előállított BGPD abszolút folytonos. Keressünk egy alkalmas $\Upsilon(x)$ transzformációt az alábbi algoritmusnak megfelelően:

I. Tekintsünk egy paraméteres baseline összefüggésmodellt, amely kielégíti az 5. egyenlet feltételeit is; (pl. szimmetrikus logisztikus/negatív logisztikus, stb.);

II. Adjunk meg egy szigorúan monoton $\Upsilon(x)$ transzformációt, amelyre $\Upsilon(x) : [0,1] \to [0,1]$, amelyre $\Upsilon(0) = 0$, $\Upsilon(1) = 1$;

III. Konstruáljunk egy új összefüggésmodellt a baseline modell segítségével $A_\Upsilon(t) = A(\Upsilon(t))$;

IV. Ellenőrizzük, hogy A_Υ továbbra is érvényes összefüggésfüggvény marad-e, kielégítve az 5. egyenlet extra feltételeit is.

A konstrukcióhoz kézenfekvő feltételezés a transzformáció $\Upsilon(t) = t + f(t)$ alakban való keresése. Ekkor, f-et úgy választva, hogy a deriváltakra $f'(0) = f'(1) = 0$ teljesül, az 5. egyenlet automatikusan érvényben marad, bár természetesen a határfeltételek és a konvexitás még ellenőrzésre szorulnak. A disszertációban két példát mutatok be. Elsőként legyen

$$f_{\psi_1,\psi_2}(t) = \psi_1(t(1-t))^{\psi_2}, \text{ for } t \in [0,1], \tag{8}$$

ahol $\psi_1 \in \mathbb{R}$ és $\psi_2 \geq 1$ az aszimmetria paraméterek. Egy másik opcióként tekinthetjük a következő, szintén 2 paraméteres függvényt, ami kielégíti az alábbi feltételeket (lásd Rakonczai, 2012).

- $f_{\phi_1,\phi_2}(0) = f_{\phi_1,\phi_2}(1) = 0$ és $f'_{\phi_1,\phi_2}(0) = f'_{\phi_1,\phi_2}(1) = 0$,
- $f_{\phi_1,\phi_2}\left(\frac{1}{\phi_2}\right) = 0$, $f_{\phi_1,\phi_2}\left(\frac{1}{2\phi_2}\right) = \phi_1$ és $f_{\phi_1,\phi_2}\left(\frac{\phi_2+1}{2\phi_2}\right) = -\phi_1$,

pl. Hermite interpoláció segítségével.

3.2. Új aszimmetrikus modellek magasabb dimenziókban

A 3.1. alszakaszban bemutatott ötlet magasabb, mint kétdimenziós esetekben is alkalmazható. Ilyenkor azt a transzformációt keressük, amely alkalmazása a baseline-ra nem sérti a többdimenziós összefüggésfüggvény kritériumait. A konvexitás ellnőrzéséhez használhatjuk a másodrendű konvexitási feltételt. A $\Psi-$transzformáció kiterjeszthetéseként (8. egyenlet) tekinthetjük a következő $\Psi(t_1, \dots, t_{d-1})$ transzformációt

$$\left(t_1 + \psi_{1,1}\left(t_1\left[1 - \sum_{i=1}^{d-1} t_i\right]\right)^{\psi_{1,2}}, \dots, t_{d-1} + \psi_{d-1,1}\left(t_{d-1}\left[1 - \sum_{i=1}^{d-1} t_i\right]\right)^{\psi_{d-1,2}}\right).$$

Míg ennek a módszernek a használata Rakonczai és Zempléni (2011) cikkében a jövőbeli kutatási célok között kerül megemlítésre, az értekezés során már tárgyalom a módszer alkalmazását 3D szél adatokon. Az `R` függvények, amelyekkel a bemutatott modellek illeszthetők az `mgpd` csomag legújabb verziójában már megtalálhatók.

3.3. Kopuláktól az autokopulákig

Az autokorrelációs függvény analógiájára bevezetem az idősorok ún. autokopuláit (lásd Rakonczai et al., 2011), amelyek segítségével az idősor különböző lag-jei között fennálló összefüggési struktúra sajátosságai feltárhatóvá válnak. Kopula illeszkedés vizsgálati módszereket alkalmazva (lásd Rakonczai és Zempléni, 2007) az összefüggésekről az autokorrelációknál sokkal árnyaltabb képet kaphatunk. A kopulák alaptétele (Sklar, 1959) szerint minden folytonos d-dimenziós H eloszlásfüggvényhez, amelynek egydimenziós marginálisait jelölje $F_i, i=1,\ldots,d$, egyértelműen létezik egy C kopula, azaz egy d-dimenziós eloszlásfüggvény az egységkockában [0,1] intervallumon egyenletes eloszlású marginálisokkal, hogy H felírható $H(x_1,...,x_d) = C(F_1(x_1),...,F_d(x_d))$ alakban.

3. Definíció (Rakonczai et al. 2011). *Adott szigorúan stacionárius Y_t idősor és $l \in \mathbb{Z}^+$ esetén a $C_{Y,l}$ l-lag autokopulát a (Y_t, Y_{t-l}) kétdimenziós valószínűségi változó kopulájaként definiáljuk. Az l-lag autokopulákat l függvényeként tekintve kapjuk az ún. autokopula függvényt.*

Alkalmazások a következő szakaszban találhatók.

4. Illeszkedésvizsgálat és Szimuláció

Ebben a szakaszban először a χ^2-teszt egy változatát mutatom be, melynek segítségével kétváltozós modellek illeszkedése tesztelhető. A teszt során a partíciókat a kétváltozós sűrűségfüggvény jól megválasztott szintvonalai szolgáltatják az eloszlásfüggvény szintvonalai (kvantilis görbék) helyett. Ezt az eljárást a későbbiek során gyakran alkalmazom. Más, speciálisan a kopula modellek illeszkedésvizsgálatára alkalmas tesztek is bemutatásra kerülnek, ezek hasznosak lesznek az előzőekben definiált ún. autokopulák tesztelésekor. Legjobb tudomásom szerint szimulációs módszerek a BGPD eloszlásból (a logisztikus modellen kívül) nem ismertek, ezért az értekezésben bevezetek egy közelítő módszert, amely a gyakorlatban könnyen alkalmazható. Végül, szimulációs vizsgálatok segítségével bizonyítom néhány modell és eljárás hatékony alkalmazhatóságát.

4.1. Illeszkedésvizsgálat

4.1.1. χ^2-Teszt Predikciós Régiókkal

A kompakt predikciós régió definícióját Hall és Tajvidi (2004) vezette be. Ezen régiók segítségével χ^2 illeszkedésvizsgálati tesztet végezhetünk, hiszen a régiók a síkot partícionálják és az egyes partíciók nullhipotézis melletti valószínűségeit is ismerjük. A partíciókon belül a modellből származó és tapasztalati gyakoriságok közötti eltérést a χ^2 statisztika segítségével kézenfekvő módon mérhetjük.

4.1.2. Kopula Illeszkedésvizsgálat

Egy speciális tesztelési algoritmussal (Rakonczai és Zempléni, 2007) amely egy alapvető módszer (Genest et al., 2006) súlyokkal módosított változata a többdimenziós teszteket egydimenziós tesztre vezethetjük vissza. Ezt a gyakorlatban az ún. Kendall-féle függvény segítségével valósíthatjuk meg $\mathcal{K}(t) = P(H(X,Y) \le t) = P(C(F(X), G(Y))) \le t) = P(C(U,V)) \le t)$. Jelölje $\mathcal{K}(t)$ tapasztalati becslését

$\mathcal{K}_n(t)$. Bár zárt formula $\mathcal{K}(t)$-re csak speciális kopula családok esetén létezik, értékeit tetszőleges pontossággal approximálhatjuk szimuláció segítségével bármilyen kopula esetén. Az alkalmazott próbastatisztikákat a 1. táblázatban foglaltam össze, ahol $(t_i)_{i=1}^m$ egy megfelelő fiomságú felosztását jelöli a (0,1) intervallumnak.

1. táblázat. Numerikusan közelített Cramér-von Mises próbastatisztikák a Kendall-féle függvényre

Fókuszált régió	Próbastatisztika
Globális	$S_1 = \frac{1}{m}\sum_{t_i \in [0+\varepsilon, 1-\varepsilon]} (\mathcal{K}(\theta_n, t_i) - \mathcal{K}_n(t_i))^2$
Felső farok	$S_2 = \frac{1}{m}\sum_{t_i \in [0+\varepsilon, 1-\varepsilon]} \frac{(\mathcal{K}(\theta_n, t_i) - \mathcal{K}_n(t_i))^2}{1-\mathcal{K}(\theta_n, t_i)}$
Alsó farok	$S_3 = \frac{1}{m}\sum_{t_i \in [0+\varepsilon, 1-\varepsilon]} \frac{(\mathcal{K}(\theta_n, t_i) - \mathcal{K}_n(t_i))^2}{\mathcal{K}(\theta_n, t_i)}$
Alsó és felső farok	$S_4 = \frac{1}{m}\sum_{t_i \in [0+\varepsilon, 1-\varepsilon]} \frac{(\mathcal{K}(\theta_n, t_i) - \mathcal{K}_n(t_i))^2}{\mathcal{K}(\theta_n, t_i)(1-\mathcal{K}(\theta_n, t_i))}$

4.2. Szimulációs vizsgálatok

Logisztikus BEVD és BGPD összehasonlítása. A BEVD és BGPD vonzási tartományából szimulált minták mutatják, hogy logisztikus összefüggésmodell esetén közepes és magas asszociáció mellett a BGPD becslések jól teljesítenek (hasonló mértékben a BEVD-hez), viszont alacsony asszociáció esetén torzítás jelenik meg a szimuláció során alkalmazott mintaelemszám mellett. Bár természetesen az említett torzítás magasabb elemszámú minták (és magasabb küszöbértékek) esetén eltűnik, a BEVD alacsony asszociáció esetén is hatékony. Hasonló vizsgálatokat más típusú összefüggési struktúrákat feltételezve is végeztem, ahol az eredmények azt mutatják, hogy valójában az aszimptotikus farokösszefüggés nagysága játszik döntő szerepet a BGPD modellek hatékonyságában. Alacsony farok összefüggés esetén a konvergencia lassabb, azaz a BGPD modell magasabb küszöb szint választása mellett ad jó közelítést. Ennek következtében több megfigyelésre van szükség a kielégítő becslések elvégzéséhez. Közepes vagy magasabb asszociáció esetén ez a probléma megszűnik és a BGPD a BEVD-hez hasonló hatékonysággal alkalmazható.

Paraméterbecslések standard hibája aszimmetrikus BGPD esetén. Magából a BGPD családból is végeztem szimuációkat a közelítő szimulációs technika segítségével. Ennek segítségével, a 3.1. alszakaszban látott konstrukciós módszerre koncentrálva, az új paraméteres BGPD családokból generáltam mintákat és kiszámoltam a becslések standard hibáját. A kis eltérések az ismert elméleti paraméter értékektől és a relatívan kicsi standard hiba a modell konstrukció helyességét támasztja alá. (Sőt tulajdonképpen ezzel közvetett módon azt bizonyítjuk, hogy a közelítő szimulációs eljárást kellő pontossággal használjuk.)

Heteroszkedaszticitás tesztelése. Azon idősorok, amelyeknek ugyanaz a gyenge AR reprezentációja, azonos autokovariancia struktúrával rendelkeznek. Ennek következtében persze semelyik autokovarianciákon alapuló teszt nem képes különbséget tenni egy ARCH- (vagy GARCH-) innováció által generált AR (későbbiekben AR-ARCH) és egy egyszerű független azonos eloszlású innovációval rendelkező AR idősor között. Az identifikáció az autokopulák használatával viszont már sokkal ígéretesebb. Elsőként n=500-as mintaméret mellett végeztem szimulációkat az AR és AR-ARCH modellekből, majd az autokopulák azonosságát különböző lag-ek mellett teszteltem a 1. táblázat statisztikáinak segítségével. A nullhipotézisként azt tettem fel, hogy a minták az AR modellből származnak. Az esetek 70%-ában az 1-lag AR-ARCH autokopulákról a fenti teszt eljárással kimutattam, hogy nem-AR autokopulák. Ez az arány eléri a 90%-ot $n = 1000$-as mintaméret, vagy akár meghaladja a 99%-ot is $n = 1500$-as mintaméret esetén. Emellett az is látható, hogy a teszt hatékonyságát a próbastatisztikák megfelelő súlyozásával növelni lehet. Különösen azok a próbastatisztikák teljesítenek jól ebben az esetben, amelyek az eloszlás farkain keletkező eltérésekre koncentrálnak.

5. Alkalmazások Szélsebesség Adatsorokra

A korábban bemutatott módszerek illusztrációjaként az 5 következő észak-németországi mérőállomás szél adatsorait vizsgáltam: Hamburg, Hannover, Bremerhaven, Fehmarn és Schleswig. A modell-illesztésen és a standard és új extrém-érték eloszlások összehasonlításán túl kétdimenziós predikciós régiók becslését is kiszámoltam, ezzel interpretálva az illesztett paramétereket. Egy előzetes egyváltozós analízist követően illesztettem standard BEVD és BGPD modelleket, és bár általánosságban az illeszkedés nem mondható rossznak, számos esetben szofisztikáltabb modell konstrukciókkal, pl. időfüggő paraméterek, vagy extra aszimmetria paraméterek használatával az illeszkedés még nagy mértékben javítható. További újdonság a kétváltozós eset mellett, a háromváltozós eset részletes vizsgálata.

Nem-stacionárius BGPD. A BGPD modell nem-stacionárius kiterjesztését vizsgálom, amely lehetőséget nyújt arra, hogy az extrém események bizonyos jellemzői időről időre változzanak, vagy függjenek más kovariánsok értékeitől (lásd Rakonczai et al., 2010). Ebben a részben 4 állomáspár szintmeghaladásait modellezem. Számos nem-stacionárius modellt illesztek a paraméterek különböző részhalmazait tekintve időfüggőnek. A likelihood-hányados próba alapján a legtöbb esetben mondhatjuk, hogy az időfüggő paraméterek szignifikánsan javítják a stacionárius modellt. Ún. cross-validation módszerrel bizonyítom, hogy az időfüggő modellekben a kvantilis előrejelzések is jobbá válnak.

Aszimmetrikus BGPD és TGPD. Három állomáspárra elsőként 6 különböző, standardnak nevezhető, összefüggésmodelt feltételezve BGPD modellt illesztek, nevezetesen logisztikus, negatív logisztikus, Coles-Tawn (Dirichlet), bilogisztikus, negatív bilogisztikus és Tajvidi (általánosított szimmetrikus logisztikus). Majd ezt követően alkalmazom a korábban bemutatott $\Psi-$ és $\Phi-$transzformációkat logisztikus és negatív logisztikus modellekre, aminek eredményeképp 4 új összefüggésmodell család jön létre. Az új aszimmetria paraméterek az esetek többségében szignifikánsnak bizonyulnak mindhárom állomáspár esetén és az új modellek a legjobban illeszkedő modellek között vannak. A háromváltozós esetben a Ψ-transzformációt alkalmaztam a 3D logisztikus és negatív logisztikus összefüggési struktúrákra, és ezeket feltételezve aszimmetrikus TGPD modelleket is illesztettem. A számítások az `mgpd` (1.99) `R`-csomag új frissített verziójával készültek, és a legfontosabb kód részletek megtalálhatók az értekezésben.

6. További kutatási irányok

Végül bemutatok néhány ötletet, melyek megvalósítása a jövőbeli kutatási céljaim között szerepel. A vizsgált módszerek igazán magas dimenzióba való kiterjesztése még megoldásra vár. Bemutatok egy 5 dimenziós MGPD illesztést ún. páronkénti composite likelihood módszerrel és egy további ötletet, ami lehetővé tenné az általános térbeli alkalmazást is.

A doktori értekezés az alábbi saját publikációkon alapul:

- Rakonczai, P. and Turkman, F. (2012) Applications of generalized Pareto processes. (Technical report under progress, OTKA outgoing mobility grant, Lisbon, Portugal)

- Rakonczai, P. (2012) Asymmetric dependence models for bivariate threshold exceedance models. *Forum Statisticum Slovacum, ISSN 1336-7420* **1**, p.25-32.
- Rakonczai, P. and Zempléni, A. (2012) Bivariate generalized Pareto distribution in practice: models and estimation. *Environmetrics*, John Wiley & Sons, **23**, p.219-227.
- Zempléni, A. and Rakonczai, P. (2011) New bivariate threshold models with hydrological applications. *Conference on Environmental Risk and Extreme Events*, Ascona, July 10-15
- Rakonczai, P., Márkus, L. and Zempléni, A. (2011) Autocopulas: investigating the interdependence structure of stationary time series. *Methodology and Computing in Applied Probability*, **14**, p.149-167.
- Rakonczai, P. (2011) `Package 'mgpd' manual`. see http://cran.r-project.org/web/packages/mgpd/mgpd.pdf
- Rakonczai, P. and Tajvidi, N. (2010) On prediction of bivariate extremes. *International Journal of Intelligent Technologies and Applied Statistics*, **3**(2), p.115-139.
- Rakonczai, P., Butler, A. and Zempléni, A. (2010) Modeling temporal trend within bivariate generalized Pareto models of logistic type. (Technical report, HPC-Europa2 Project, Edinburgh, UK, available at http://www.math.elte.hu/∼paulo/pdf/)
- Rakonczai, P. (2009) On Modelling and Prediction of Multivariate Extremes, with applications to environmental data. Centrum Scientiarum Mathematicarum, Licentiate Theses in Mathematical Sciences 2009:05
- Rakonczai, P., Márkus L. and Zempléni, A. (2008a) Goodness of Fit for Auto-Copulas: Testing the Adequacy of Time Series Models, *Proceedings of the 4th International Workshop in Applied Probability* CD-ROM, paper No.73., 6 pages, Compiegne, France
- Rakonczai, P., Márkus L. and Zempléni, A. (2008b) Adequacy of Time Series Models, Tested by Goodness of Fit for Auto-Copulas, *Proceedings of the COMPSTAT2008 conference*, Porto, Portugal
- Rakonczai, P. and Zempléni, A. (2007) Copulas and goodness of fit tests. *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific, Hackensack, NJ, p.198-206.
- Bozsó, D., Rakonczai, P. and Zempléni, A. (2005) Árvizek a Tiszán és néhány mellékfolyóján. *Statisztikai Szemle*, **83**(10-11), p.919-936.

Hivatkozások

[1] Genest, C. Quessy, J.-F. and Rémillard, B. (2006) Goodnes-of-fit Procedures for Copula Models Based on the Integral Probability Transformation. *Scandinavian J. of Statistics*, **33**, p.337-366.

[2] Hall, P. and Tajvidi, N. (2004) Prediction regions for bivariate extreme events. *Aust. N. Z. J. Stat.* **46**(1), p.99-112.

[3] de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. Probab.*, **12**, p.1194-1204

[4] de Haan, L. and Pickands, J. (1986) Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, **72**, p.477-492.

[5] de Haan, L. and Resnick, S. I. (1987) On regular variation of probability densities. *Stochastic Processes and Their Applications*, **25**, p.83-93

[6] Krusper (2011) Többdimenziós extrém érték eloszlások alkalmazása biztosítási és pénzügyi adatokra. *MSc thesis*, Eötvös Loránd University, Budapest (in Hungarian) (http://www.cs.elte.hu/blobs/diplomamunkak/mat/2011/krusper_marta.pdf)

[7] Pickands, J. (1981) Multivariate extreme value distributions. *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session*: p.859-878.

[8] Rootzén, H. and Tajvidi, N. (2006) The multivariate generalized Pareto distribution. *Bernoulli* **12**, p.917-930.

[9] Sklar, A. (1959) Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, p.229-231.

[10] Vatan, P. (1985) Max-infinite divisibility and max-stability in infinite dimensions. *Probability in Banach Spaces V.*, ed. A. Beck et al. Lecture Notes in Mathematics 1153, Springer, Berlin, p.400-425.