



BUDAPESTI CORVINUS EGYETEM
EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Szakdolgozat

Öngerjesztő pontfolyamatok alkalmazása extrém árvízkárok modellezésére

Készítette:

NAGYBALI TAMÁS

Biztosítási és pénzügyi matematika MSc.
aktuárius specializáció

Konzulens:

ARATÓ MIKLÓS

egyetemi docens
Valószínűségelméleti és Statisztika
Tanszék

Budapest, 2025

Tartalomjegyzék

1. Bevezetés	2
2. Adatbázis bemutatása	6
3. Elméleti háttér	8
3.1. Klasszikus elmélet	9
3.2. Extrém értékek leírása pontfolyamatokkal	13
3.3. Jelölt pontfolyamatok	16
3.3.1. Matematikai háttér	16
3.3.2. Likelihood-függvény kiszámítása MPP modellre	18
4. SEMPP modell	21
4.1. A modell felírása	21
4.2. A jelek sűrűségfüggvénye - EGPD	23
4.3. Modellillesztés, illeszkedésvizsgálat, klaszterreprezentáció, szimuláció .	25
4.3.1. A loglikelihood-függvény felírása	25
4.3.2. Szimuláció a klaszterreprezentáció segítségével	27
5. Modellezés eredményei	28
5.1. SEMPP modell illesztése	28
5.1.1. Az alapintenzitás-függvény meghatározása	30
5.1.2. Becsült paraméterek	31
5.2. Illeszkedésvizsgálat, összehasonlítás a klasszikus modellel	33
6. Szimulációs eredmények, viszontbiztosítási példa a modell alkalmazására	35
6.1. Szimuláció	35
6.2. Viszontbiztosítás	39
6.2.1. CatXL és ECOMOR szerződések	39
6.2.2. Szerződések árazása	40
7. Összefoglalás	43

1. Bevezetés

A biztosítás, mint a különböző kockázatok elleni kooperatív védekezés egy formája, megkerülhetetlen része mindennapi életünknek. Az emberiség már az ókorban felismerte, hogy cselekvéseiket és vállalkozásaikat kísérő - első sorban anyagi - bizonytalanság ellen célszerű együttesen védekezni. A biztosítás gyakran emlegetett korai példái közé tartoznak az időszámításunk előtti második és harmadik évezredben elsőként a kínaiak és az indiaiak által alkalmazott praktikák (Vaughan 1996). Az áruszállítás veszélyei okán kereskedők portékájukat több hajón helyezték el, ezzel védve magukat teljes vagyonuk esetleges elvesztésétől. A nagy károk elleni védekezésről számtalan ehhez hasonló, némileg továbbfejlesztett kezdetleges biztosítási példát találunk a történelemben, melyekről például Longnaker cikkében olvashatunk (Longnaker 1962). Biztosítással tehát olyan kockázatok is kezelhetővé tehetünk, melyeknek pénzügyi következményei túl nagy terhet rónának az egyénre. Közgazdasági fogalmakkal élve egy modern biztosítási rendszerben a gazdaság egyes ágensei a kockázatporlasztás érdekében (homogén) veszélyközösséget formálnak, egy biztosító pedig egy előre, statisztikai módszerekkel meghatározott díj ellenében átvállalja tőlük a kockázat nagy részét. A biztosítótársaságok - legyen szó élet-, nem-élet- vagy kompozit biztosítókról - a pénzügyi piacok fontos szereplői, hiszen működésük során a gazdaság kis tőkéit összegyűjtik, és nagy, befektethető tőkévé akumulálják. Fejlett piacgazdaságokban a biztosítók a legnagyobb befektetők közé tartoznak, így gazdaságélénkítő és munkahelyteremtő képességük nem elhanyagolható (Banyár 2016). A pénzügyi piacok szerves részét alkotó entitásokként tehát inszolvenciájuk nagyon súlyos következménnyel járhat egy adott ország, régió, de az egyre erősebb globalizáció és a komplex viszontbiztosítási piac következményeként az egész világ gazdaságára (McDonald és Paulson 2015). Ezt elkerülendő, a biztosítók igyekeznek megfelelően diverzifikálni portfóliójukat, valamint adekvát tartalékolással garantálni a prudens működést.

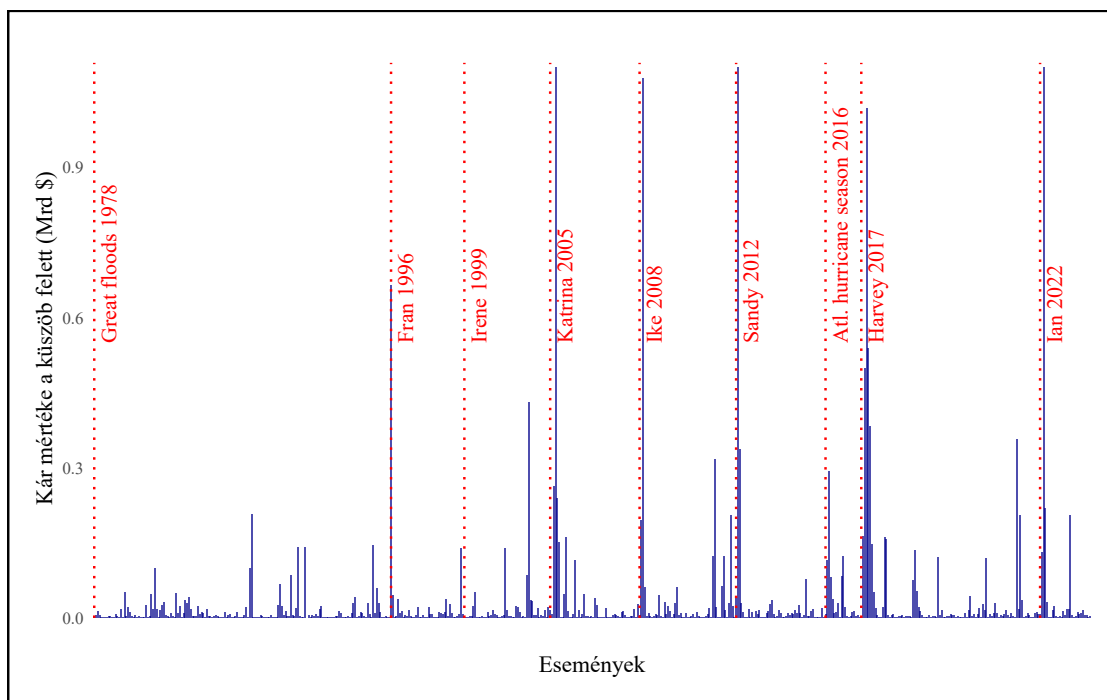
Az előző bekezdésben leírtakból és a biztosítási piac jellegéből adódóan biztosításmatematikai szempontból a kiemelkedően nagy károk leírása, számszerűsítése, valószínűségszámítási és statisztikai modellezése a terület legfontosabb kérdései közé

tartoznak. Egy jól diverzifikált, stabil portfóliót is érhetnek olyan károk (elsősorban nem-életbiztosítási területen), melyek veszélyt jelenthetnek a biztosító szolvenciájára. Jellemzően ilyen, a portfólió nagy részére kiterjedő, valamint időben csoportosuló károkat különböző természeti katasztrófák idézhetnek elő. Általában 10 fajta természeti katasztrófát különböztetünk meg, ezek a cunami, hurrikán, földrengés, jégeső, vulkánkitörés, tornádó, hóihar, földcsuszamlás, futótűz és árvíz. A felsorolt események (és általánosságban az extrém károk) elleni pénzügyi védekezés kiemelt fontosságú a biztosítótársaságok szempontjából, melynek adekvátságát szabályozó szervek ellenőrzik, a követendő irányelveket a Szolvencia II szabályozásban rögzítették (EIOPA 2020). Az inszolvenca elkerülése érdekében a biztosítók portfóliójuk egy részére viszontbiztosítási szerződéseket kötnek, azaz adott díj fejében szolgáltatásuk egy részét átvállalják más biztosítótársaságok. Klasszikus formájában a viszontbiztosítás védelem a kifizetendő kárösszeg nem várt növekedése ellen, megjelenése nagyjából egyidős a szervezett biztosítótársaságok létrejöttével (Kerényi 2011). Az első viszontbiztosító társaság az 1842-es nagy hamburgi tűzvész okán, 1846-ban jött létre Kölnben, mely jól szemlélteti, hogy nem tárgyalhatjuk a természeti katasztrófák biztosítási piacra gyakorolt hatását a viszontbiztosítás említése nélkül.

A megszokottnál jóval nagyobb károk elemzése napjainkban mind a biztosításban, mind a pénzügyi piacokon bevett gyakorlat. Ennek eszköze az extrémértékelmélet, melynek fő úttörői Leonard H. C. Tippett, Emil J. Gumbel és Sir Ronald A. Fisher voltak. Az egyváltozós elmélet alaptételének tekinthető a *Fisher-Tippett-Gnedenko tétel*, mely a független, azonos eloszlású valószínűségi változókból kapott rendezett statisztikák aszimptotikus viselkedését karakterizálja (Fréchet 1927; Fisher és Tippett 1928; Mises 1936; Gnedenko 1943). A klasszikus elméletet kiválóan összefoglalják Resnick (2008), Embrechts és tsai. (1997), Coles (2001) és Haan és Ferreira (2006) munkái, az elméleti háttér bevezetésénél ezen könyvek tárgyalásmenetét követjük. Az extrémérték-elemzés szintén elterjedt módszertan a mérnöki tudományokban, a közgazdaságtanban, a földtanban, hidrológiában, a biológiában és a virológiában, valamint a kiberbiztonság és kommunikációs hálózatok, sőt a sportok területén is (Coles 2001; Reinhart 2018; Castillo-Mateo és tsai. 2023). Ezen dolgozatban az árvízkárok extrémérték-modellezésével foglalkozunk. Azt vizsgáljuk, hogy

egy biztosítási portfóliót érő károk eloszlásának jobboldali farka hogyan viselkedik. Az elemzés során az USA National Flood Insurance Program (NFIP) által biztosított ingatlanok kárbejelentés adatbázisát használjuk. Az adatbázis az Amerikai Egyesült Államok egész területén történt árvízkárokról tartalmaz napi adatokat 1978 január és - a letöltés időpontjában - 2024 augusztus között.

Az extrémérték-elemzés elméletét régóta alkalmazzák természeti katasztrófák által okozott károk eloszlásának modellezésére. A klasszikus elméletnek azonban természetesen vannak korlátai. Feltételezi a kár nagyságok függetlenségét és azonos eloszlását, de ahogy az korábban említésre került, a természeti katasztrófák okozta károkból gyakran megfigyelhető az extrém események időbeli klasztereződése, valamint logikus, hogy például különböző magnitúdójú földrengések várhatóan eltérő nagyságú károkat okoznak. Ezt a megfontolást támasztja alá az alábbi ábra, melyen az általunk használt adatbázis extrém kárai szerepelnek:



1. ábra. Extrém árvízkárok időbeli eloszlása és nagysága. A piros szaggatott vonalak az NFIP bejelentései alapján a 10 legnagyobb kárt okozó természeti katasztrófa kezdetének időpontjait jelzik

Az u küszöbérték meghatározását, ami felett egy kárt extrémnek tekintünk, a későbbi fejezetekben fejtjük ki. Láthatjuk, hogy az extrém károk nem egyenletesen, hanem - főként a nagyobb természeti katasztrófák körül - klasztereződve következnek be, valamint a nagyságaik is igen eltérőek (megjegyzendő, hogy a könnyebb átláthatóság érdekében az ábrán a függőleges tengelyen 1.2 milliárd USD-ig szerepelnek a küszöbérték feletti kárnagyságok, azonban az Ian, Sandy és Katrina hurrikánok következtében ennél magasabb volt a maximális érték, utóbbi esetében például 6.12 milliárd USD). Így tehát azt mondhatjuk, hogy általában mind a függetlenség, mind az azonos eloszlás feltétele sérül. Ez természetesen azt jelenti, hogy a *Fisher-Tippett-Gnedenko tétel* aszimptotikus eredményei sem lesznek érvényesek, bár Leadbetter és tsai. (1983) bebizonyítják, hogy bizonyos enyhe keverő feltételek mellett stacionárius idősorokra is kiterjeszthető az elmélet, azonban ezen feltételek ellenőrzése a gyakorlatban igen nehéz (Embrechts és tsai. 1997). Ezt a leggyakrabban úgy orvosolják, hogy a rendelkezésre álló idősort deklaszterezik, azaz ha például egy adott küszöbérték felett egymás után három érték is található, akkor csak a legnagyobbat tartják meg, a másik kettőt a küszöbvel teszik egyenlővé. Ez a módszer természetesen jelentős információvesztéssel jár, valamint a küszöbérték feletti megfigyelések értékeinek csonkításával a deklaszterezett adatokra illesztett modell hajlamos a frekvenciát és kárnagyságokat mozgató paraméterek alábecsülésére. Ezen megfontolások alapján ebben a dolgozatban a pontfolyamatok eszköztárát hívjuk segítségül. Az extrém értékek pontfolyamat karakterizációja egyesíti a tradicionális Block-maxima és a küszöbérték túllépések (Peaks-Over-Threshold, POT) módszereket (Coles 2001), így használata igencsak elterjedt. Azonban a független, azonos eloszlás itt is feltétel, így nagyobb általánosításra lesz szükségünk. A dolgozat gerincét egy sztochasztikus öngerjesztő jelölt pontfolyamat (Self-exciting Marked Point Process, a továbbiakban SEMPP) modell adja, melyhez a kárnagyságokról semmilyen feltételezéssel nem kell élnünk. A pontfolyamatok elméletének egy összefoglalóját megtalálhatjuk például Daley és Vere-Jones (2006) munkájában. Elemzésünk nagyban támaszkodik Li és tsai. (2021) tanulmányára, melyben a kanadai Ottawa folyó Quebec tartományban található Rivière des Mille Îles nevezetű csatornájának kritikus érték alatti vízszintjének eloszlását vizsgálják SEMPP modellezéssel. A vizsgált adatbázisról számtalan

elemzés és tanulmány készült, azonban az itt alkalmazott módszertan legjobb tudomásunk szerint egy újfajta megközelítése a témának.

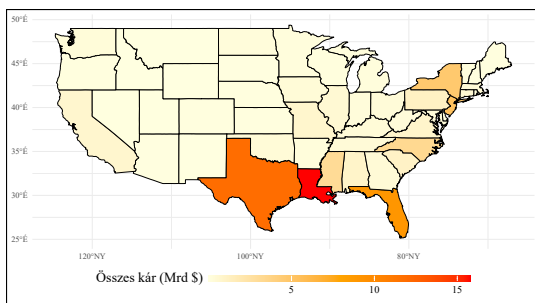
A dolgozat felépítése a következő: a második fejezetben röviden bemutatásra kerül a használt adatbázis, valamint az átalakítások, amelyeket az idősor létrehozásához végzünk. A harmadik fejezetben ismertetjük a klasszikus extrém érték elmélet számunkra leginkább releváns definícióit és tételeit, majd rátérünk az SEMP modellezéshez szükséges fogalmakra. A negyedik fejezet során felírjuk a modell alapintenzitás-függvényének alakját, a jelek sűrűségfüggvényét, majd a maximalizálható közös likelihood-függvényt és bemutatjuk az illeszkedésvizsgálatnál használt módszerek mellett a szimulációs elemzéshez tartozó algoritmust. Az ötödik fejezetben ismertetjük a modell illesztésének menetét valamint a becsült paramétereket és az illeszkedésvizsgálatok eredményét. A hatodik fejezetben rátérünk a modell egy potenciális alkalmazására, a viszontbiztosításra, ahol szimuláció segítségével vizsgáljuk CatXL és ECOMOR viszontbiztosítási szerződések pénzáramlásait, és várhatóértékkel meghatározzuk a viszontbiztosítás díját különböző paraméterek mellett. A dolgozatot egy összefoglaló fejezet zárja, mely a legfontosabb eredmények és következtetések tárgyalása mellett kitér a modell továbbfejlesztéseinek lehetőségeire és az egyéb felmerülő kutatási kérdésekre. Az elemzést RStudio-ban végezzük főként a **PtProcess** (Harte 2010) és **extRemes** (Gilleland és Katz 2016) csomagokra támaszkodva. Az alkalmazott függvények nagy része saját készítésű, mivel amellet, hogy a **PtProcess** csomag könnyen használható környezetet nyújt MPP modellek illesztésére, diszkrét időskálán történő modellezést nem támogatja.

2. Adatbázis bemutatása

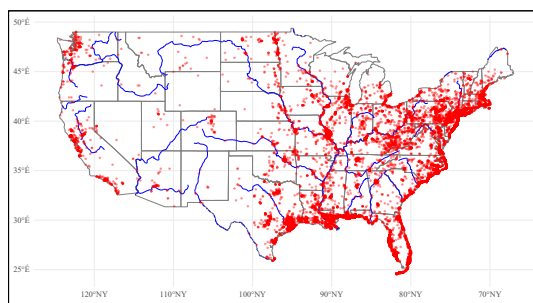
A National Flood Insurance Program (NFIP) az amerikai Federal Emergency Management Agency (FEMA) által irányított program, melynek létrehozásáról az USA kongresszusa 1968-ban döntött. Az NFIP két fő célja az árvízkárok megosztása a veszélyeztetett területen élők között, valamint a potenciális árvízkárok csökkentése az ártéri építkezések terjedésének korlátozásával (Horn és Webel 2018). A program lehetővé teszi, hogy helyi közösségek közösen vásároljanak árvízkárok elleni biztosítást.

Jelenleg több mint 23 000 ilyen kisebb-nagyobb konglomerátum része a programnak. A program különböző biztosítási csomagjait több mint 50 biztosító árulja. Az adatbázis részletes ismertetése és elemzése megtalálható Dombrowski és tsai. (2021), valamint Horn és Brown (2017) munkáiban.

A FEMA oldalán nyilvánosan elérhető adatbázis a letöltés pillanatában több mint 2.6 millió káreseményt tartalmazott az USA összes tagállamából, az egyes sorokhoz 73 változó tartozik. A biztosított ingatlant ért kár-, és kárkifizetés, valamint a káresemény időpontján kívül megtalálhatóak benne többek között az ingatlan jellegét, használatának módját, a kárt okozó természeti katasztrófa nevét, az árvíz típusát, a hosszúsági és szélességi - az anonimitás megőrzése érdekében 3 tizedesjegyre kerekítve - fokok értékét tartalmazó oszlopok. Az adatbázist havi rendszerességgel, 1-2 hónap késleltetés mellett frissítik, a károk és a fizetett kárkövetelések dollárban értendők. A dolgozat során az árapályvíz, valamint patak, folyó vagy tó áradása okozta károkat modellezzük. A káreseményeket első körben napi szinten aggregáljuk; a vizsgált intervallumban összesen 11 143 napon történt káresemény a 17 032-ből. A könnyebb kezelhetőség érdekében a kárnagyságokat az elemzés során milliárd dollárra konvertáljuk. Az alábbi két ábra az államonként aggregált kárnagyságokat és az egyedi károk földrajzi elhelyezkedését tartalmazza (utóbbi esetében 25 000 elemszámú mintát veszünk, hogy ábrázolható legyen):



(a) Összesített kárnagyság államonként



(b) Az egyes árvízkárok földrajzi elhelyezkedése

2. ábra. Az árvízkárok aggregált értéke 1978 január és 2024 augusztus között (balra) és az egyes károk földrajzi elhelyezkedése (jobbra)

Látható, hogy a vizsgált időszakban a déli parti államokat, legfőképp Texast, Louisiana-t és Floridát érték a legnagyobb károk. A jobboldali ábra alapján az árvízkárok az előbb említett államok mellett főként a nagyobb folyók menti területekre és a keleti partra összpontosulnak. A nyugati parton három nagyobb gócpont található, míg a szárazabb középső államokban csak elvétve történnek károk. Az átlagos napi kárnagyság 1 806 956 dollár, a medián 17 247 dollár, a maximális egy napra eső kárnagyság 6.2 milliárd dollár, az adatsor tehát erősen balra dőlő és jobbra elnyúló. A Ljung-Box teszt szerint a függetlenség hipotézisét minden szignifikanciaszint mellett elvethetjük. Ami a stacionaritást illeti, az ADF teszt alapján a folyamat stacionárius, azonban a KPSS 0.0478-as p-értékét tekintve ez az állítás megkérdőjelezhető. Az extrém értékek időbeli függőségi kapcsolatát Reiss és Thomas (2007) által bevezetett - itt nem ábrázolt - úgynevezett *auto-tail dependence* - *ATD* függvény segítségével is ellenőrizhetjük. Amennyiben az ATD függvény értékei különböző késleltetések mellett szignifikánsan eltérnek nullától, akkor az extrém értékek időben korrelálnak. Ahogy az várható, az általunk vizsgált idősor esetében pontosan ezt tapasztalhatjuk.

3. Elméleti háttér

Ebben a fejezetben ismertetjük az extrémérték-elemzéshez kapcsolódó legfontosabb definíciókat és tételeket. A fejezet három részből tevődik össze. Az első alponthozban a klasszikus elméletet tárgyaljuk, a teljesség igénye nélkül bevezetjük a valószínűségi változók sorozatán vett maximumhoz kapcsolódó alapvető fogalmakat és az extrémérték-eloszlásokat, valamint a POT módszerét. Ezután definiáljuk a pontfolyamatokat, megadjuk az extrém értékek pontfolyamat karakterizációját és bemutatjuk kapcsolatát a klasszikus módszerekkel. A harmadik alfejezetben bevezetjük a jelelt pontfolyamatot (Marked Point Process, innen MPP) és bemutatjuk alkalmazhatóságát az extrémérték-modellezésben. Az általános elméletet folytonos modellek esetében vezetjük be, mivel ezek bemutatása után könnyebben származtathatóak az analóg diszkrét definíciók és számítások. Az eloszlásfüggvényeknél az angolszász konvenciót követjük, azaz X valószínűségi változó esetében $F(x) = P(X \leq x)$.

3.1. Klasszikus elmélet

Az extrémérték-elemzésen belül alapvetően két módszert különböztetünk meg, ezek a blokk-maxima és szinttúlépések módszerei. Előbbinél felosztjuk a megfigyelt intervallumot egyenlő méretű, nem-átfedő blokkokra, és figyelmünket az egyes blokkok maximális értékeire fordítjuk (Ferreira és Haan 2015). Megfelelő feltételek mellett ezen értékek egy extrémérték-eloszlást követnek, melyeket rövidesen bemutatunk. A módszer hátránya természetesen a jelentős információvesztés. A POT módszer esetében azokra az értékekre fókuszálunk, melyek meghaladtak egy adott küszöbértéket. Megfelelő feltételek mellett ezen küszöb feletti változók eloszlása általánosított Pareto eloszlást - GPD - követ. Az általunk használt modell ehhez áll közelebb. A továbbiakban $X, X_1, X_2, \dots$ nem-elfajuló, független, azonos eloszlású (IID) valószínűségi változókat jelöl. A mintabeli maximumot az alábbi módon jelöljük:

$$M_1 = X_1, \quad M_n = \max(X_1, \dots, X_n).$$

Első lépésben ennek a maximumnak az aszimptotikus viselkedését leíró, már említett alaptétel kerül bemutatásra. Felmerül a kérdés, hogy lehetséges-e a maximum eloszlásáról a centrális határeloszlás tételhez hasonló állítást megfogalmazni. A válasz igenlő, megfelelő (c_n) és (d_n) normáló konstansok sorozatának segítségével.

1. Tétel. (*Fisher-Tippett-Gnedenko, a maximum határértékének tétele*)

Legyen (X_n) IID valószínűségi változók sorozata. Ha léteznek $(c_n) > 0$ és $(d_n) \in \mathbb{R}$ normáló konstansok, valamint egy nem elfajult H eloszlásfüggvény úgy, hogy

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

Akkor H az alábbi három eloszlástípus egyikéhez tartozik:

1. *Fréchet:*

$$\Phi_\alpha(x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ \exp\{-x^{-\alpha}\}, & \text{különben} \end{cases} \tag{2}$$

2. *Weibull*:

$$\Psi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & \text{ha } x \leq 0 \\ 1, & x > 0 \end{cases} \quad (3)$$

3. *Gumbel*:

$$\Lambda_\alpha(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}. \quad (4)$$

Fontos megjegyezni, hogy annak ellenére, hogy modellezési szempontból a három eloszlás eltérő, könnyen lehet közöttük matematikai kapcsolatot találni. Valóban, ha X pozitív valószínűségi változó, akkor

$$X \sim \Phi_\alpha \Leftrightarrow \ln(X^\alpha) \sim \Lambda \Leftrightarrow -X^{-1} \sim \Psi_\alpha.$$

Az **1. Tételben** szereplő eloszlásokat sztenderd extrémérték-eloszlásoknak nevezzük. Az extrémérték-eloszlások részletesebb vizsgálatára nem térünk ki, ez megtalálható Resnick (2008) és Embrechts és tsai. (1997) könyveiben. Bevezethető azonban $(H_\xi)_{\xi \in \mathbb{R}}$ eloszlásfüggvények egy parametrikus családja, mely tartalmazza a sztenderd extrémérték-eloszlásokat:

$$H_\xi = \begin{cases} \Phi_{\frac{1}{\xi}} & \text{ha } \xi > 0, \\ \Lambda & \text{ha } \xi = 0, \\ \Psi_{-\frac{1}{\xi}} & \text{ha } \xi < 0. \end{cases} \quad (5)$$

Ezt az eloszlásfüggvényt *általánosított extrémérték-eloszlásnak* hívjuk (innentől GEV, ami az angol Generalised Extreme Value rövidítése). A következő a legszélesebb körben elfogadott definíció a szakirodalomban:

1. Definíció. (*Jenkinson-Von Mises általánosítás*)

Definiáljuk a H_ξ eloszlásfüggvényt a következőképpen:

$$H_\xi(x) = \begin{cases} \exp\left\{-(1+\xi x)^{-\frac{1}{\xi}}\right\} & \text{ha } \xi \neq 0, \\ \exp\{-\exp\{-x\}\} & \text{ha } \xi = 0, \end{cases} \quad (6)$$

ahol $1 + \xi x > 0$.

A fenti H_ξ helyére bevezethetjük a lokáció-skála paraméterezett $H_{\xi;\mu;\psi}$ eloszlás-családot az x helyébe $(x - \mu)/\psi$ -t írva, $\mu \in \mathbb{R}$, $\psi > 0$ mellett. A $H_{\xi;\mu;\psi}$ függvényre is GEV-ként hivatkozunk. A H_0 eloszlásfüggvényt $\lim_{\xi \rightarrow 0} H_\xi$ határértékként definiáljuk. Ezzel az értelmezéssel a

$$H_\xi(x) = \exp\left\{-(1+\xi x)^{-\frac{1}{\xi}}\right\}, \quad 1 + \xi x > 0,$$

eloszlásfüggvény reprezentálja az extrém érték eloszlásokat minden $\xi \in \mathbb{R}$ esetén.

2. Definíció. (Szintmeghaladás eloszlásfüggvénye)

Legyen X valószínűségi változó x_F jobboldali végpontú F eloszlásfüggvényvel. Rögzített $a < x_F$ -re az

$$F_u(x) = P(X - u \leq x | X > u), \quad x \geq 0, \quad (7)$$

képlettel definiált függvényt az X változó u küszöb melletti szintmeghaladás eloszlásfüggvényének, az

$$e(u) = E(X - u | X > 0)$$

függvényt pedig az átlagos szintmeghaladási függvénynek hívjuk.

Az u küszöb melletti feltételes eloszlásfüggvény esetében Balkema és Haan (1974), valamint Pickands (1975) megmutatják, hogy amennyiben $X_1, X_2, \dots$ valószínűségi változók F eloszlásfüggvénye egy H_ξ extrémérték-eloszlásfüggvény úgynevezett maximális vonzási tartományába tartozik, akkor $u \rightarrow \infty$ mellett az F_u eloszlása aszimptotikusan az Általánosított Pareto eloszláshoz tart. Ennek definíciója az alábbi.

3. Definíció. (Általánosított Pareto eloszlás, GPD)

Definiáljuk a G_ξ eloszlásfüggvényt az alábbi módon:

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}} & \text{ha } \xi \neq 0, \\ 1 - e^{-x} & \text{ha } \xi = 0. \end{cases} \quad (8)$$

G_ξ -t sztenderd általánosított Pareto eloszlásnak (*Generalised Pareto Distribution*, innen *GPD*) hívjuk. A GEV-nél láttotakhoz analóg módon bevezethető a lokáció-skála paraméterezett $G_{\xi;\nu;\beta}$ eloszláscsalád az $x = \frac{x-\nu}{\beta}$, $\nu \in \mathbb{R}$, $\beta > 0$ helyettesítéssel. $G_{\xi;\nu;\beta}$ -ra is GPD-ként hivatkozunk.

Ahogy a GEV esetében, G_0 -t itt is $\lim_{\xi \rightarrow 0} G_\xi$ -ként értelmezzük. Ezzel

$$G_{\xi;\beta}(x) = 1 - (1 + \xi x)^{-\frac{1}{\xi}}, \quad x \in D(\xi; \beta), \quad (9)$$

Ahol

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

Összefoglalva tehát a GEV a normált maximum határeloszlását, a GPD pedig a küszöbérték feletti értékek eloszlását írja le. Az u küszöbértéknek megválasztása kiemelt fontosságú az extrémérték-elemzésben, mivel modell illesztés esetében egy variancia/torzítás trade-off szituációval találjuk szemben magunkat. Ha u értéke túl alacsony, akkor nem csak extrém értékek kerülnek az adatsorunkba, így a becsült paraméterek torzítottak lesznek, mivel nem teljesülnek az aszimptotikus feltételek. Ha pedig u túl magas, akkor az adatpontok hiánya miatt a becslés varianciája nagyra nőhet (Coles 2001). A küszöbérték megválasztásában az empirikus átlagos szintmeghaladási függvény (8) ábrázolása segíthet. általánosított Pareto eloszlás $e(u)$ lineáris, így olyan intervallumon kell küszöbértéket választanunk, ahol $\bar{e}(u)$ vonala nagyjából egyenes.

Embrechts és tsai. (1997) 3.1.1-es állítása X_n független azonos eloszlású valószínűségi változók maximumának határeloszlásáról fogalmaz meg egy tételt a túlélési függvény és alkalmas (u_n) valós sorozat segítségével. Szintén ezen könyv 8.1-es al-

fejezetében több olyan példát is találhatunk, melyben ezen IID változókból képzett - stacionárius - valószínűségi változókra egy kicsivel eltérő aszimptotikus eredmény érvényes. Felmerül a kérdés, hogy bizonyos függőségi struktúra mellett milyen tétel fogalmazható meg a maximum határeloszlásáról. Erre ad választ az extrém index definíciója, mely a korábban említett állítás kiterjesztése erősen stacionárius idősorokra:

4. Definíció. *Extrém index*

Legyen X_n erősen stacionárius F eloszlásfüggvényű valószínűségi változók sorozata, θ pedig egy nemnegatív valós szám. Jelölje $\bar{F}$ az F eloszlásfüggvény túlélési függvényét, vagyis $\bar{F} = 1 - F$. Tegyük fel, hogy minden $\tau > 0$ esetén létezik u_n pozitív sorozat, hogy

$$\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$$

és

$$\lim_{n \rightarrow \infty} P(M_n \leq u_n) = e^{-\theta\tau}.$$

Ekkor θ -t a az idősor extrém indexének nevezzük.

Az extrém index intuitív módon karakterizálja a sorozat függőségi struktúrája és extrémérték-viselkedése közötti kapcsolatot. Értéke minden esetben a $[0, 1]$ intervallumon van, a $\theta = 0$ bizonyos értelemben patológiai eset, de ahogy θ tart a 0-hoz, úgy erősödik az összefüggés az extrém értékek között. Teljes függetlenség esetén θ értelemszerűen 1. Az index konkrét értéke többféleképpen értelmezhető, az egyik leggyakoribb és legkézenfekvőbb úgy tekinteni θ -ra, mint az átlagos klaszterméret reciproka. Az extrém index becslései módszereiről ad áttekintést Ferro és Segers (2003) munkája.

3.2. Extrém értékek leírása pontfolyamatokkal

Egy pontfolyamatra gondolhatunk úgy, mint X_i pontok véletlenszerű eloszlására a térben. Ebben az alponthan ezen folyamatokkal foglalkozunk, a tárgyalás Resnick (2008) és Embrechts és tsai. (1997) menetét követi. Első lépésben bevezetjük az

extrém értékek sztenderd pontfolyamat karakterizációját, majd az alapvető fogalmak bemutatása után kitérünk a jelölt pontfolyamatok elméletére. A továbbiakban az eddigiekhez hasonlóan, itt is az angol rövidítéseket használjuk.

A teret, melyben a pontok szóródhatnak állapot térnek hívjuk és E -vel jelöljük. Annak ellenére, hogy az itt leírtak általánosabb tereken is érvényesek, a dolgozat kontextusában érdemes E -re a pozitív félegyenes adott intervallumaként gondolni. E egy kompakt halmaz, jelölje $\mathcal{E}$ az E részhalmazaiából álló Borel σ -algebrát, ϵ_x pedig a Dirac mértéket $\mathcal{E}$ -n, vagyis

$$\epsilon_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \quad A \in \mathcal{E}.$$

Adott $(x_i)_{i \geq 1} \in E$ sorozatra legyen

$$m(A) = \text{card}\{i : x_i \in A\} = \sum_{i=1}^{\infty} \epsilon_{x_i}(A), \quad A \in \mathcal{E}.$$

$m(A)$ egy számláló mérték $\mathcal{E}$ -n, amelyet *pont mértéknek* hívunk, amennyiben $m(H) < \infty$, $\forall H \in \mathcal{E}$ halmazra. Jelölje $K_p(E)$ az E állapot téren értelmezett pont mértékek terét, és definiáljuk a $\mathcal{K}_p(E) := \sigma\{m \in K_p(E) : m(F) \in B\}$ σ -algebrát, tetszőleges $F \in \mathcal{E}$ és $B \in \mathcal{B}([0; \infty])$. mellett. Természetesen a céljainkhoz $B \in \mathcal{B}([0; \infty])$ a szükségesnél nagyobb általánosítás, de konzisztensek maradunk a szakirodalmakban leírtakkal. Ezek után már megadhatjuk a pontfolyamat definícióját:

5. Definíció. *pontfolyamatnak nevezzük adott E halmazon egy valószínűségi mezőből az*

$$N : (\Omega, \mathcal{F}, P) \rightarrow (K_p(E), \mathcal{K}_p(E))$$

mérhető leképezést.

Egy pontfolyamat *egyszerű*, ha $x \in E$ esetén $P(N(\{x\}) \leq 1) = 1$. A számunkra legfontosabb ilyen leképezés a szintmeghaladások pontfolyamata, melyet az alábbi

módon definiálhatunk:

$$N_n(\cdot) = \sum_{i=1}^n \epsilon_{\frac{i}{n}}(\cdot) I_{\{X_i > u\}}, \quad n = 1, 2, \dots, \quad (10)$$

Ennek a pontfolyamatnak az állapottere $E = (0, 1]$, és azt adja meg, hogy az $X_1, \dots, X_n$ - ebben az alpontban még IID - valószínűségi változók sorozata hány-szor lépi túl az u szintet. Ezt könnyen párhuzamba állíthatjuk az első alpontban leírtakkal; jelölje $X_{n,n} \leq \dots \leq X_{1,n}$ az $X_1, \dots, X_n$ változók rendezett mintáját, ekkor:

$$\begin{aligned} \{N_n(0, 1] < k\} &= \{\text{card}\{i \leq n : X_i > u\} < k\} \\ &= \{\text{Az } X_i \text{ változók közül kevesebb mint } k \text{ nagyobb } u\text{-nál}\} \\ &= \{X_{k,n} \leq u\}. \end{aligned}$$

Ha egy kísérletet sokszor megismétlünk, akkor a "sikeres" kimenetek száma binomiális eloszlású, n és p paraméterekkel, azaz $\text{Binom}(n, p)$. A Poisson-eloszlás egyik alapvető tulajdonsága, hogy előáll a binomiális eloszlás határértékeként n végtelenbe tartása mellett, feltéve, hogy az np szorzat konstans marad. Hasonló állapítható meg a pontfolyamatok esetében is, ahol esetünkben a sikert egy adott küszöb meghaladása jelenti. A következő definícióban Λ egy Radon mérték $\mathcal{E}$ -n, azaz $\Lambda(A) < \infty$ minden $A \subset E$ kompakt halmazra.

6. Definíció. (Poisson pontfolyamat)

Egy N pontfolyamatot Λ paraméterű Poisson pontfolyamatnak hívunk ($\text{PPM}(\Lambda)$ -vel jelöljük), ha a következő két feltétel teljesül:

(a) $A \in \mathcal{E}$ esetén

$$P(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{(\Lambda(A))^k}{k!} & \text{ha } \Lambda(A) < \infty, \\ 0 & \text{ha } \Lambda(A) = \infty, \end{cases} \quad k \geq 0. \quad (11)$$

(b) Ha $A_1, \dots, A_m$ páronként diszjunkt halmazok $\mathcal{E}$ -ben, akkor $N(A_1), \dots, N(A_m)$

független valószínűségi változók.

A folyamat intenzitásfüggvényén a Λ mérték Radon-Nikodym deriváltját értjük (vagyis $\forall A \subset \mathcal{E}$ halmazra $\Lambda(A) = \int_A \lambda(z) dz$), melyet λ -val jelölünk. A Poisson pontfolyamatra az SEMPP modell illeszkedésvizsgálatánál lesz szükségünk. A szintmeghaladások pontfolyamata enyhe feltételek mellett gyengén konvergál a Poisson pontfolyamathoz (Resnick 2008).

3.3. Jelölt pontfolyamatok

3.3.1. Matematikai háttér

Sok esetben nem csak adott események időbeli eloszlása érdekel minket. Káresemények bekövetkezésénél fontos a kárnagyságok eloszlásának vizsgálata, a biztosító portfólióját fenyegető kockázatok pontosabb felmérése érdekében. Az előző alpontban leírtak érvényesek maradnak abban az esetben is, ha a $Z_1, Z_2, \dots$ valószínűségi változók általánosabb állapottéren szóródnak. Tegyük fel, hogy minden $n \in \mathbb{N}$ -re, $Z_n = (T_n, M_n)$, ahol $T_n \in [0, T]$ jelzi az extrém káresemény bekövetkezésének időpontját, $M_n \in \mathcal{M} = (u, \infty)$ pedig a kár nagyságát egy előre meghatározott u küszöbérték felett. Bevezethető az alábbi definíció:

7. Definíció. (Jelölt pontfolyamat)

Jelölt pontfolyamatnak (MPP) nevezzük az olyan két vagy többdimenziós pontfolyamatot, melyben az időbeli pontok az $\mathcal{X}$, a "jelek" pedig az $\mathcal{M}$ teljes szeparábilis metrikus terek pontjai, tehát egy olyan pontfolyamatról beszélünk, melyben a (x_i, m_i) párok az $\mathcal{X} \times \mathcal{M}$ térben helyezkednek el, azzal a megkötéssel, hogy az alapfolyamat $N_g(\cdot)$ maga is egy pontfolyamat, vagyis korlátos $A \in \mathcal{B}_{\mathcal{X}}$ -re, $N_g(A) = N(A \times \mathcal{M}) < \infty$.

Nem minden szorzattéren értelmezett pontfolyamat jelölt pontfolyamat, mivel például a kétváltozós Poisson folyamat $\mathbb{R}^2$ -en 1 valószínűséggel nem korlátos minden Borel halmazon (Daley és Vere-Jones 2006). Egy MPP egyszerű, ha az N_g alapfolya-

mata egyszerű. A következő állítás az $E = [0, 1] \times \mathcal{M}_{\mu, \sigma, \xi}$ állapottéren értelmezett

$$N_n = \sum_{i=1}^n \epsilon_{\left(\frac{i}{n}, \frac{X_i - a_n}{b_n}\right)}$$

MPP gyenge konvergenciáját írja le, mely Li és tsai. (2021) munkájában található 1. állítás.

1. Állítás. *Legyen $X_1, X_2, \dots$ IID valószínűségi változók sorozata megfelelő $H_{\mu, \sigma, \xi}$ eloszlásfüggvénnyel. Ekkor, $n \rightarrow \infty$ mellett a fenti N_n sorozat gyengén konvergál egy, a (11)-es egyenletben definiált N Poisson pontfolyamathoz Λ intenzitásfüggvénnyel, melyet a $\Lambda(A) = -(t_2 - t_1) \ln \{H_{\mu, \sigma, \xi}(x)\}$ egyenlőség ad, tetszőleges $A = [t_1, t_2] \times (x, m_2]$ halmazra, $0 \leq t_1 < t_2 \leq 1$ és $x \in (m_1, m_2)$ mellett.*

Az **1. Állításból** közvetlenül levezethető, hogy a jelölt pontfolyamatok keretrendszere magában foglalja a tradicionális blokk-maxima és szinttúlépési módszereket (Li és tsai. 2021). Ahogy azt a bevezetésben is említettük, a használt adatbázisban napi adatok állnak rendelkezésünkre. Az eddig bevezetett pontfolyamatok mind folytonos időben értelmezettek. Amennyiben az időt diszkrét skálán mérjük - például $\mathbb{N}$ -en - az állapot tér megszámlálható, és egy $A \subseteq \mathcal{E}$ halmaz akkor és csak akkor kompakt, ha véges. A $T_1, T_2, \dots$ változók ez esetben majdnem biztosan eltérő véletlen időpontok $\mathcal{E}$ -ben. Így a szintmeghaladások jelölt pontfolyamatának N_g alapfolyamata reprezentálható $Y_1, Y_2, \dots$ valószínűségi változók sorozatával, ahol $Y_n = 1$ amennyiben esemény történt az $n \in \mathcal{E}$ időpontban, és $Y_n = 0$ különben.

Egy N jelölt pontfolyamatot egyértelműen meghatároz az alapintenzitás-függvénye valamint a jelek eloszlásfüggvénye, lásd Daley és Vere-Jones (2006) 7.3.IV állítás. Az alapintenzitás-függvény leírja egy adott esemény bekövetkezésének gyakoriságát. Fontos tulajdonsága, hogy tetszőleges t időpontban nem csak t értéke, hanem a folyamat alakulása valamely $t_i < t$ időpontig bezárólag is befolyásolhatja. A továbbiakban jelölje H_t a megfigyelt események halmazát a t időpontig, vagyis $H_t = \{(t_i, m_i) \mid \forall i : t_i < t\}$, $\mathcal{H}_t$ pedig $N(t, m)$ természetes filtrációját, és legyen

$\mathcal{H}_{t-} = \sigma\{N(s, m) : s < t\}$. Egy jelölt pontfolyamat feltételes intenzitás-függvényét a

$$\lambda(t, m \mid \mathcal{H}_{t-}) = \lim_{\delta, \eta \rightarrow 0} \frac{1}{\delta\eta} P(N_{\delta, \eta}(t, m) > 0 \mid \mathcal{H}_{t-}) \quad (12)$$

egyenlettel definiáljuk, ahol $N_{\delta, \eta}(t, m)$ a $[t, t+\delta) \times [m, m+\eta)$ halmazba eső események számát adja meg. Az alapintenzitás-függvény pedig a

$$\lambda_g(t, \theta_1, \dots, \theta_m \mid \mathcal{H}_{t-}) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P(N_\delta(t) > 0 \mid \mathcal{H}_{t-}) \quad (13)$$

alakot ölti, ahol $\theta_1, \dots, \theta_p \in \Theta_p$ a paraméterek. Kényelmi okokból a paramétereket a továbbiakban elhagyjuk, és csak $\lambda_g(t \mid \mathcal{H}_{t-})$ -vel jelöljük. Egy jelölt pontfolyamat feltételes intenzitásfüggvénye mindig előáll az alapintenzitás-függvény és a jelek sűrűségfüggvényének szorzataként (Daley és Vere-Jones 2006), tehát

$$\lambda(t, m \mid \mathcal{H}_{t-}) = \lambda(t \mid \mathcal{H}_{t-}) f(m \mid \mathcal{H}_{t-}) \quad (14)$$

Természetesen a jelek feltételes sűrűségfüggvénye alatt is $f(m, \psi_1, \dots, \psi_q \mid \mathcal{H}_{t-})$ -t értjük, ahol $\psi_1, \dots, \psi_q \in \Psi_q$ a jelekhez tartozó paraméterek. A következő feladat egy jelölt pontfolyamat likelihood-függvényének kiszámítása.

3.3.2. Likelihood-függvény kiszámítása MPP modellre

Első körben levezetjük egy egyszerű folytonos pontfolyamat likelihood-függvényét, ahol az alapintenzitás-függvény csak az időtől függ, vagyis olyan alakú, mint a (13)-as egyenletben. A likelihood levezetéséhez szükséges az N folyamat regularitásának feltételezése, definícióért lásd Daley és Vere-Jones (2006). Jelölje tehát τ a t időpont előtti utolsó esemény időpontját, $\emptyset_{(\tau, t)}$ pedig a nulla kimenetet, tehát azt, hogy (τ, t) intervallumon nem történt esemény. Jelöljük W -vel a következő esemény feltételes eloszlását, azaz:

$$W(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau, t)}) = P\{T^* \geq t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau, t)}\},$$

$w(t \mid \mathcal{H}_t \cap \emptyset_{(\tau,t)})$ -vel pedig a hozzá tartozó feltételes sűrűségfüggvényt. Ekkor az alapintenzitás-függvény előáll az alábbi alakban:

$$\lambda_g(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)}) = \frac{w(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)})}{1 - W(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)})}.$$

A kapott differenciálegyenletet megoldva kapjuk, hogy

$$W(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)}) = 1 - \exp \left\{ - \int_\tau^t \lambda_g(u \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,u)}) du \right\}$$

Felhasználva a fenti összefüggést, az egyenletből kifejezhetjük a feltételes sűrűségfüggvényt:

$$w(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)}) = \lambda_g(t \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,t)}) \exp \left\{ - \int_\tau^t \lambda_g(u \mid \mathcal{H}_\tau \cap \emptyset_{(\tau,u)}) du \right\}$$

Vizsgáljuk most a loglikelihood-függvényt a $[0, T]$ intervallumon (általánosabban ezt megtehetnénk tetszőleges $[T_1, T_2]$ intervallumon is, ahol az intervallum előtti értékek explicit módon bekerülnek a likelihoodba). Legyenek $0 < t_1 < \dots < t_n < T$ az $i \in \mathbb{N}$ események megfigyelt időpontjai, ekkor

$$\begin{aligned}
\log L &= \log w(t_1 \mid \mathcal{H}_0 \cap \emptyset_{(0,t_1)}) + \sum_{i=2}^n \log w(t_i \mid \mathcal{H}_{t_{i-1}} \cap \emptyset_{(t_{i-1},t_i)}) \\
&\quad + \log (1 - W(T \mid \mathcal{H}_{t_n} \cap \emptyset_{(t_n,T)})) \\
&= \sum_{i=1}^n \log \lambda_g(t_i \mid \mathcal{H}_{t_i}) - \int_0^{t_1} \lambda_g(u \mid \mathcal{H}_0 \cap \emptyset_{(0,u)}) du \\
&\quad - \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \lambda_g(u \mid \mathcal{H}_{t_{i-1}} \cap \emptyset_{(t_{i-1},u)}) du \\
&\quad - \int_{t_n}^T \lambda_g(u \mid \mathcal{H}_{t_n} \cap \emptyset_{(t_n,u)}) du \\
&= \sum_{i:0 \leq t_i \leq T} \log \lambda_g(t_i \mid \mathcal{H}_{t_i}) - \int_0^T \lambda_g(t \mid \mathcal{H}_t) dt.
\end{aligned}$$

A fenti gondolatmenet kiterjeszthető jelölt pontfolyamatokra, lásd Daley és Vere-Jones (2006) 7.3. fejezetét. Így tehát az alábbi tételhez jutunk:

2. Tétel. (*MPP likelihood-függvénye*)

Legyen N egy reguláris MPP a $[0, T] \times \mathcal{M}$ állapottéren véges pozitív T érték mellett, $\{(t_1, m_1), \dots, (t_{N_g(T)}, m_{N_g(T)})\}$ pedig a folyamat realizációi. Ekkor, a megfigyelt folyamat likelihood-függvénye az alábbi alakot ölti:

$$\begin{aligned}
L &= \left[\prod_{i=1}^{N_g(T)} \lambda(t_i, m_i) \right] \exp \left(- \int_0^T \int_{\mathcal{M}} \lambda(u, m) du \ell_{\mathcal{M}}(dm) \right) \\
&= \left[\prod_{i=1}^{N_g(T)} \lambda_g(t_i) \right] \left[\prod_{i=1}^{N_g(T)} f(m_i \mid t_i) \right] \exp \left(- \int_0^T \lambda_g(u) du \right), \tag{15}
\end{aligned}$$

ahol $\ell_{\mathcal{M}}$ az $\mathcal{M}$ térben a Lebesgue mértékre vett referenciamérték.

Az egyenlőség jobb oldalát azon feltevés mellett kapjuk, hogy a jelek sűrűségfüggvénye $\mathcal{M}$ felett kiintegrálva 1-et ad. A jelek feltételes sűrűségfüggvényét tartalmazó

tagot L_1 , az alapintenzitás-függvényt tartalmazót pedig L_2 jelölje. A fontosabb fogalmak és számítások tisztázása után áttérhetünk az SEMPP modell tárgyalására, a diszkrét maximum likelihood becslés eljárására, valamint az illeszkedésvizsgálathoz és szimulációhoz szükséges elméletre.

4. SEMPP modell

4.1. A modell felírása

Ahogy azt már korábban említettük, a természeti katasztrófák okozta károkra általában nem reális feltételezés, hogy független, azonos eloszlásúak. Ezt láttuk a 2.2-es fejezetben is, az általunk vizsgált idősorban egyértelműen van függőségi struktúra, a megfigyeléseknél még a stacionaritás is megkérdőjelezhető. Az extrém index 0.47-es értéke azt sejteti, hogy az extrém értékek között is jelentős a kapcsolat, ezt az 1. ábrán is láthatjuk; egy küszöb felett a megfigyelések gyakran klasztereződnek.

Az SEMPP keretrendszere lehetővé teszi ilyen idősorok szélsőséges viselkedésének modellezését. Természetesen ennek a módszernek is megvan a maga hátránya; a gyakorlatban rendszerint nehéz találni egy kellően rugalmas és az adott idősorra jól illeszthető alakot a folyamat sztochasztikájának leírására (Chavez-Demoulin és tsai. 2005). Tanulmányunkban az először Alan G. Hawkes (Hawkes 1971; Hawkes és Oakes 1974) által bevezetett, és róla elnevezett Hawkes folyamat struktúráját vesszük alapul. A következő egyenletek Li és tsai. (2021) jelölésrendszerét követik. Egy (folytonos) Hawkes folyamat olyan pontfolyamat, melynek sztochasztikus intenzitásfüggvénye az alábbi:

$$\lambda(t) = \zeta + \nu(t), \quad \forall t \in [0, T], \quad (16)$$

ahol ζ a folyamat háttérintenzitása, amely az egyes klaszterek átlagos érkezési idejét tükrözi, valamint

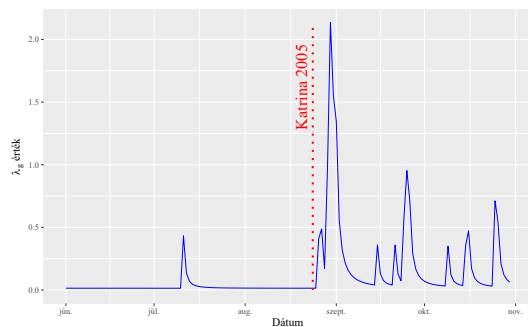
$$\nu(t) = \sum_{i:t_i < t} \phi(t - t_i) = \int_0^t \phi(t - s) dN(s) \quad (17)$$

az öngerjesztő komponens. A (16)-os egyenletben leírt Hawkes folyamatot lineárisnak nevezzük, ϕ a klasztereződés tendenciáját szabályozza, vagyis azt, hogy mennyi idő alatt áll vissza λ a háttérarára. A fenti folyamat tisztán öngerjesztő pontfolyamat (SEPP), ennek a fogalmát terjeszti ki az SEMPP, mely számításba veszi a kölcsönhatást az alapfolyamat és a jelek sűrűségfüggvénye között. Egy SEMPP alapfolyamata is $\lambda_g = \zeta + \nu(t)$ alakú, azonban a (17)-es egyenletben definiált öngerjesztő tag kiegészül az alábbi módon:

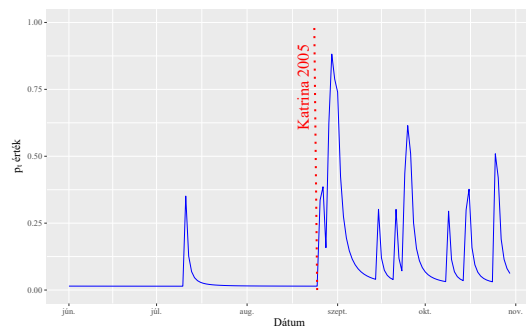
$$\nu(t) = \sum_{i:t_i < t} \phi(t - t_i) g(m_i) = \iint_{(0,t) \times \mathcal{M}} \phi(t - s) g(m) dN(s, m). \quad (18)$$

A g függvény tetszőleges t_i időpontban bekövetkezett m_i esemény intenzitásfüggvényre gyakorolt hatását fejezi ki. Ezen a ponton érdemes szót ejteni a folytonos és diszkrét modellek közötti eltérésről. Míg a folytonos modellben $\lambda_g(s)$ értelmezhető annak a feltételes valószínűségként, hogy ismerve a folyamat múltját az s időpontig bezárólag, az s időpontot követő kicsiny $[s, s + \delta]$ intervallumon bekövetkezik egy esemény, addig ez a diszkrét modell esetében nincs így. Utóbbi esetben az események csak $t_1, t_2, \dots$ egész számot felvevő időpontokban történhetnek, így az alapintenzitásfüggvény diszkrét pontfolyamat lesz, az események bekövetkezését pedig - ahogy azt korábban említettük - bináris $Y_1, \dots, Y_T$ sorozattal reprezentálhatjuk. A fent említett feltételes valószínűségek diszkrét megfelelője a $p_1, p_2, \dots, p_T$ sorozat, melyek $\forall t \in \mathbb{N}$ időpontban $p_t = \mathbb{E}(Y_t \mid \mathcal{H}_{t-1}) = P(Y_t = 1 \mid \mathcal{H}_{t-1})$ képlettel definiálhatók. Ezeket a p_t értékeket természetesen a diszkrét folyamat alapintenzitás-függvényből származtatjuk, így szükségünk van egy $\varphi : \mathbb{R} \rightarrow [0, 1]$ függvényre, amellyel annak minden t időpontbeli értékét a megfelelő intervallumra transzformálhatjuk. Ha összeadjuk a $t_1, t_2, \dots, t_T$ időpontokban a p_t értékeket, akkor megkapjuk az adott megfigyelési időre az események számának várható értékét, így a valós értékhez közeli $\sum_{t=1}^T p_t$ a modell megfelelő illeszkedését jelezheti. A φ függvénynek számtalan alakját használják az irodalomban (Daley és Vere-Jones 2006), mi Li és tsai. (2021) tanulmányában alkalmazott $\varphi = 1 - e^{-\lambda_g}$ transzformációs függvényt választjuk. Diszkrét és folytonos esetben is tetszőleges t időpont előtti események kihatnak az alapintenzitás-függvény jelenlegi értékére, megnövelve ezzel t -ben és az azt követő időpontokban egy esemény

bekövetkezésének valószínűségét. A ϕ úgynevezett magfüggvény és a g függvény tehát együttesen szabályozzák az események hatását a folyamat alakulására. Az alapintenzitás-függvény változását, valamint az ide tartozó feltételes valószínűségeket az általunk vizsgált idősoron az alábbi ábra szemlélteti:



(a) Intenzitásfüggvény alakulása 2005 júniusa és októbere között



(b) Események bekövetkezésének feltételes valószínűségei 2005 júniusa és októbere között

3. ábra. Az intenzitásfüggvény és hozzátartozó feltételes valószínűségek alakulása

Ahogy az várható, a Katrina hurrikán érkezésével jelentősen megugrik az alapintenzitás-függvény értéke, mely csak hosszabb idő után áll vissza a normál szintre.

4.2. A jelek sűrűségfüggvénye - EGPD

Papastathopoulos és Tawn (2013) bevezetnek egy új eloszláscsaládot, amely az Általános Pareto eloszláscsalád bővítése (Extended General Pareto Distribution, EGPD), mellyel céljuk a POT módszertan stabilitásának javítása. Ezt egy további κ skálaparaméter hozzáadásának segítségével érik el, mellyel az eloszlás pontosabban leírható, széleinek karakterisztikáit változtatlanul hagyva. Minden egyes EGPD eloszlás egy $X = \sigma H_{\xi}^{-1}(V)$ integráltranszformáción keresztül definiálható, ahol H_{ξ}^{-1} az általánosított Pareto eloszlás kvantilisfüggvénye és V egy $(0, 1)$ intervallumon értelmezett valószínűségi változó. Ekkor $F_{\eta}(x) = G\left\{H_{\xi}\left(\frac{x}{\sigma}\right)\right\}$, $\eta = (\xi, \sigma)$, a megfelelő sűrűségfüggvény pedig $x \in (0, \infty)$ mellett $f_{\eta}(x) = \sigma^{-1} h_{\xi}\left(\frac{x}{\sigma}\right) g\left\{H_{\xi}\left(\frac{x}{\sigma}\right)\right\}$. Annak érdekében,

hogy egy EGPD kellően megragadja az eloszlás alsó végét, amellet, hogy a jobboldali farokviselkedés karakterizációja változatlan maradjon, szükségünk van megfelelő G_κ kapocsfüggvény találására. Naveau és tsai. (2016) küszöbérték választása nélkül párhuzamosan modellezik az alacsony, közepes és magas intenzitású csapadékokat. Mivel a vizsgált eloszlás az egész pozitív félegyenesre kiterjed, így az alsó és felső farokviselkedés szimultán karakterizációja kiemelt figyelmet kap. Ennek érdekében elemzésükben 3 fajta kapocsfüggvényt tárgyalnak. Li és tsai. (2021) az extrém alacsony vízszintek küszöb körüli koncentrációjából, és az eloszlás jobboldali szélének vastagságából kifolyólag hasonló problémával szembesülnek. Mindkét tanulmányban sikeresen alkalmazzák a béta kapocsfüggvényt a kutatási kérdés vizsgálatához. Ahogy az a 2. fejezetben bemutatásra került, az NFIP adatbázisa jelentősen balra dőlő, jobbra elnyúló, valamint a küszöbérték környezetében is nagy az extrém károk koncentrációja, így jogos a feltételezés, hogy itt is az előző kettő tanulmányhoz hasonló problémával találjuk szemben magunkat.

Továbbra is célunk, hogy a folyamat múltbeli alakulása egyaránt hatással legyen a jövőbeli események bekövetkezésének időpontjaira, valamint a hozzájuk tartozó kárnagyságok mértékére. Ennek érdekében a jelek eloszlásának skálaparaméterét időtől függővé tesszük $\nu(t)$, azaz az öngerjesztő kapocsfüggvény segítségével, melyet a (17)-es egyenletben definiáltunk. A skálaparaméter tehát a $\sigma_t = \beta_0 + \beta_1\nu(t)$ alakot ölti. Ennek köszönhetően az extrém károk klasztereződési tendenciája leírhatóvá válik a SEMPP modell illesztésével. A jelek feltételes sűrűségfüggvénye tehát az alábbi:

$$f_M(m \mid t, \mathcal{H}_t) = \sigma_t^{-1} g_\kappa \left\{ H_\xi \left(\frac{m}{\sigma_t} \right) \right\} h_\xi \left(\frac{m}{\sigma_t} \right), \quad (19)$$

ahol $g_\kappa = G'_\kappa$.

4.3. Modellillesztés, illeszkedésvizsgálat, klaszterreprezentáció, szimuláció

4.3.1. A loglikelihood-függvény felírása

Az MPP modellezésnél leírtaknak megfelelően egy SEMPP modellt is teljesen karakterizál az alapintenzitás-függvényének (16) és a jelek feltételes sűrűségfüggvényének (19) szorzata. $t \in \{1, \dots, T\}$ esetén a diszkrét idejű SEMPP intenzitásfüggvénye az alábbi alakot ölti:

$$\lambda_{Y_t, M_t | \mathcal{H}_{t-1}}(t, m) = \{p_t f_M(m_t | Y_t = 1, \mathcal{H}_{t-1})\}^{y_t} (1 - p_t)^{1-y_t}. \quad (20)$$

Adott u küszöbérték mellett jelölje az események számát n_u . Ekkor a diszkrét modell paramétereinek becslését megkaphatjuk az $\ln(L_1) + \ln(L_2)$ loglikelihood-függvény maximalizálásával, ahol

$$\ln(L_1) = \sum_{i=1}^{n_u} \ln(f(m_i | t_i))$$

és

$$\ln(L_2) = \sum_{t=1}^T \{y_t \ln(p_t) + (1 - y_t) \ln(1 - p_t)\}$$

ahol természetesen $p_t = \varphi\{\zeta + \nu(t)\}$, y_t értéke pedig 1 ha történt esemény, 0 különben. Az egyenletekből kiolvashatjuk, hogy $\ln(L_1)$ és $\ln(L_2)$ értékei nem függetlenek, mivel a közös $\nu(t)$ öngerjesztő tag összeköti őket, így együttesen maximalizálandók.

A modell illeszkedése két szempontból ellenőrizendő; egyrészt szükséges megbizonyosodni arról, hogy sikerült az extrém káresemények időbeli eloszlását megfelelően megragadni, valamint a jelek becsült eloszlásának is kielégítőnek kell lennie. Az idődimenzióbeli illeszkedésvizsgálatot szokás reziduális elemzésnek is nevezni, melynek módszere véletlen időtranszformáción alapszik. Az elméleti háttérét Daley és Vere-Jones (2006) könyvének 7.4.IV állítása adja:

2. Állítás. *Legyen N egy pontfolyamat folytonos λ intenzitásfüggvénnyel és legyen*

$\Lambda(t) = \int_0^t \lambda(s) ds$. Ha a t_i realizációk N pontfolyamatból származnak, akkor 1 valószínűséggel a $\tau_i = \Lambda(t_i)$ transzformált sorozat egy egységrátájú PPM realizációja.

Diszkrét pontfolyamat esetében az integrált egyszerűen felváltja egy véges szuma (Li és tsai. 2021), tehát $\Lambda(t_i) = p_1 + \dots + p_{t_i}$. Ezáltal az illeszkedésvizsgálat kimerül annyiban, hogy adott becsült intenzitásfüggvény mellett megfigyeljük a transzformált időpontok eltávolodását egységrátájú PPM-től. Az utóbbi folyamat jól ismert tulajdonsága azonban, hogy az események számára feltételelesen az érkezési idők egyenletes eloszlásúak és függetlenek, azaz a transzformált időpontok $\tau_1 = \Lambda(t_1), \dots, \tau_n = \Lambda(t_n)$ sora egy egységrátájú Poisson folyamat érkezési idejei (Michaletsky 2001). Ezzel ekvivalensen a skálázott $\omega_1 = \tau_1/\Lambda(T) < \dots < \omega_n = \tau_n/\Lambda(T)$ sorozat rendezett mintát alkot egy $U(0, 1)$ eloszlásból. Ezt a hipotézist könnyen ellenőrizhetjük vizuálisan egy valószínűségi ábra segítségével az $(1/n, \hat{\omega}_1), \dots, (1, \hat{\omega}_n)$ párok ábrázolásával, statisztikailag pedig a Kolmogorov-Smirnov teszt elvégzésével, ahol a nullhipotézis elfogadása mellett megfelelőnek tekinthető az illeszkedés (Li és tsai. 2021).

Ami a jelek eloszlásának illeszkedésvizsgálatát illeti, ezt ellenőrizhetjük a szokásos módon: valószínűségi és kvantilis ábrák segítségével. Azonban a modell jelei olyan EGP eloszlásfüggvényűek, melynek forma-paramétere konstans, viszont skálaparamétere időben változó és $\nu(t)$ -től függ. Ez heterogenitást idéz elő a jelek eloszlásában, amelytől meg kell szabadulni az illeszkedésvizsgálat elvégzése előtt. Ezt úgy tesszük, hogy a jeleket egy valószínűségi integráltranszformáción keresztül sztenderdizáljuk, azaz elvégezzük az

$$U_t = G_\kappa \left\{ H_\xi \left(\frac{X_t}{\sigma_t} \right) \right\} = G_\kappa \left\{ 1 - \left(1 + \xi \frac{X_t}{\sigma_t} \right)^{-1/\xi} \right\}$$

átalakítást. A paraméterek becslése után a sztenderdizált jelek megközelítőleg függetlenek és $U(0, 1)$ eloszlásúak. Az ábrák elkészítéséhez választanunk kell egy referencia eloszlásfüggvényt (jelöljük ezt F_Y -nal), majd az $\tilde{X}_t = F_Y^{-1}(U_t)$ transzformáció elvégzése után az $\tilde{X}_t$ jelek körülbelül F_Y eloszlásúak. Így már készíthetünk egy va-

lósínűségi ábrát minden $i \in \{1, \dots, n\}$ index mellett a

$$\left(\frac{i}{n+1}, F_Y(\tilde{x}_i) \right) = \left(\frac{i}{n+1}, F_Y \{F_Y^{-1}(u_{(i)})\} \right) = \left(\frac{i}{n+1}, u_{(i)} \right)$$

párokkal. Ez az ábra nem függ az F_Y referencia-eloszlás megválasztásától (Li és tsai. 2021). Ezzel szemben a kvantilis ábra $(F_Y^{-1} \{ \frac{i}{n+1} \}, F_Y^{-1}(u_{(i)}))$ minden $i \in \{1, \dots, n\}$ -re, így ez már függ F_Y -tól. Gyakori választás az F_Y függvényre az egységrátájú exponenciális eloszlás, hiszen ez előáll a GPD határeloszlásaként $\xi \rightarrow 0$ mellett (Embrechts és tsai. 1997). Ezzel a választással $\tilde{X}_t = -\ln(1 - U_t)$. A kvantilis ábra így $i \in \{1, \dots, n\}$ mellett a

$$\left(-\ln \left(1 - \frac{i}{n+1} \right), -\ln(1 - u_{(i)}) \right) = \left(-\ln \left(1 - \frac{i}{n+1} \right), \tilde{x}_{(i)} \right)$$

párokat tartalmazza. Mindkettő alkalmas a modell validációjának elvégzésére, a kvantilis ábra főleg abban az esetben hasznos, ha az eloszlás felső értékeinek illeszkedését vizsgáljuk.

4.3.2. Szimuláció a klaszterreprezentáció segítségével

Folytonos esetben a legelterjedtebb módszer jelölt pontfolyamatok szimulálására a ritkítási algoritmus, melyet Ogata (1981) vezetett be. A módszer lényege, hogy a lehetséges klaszterközepeket egy homogén PPM generálja, majd ezen pontokat valamilyen előre meghatározott sztochasztikus szabály alapján megritkítjuk. A ritkításra azért van szükség, mert a szimuláció során használt homogén PPM paraméterének az illesztett modell intenzitásfüggvényének valamely felső határát szokták választani. Minthogy ezzel az eljárással nem csak $c_i \in \mathbb{N}$ klaszterközepeket kapunk, így ez a módszer nem alkalmazható diszkrét esetben. Szerencsére a folyamat klaszterreprezentációján alapuló szimuláció egyaránt alkalmazható mind folytonos, mind diszkrét modelleknél.

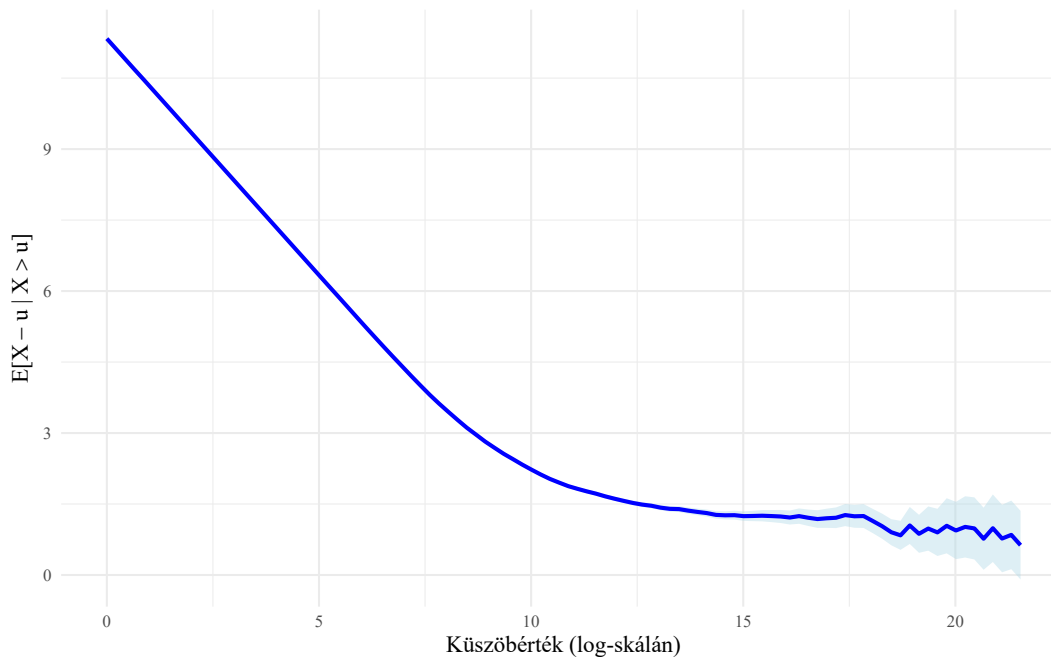
Elméleti és gyakorlati szempontból a klaszter folyamat a pontfolyamatok között az egyik legszélesebb körben alkalmazott és kutatott folyamat. Alkalmazási területé-

nek palettája igen sokszínű, mivel a két-, és háromdimenziós térben elhelyezkedő jelenségek modellezésének természetesen megjelenő objektumai. Segítségükkel vizsgálják többek között vulkánkitörések, sorbanállási rendszerek, populációk, csillagok viselkedését. Egy klaszter folyamat felépítése két fő alkotóelemből áll össze, ezek a klaszterek, valamint és az egyes klaszterek elemeinek elhelyezkedése, melyeket N_c , illetve $N(\cdot \mid c_i)$ folyamatok generálják (Daley és Vere-Jones 2006). Hawkes és Oakes (1974) bebizonyítják, hogy tetszőleges véges intenzitásfüggvényű, stacionárius öngerjesztő folyamat reprezentálható klaszter folyamatként. Ez intuitív módon a következőképpen zajlik. Minden egyes esemény besorolható bevándorló illetve utód kategóriába. Minden bevándorló a folyamat ágazási struktúrája szerint generál utódokat. Egy bevándorló az utódjaival együtt alkot egy klasztert. Diszkrét esetben a klaszterreprezentáció egy igen egyszerű folyamatot eredményez; a bevándorlókat - azaz a klaszterek kezdőpontjait - egy homogén Bernoulli folyamat generálja, melynek paramétere $\varphi(\zeta)$, így a klaszterszám a $[0, T]$ intervallumon binomiális eloszlású $\varphi(\zeta)T$ várható értékkel. Minden bevándorló generál egy utódot egy inhomogén Bernoulli folyamat szerint, melynek a paraméterét a folyamat öngerjesztő tagja, vagyis ν_t modulálja és a "siker" valószínűsége $\varphi\{\lambda_g(t)\}$. Egy klaszter megszűnik, amint egy új esemény bekövetkezésének valószínűsége visszaáll az alaphelyzetbe. A diszkrét klaszterreprezentáció felírásából adódóan a különböző statisztikák - például a szintmeghaladás hossza, a visszatérési idő - aránylag egyszerűen meghatározhatóak.

5. Modellezés eredményei

5.1. SEMPP modell illesztése

A második fejezetben leírtaknak megfelelően az elemzés során aggregált napi adatokkal dolgozunk. Statisztikai szempontból egy adott napra eső árvízkiárt akkor tekintünk extrémnek, ha értéke meghaladja az u küszöbértéket. Ezen küszöbérték helyes megválasztásában nyújt segítséget a (7)-es egyenletben definiált átlagos szintmeghaladási függvény ábrázolása az empirikus adatokat felhasználva. Az alábbi ábra a könnyebb interpretáció érdekében a logaritmizált adatsor értékei alapján készült.



4. ábra. Empirikus átlagos szintmeghaladási függvénye a napi szinten aggregált NFIP árvízkároknak

Az ábrán látható világoskék sáv a 95 százalékos konfidenciaintervallumot jelöli. Látható, hogy az átlagos szintmeghaladás függvény nagyjából 15 körül válik lineárisra, így az u küszöbértéknek $e^{15} = 3\,269\,017$ USD, vagyis 0.0032 milliárd dollárt választjuk. Ezzel az adatsorunkba 561 megfigyelés (t_i, m_i) kerül, ahol ha az adott nap kárnagyságát x_i -vel jelöljük akkor $m_i = x_i - u$. Az extrém értékek idősorának létrehozásakor az adatokon SEMPP modell esetén nem végzünk deklaszterezést, így az klaszter-tendenciára vonatkozó információtartalom megmarad. Idősorunk esetében az alkalmazott u küszöbérték mellett 0.47-es értéket kapunk az extrém index értékre, azaz szignifikánsan klasztereződnek az extrém értékek az u küszöbérték felett, a becsült átlagos klaszternagyság tehát $1/0.47 = 2.12$.

5.1.1. Az alapintenzitás-függvény meghatározása

Ahhoz, hogy az SEMP alapfolyamatát meghatározzuk, választanunk kell egy g hatásfüggvényt és egy ϕ magfüggvényt a (18)-as egyenlet alapján. Ezek megválasztása általában nem egyértelmű. Gyakran használt változatai g -nek különböző exponenciális ($e^{\delta m}$), lineáris $(1 + \delta m)$ függvények, ϕ -nek pedig exponenciális lecsengésű, logit, esetleg autoregresszív tagokat magában foglaló függvényeket is szoktak választani. Li és tsai. (2021) tanulmányában a (16)-os egyenletben található az alkalmazott alapintenzitás-függvény alakja, $g = 1 + \tanh(\delta m)$ és $\phi = \exp\{-\gamma(t_i - t)\}$ függvényekkel. Cikkükben az aszály jelenségével foglalkoztak, melynek nagysága természetéből adódóan szigorú felső határa van, valamint folytonos modellt is illesztettek, így abban a stacionaritás feltételének teljesüléséhez elengedhetetlen volt korlátozott g függvény választása. Minket az utóbbi probléma nem érint, mivel diszkrét modellt illesztünk, valamint a kárnagyságok jellegéből adódóan inkább $g(m) = \{1 + \log(1 + \delta m)\}$ hatásfüggvényt mellett döntünk. Ami a magfüggvényt illeti, hosszas kísérletezés után, hatványosan lecsengő $\phi(t) = \frac{1}{1 + \gamma(t_i - t)^2}$ függvényt választjuk, mely jobban összhangban van a klaszternagyságok alacsony, 2.12-es várható értékével. Li és tsai. (2021) aszályok viselkedésének elemzésénél az erőteljes szezonális jelenléte miatt a téli hónapokra ζ -ra korlátozzák az alapintenzitás-függvényt, azonban a mi idősorunkban ilyen jellegű probléma nem áll fenn, hiszen nagyjából azonos arányban történtek extrém káresetek az egyes évszakokban. A folyamat végső alapintenzitás-függvénye tehát az alábbi alakot ölti:

$$\lambda_g(t) = \zeta + \psi\nu(t) = \zeta + \psi \sum_{i:t_i < t} \{1 + \log(1 + \delta m_i)\} \frac{1}{(1 + \gamma(t_i - t))^2}, \quad (21)$$

ahol $\psi \in [0, \infty)$, $\delta \in [0, \infty)$, $\gamma \in [0, \infty)$ paraméterek, ψ -re azért van szükség, hogy $\delta = 0$ rögzítése mellett SEPP modellt kapjunk. A transzformáció-függvény a korábban említett, így $p_t = 1 - e^{\lambda_g(t)}$.

5.1.2. Becsült paraméterek

A 4.3.1-es fejezetben, a (20)-as egyenletből levezetett loglikelihood-függvény maximalizálása numerikus módszerekkel végezhető el, melyhez eszközül a Newton-módszerrel kalkuláló (Dennis és Schnabel 1982) `nlm` beépített optimalizáló függvényt használjuk. A függvény érzékeny a kiinduló paraméterekre, így a konvergenciahibák kiküszöbölése érdekében a különböző paraméterekkel és sűrűségfüggvényekkel rendelkező modellek mindegyikénél először a BFGS (Nocedal és Wright 2006) algoritmust alkalmazzuk, majd az így kapott értékekből indítjuk az optimalizációt. A jelek sűrűségfüggvényéhez tartozó kapocsfüggvénynek Li és tsai. (2021) munkájához hasonlóan béta-függvényt választunk, tehát $G_\kappa(v) = v^\kappa$. A loglikelihood-függvény maximalizálásával becsült paramétereket a következő táblázat tartalmazza, a Hesse-mátrixból számolt sztenderd hibák és t -értékek mellett:

Paraméter	ζ	ψ	γ	δ	ξ	β_0	β_1	κ
Becsülés	0.014	228.24	30.561	164.87	1.083	0.003	1.769	1.039
Std. hiba	0.001	202.72	14.027	130.03	0.107	0.001	1.656	0.127
t -érték	14.187	1.126	2.179	1.268	10.085	3.816	1.068	8.201

1. táblázat. A modell illesztett paramétereinek becslései, sztenderd hibái és t -értékei

A loglikelihood-függvény maximális értéke -28.4. $\zeta = 0.014$ azt jelenti, hogy az évente jelentkező bevándorlók, azaz klaszterek száma $0.014 \times 365 = 5.1$. A további érdekes információtartalommal rendelkező paraméterek ξ és κ . ξ becsült értéke 1-nél nagyobb, ami azt jelenti, hogy az illesztett eloszlásnak nem létezik egyik momentuma sem, vagyis már várható értéke is végtelen. Ez nem meglepő eredmény, ha figyelembe vesszük a modellezett károk jellegét. Ahogy Papastathopoulos és Tawn (2013) bemutatja, $\kappa = 1$ mellett az EGPD eloszlás ekvivalens a GPD eloszlással. A megfelelő sztenderd hiba értékéből láthatjuk, hogy az illesztett κ érték nem tér el szignifikánsan 1-től, azaz a GPD eloszlás megfelelő az extrém árvízkárok leírására. A β_0 paraméterből leolvasható, hogy a 4.2-es fejezetben definiált időtől függő σ_t skálaparaméter minimális értéke 0.03. Azonban feltűnhet, hogy egyes paraméterek, például ψ, γ valamint δ értékei meglehetősen nagyok, amelelt, hogy a hozzájuk

tartozó t -értékek alacsonyak. Ez két okból is gondot okozhat. Elsősorban a (21)-es egyenletből arra következtethetünk, hogy a nagy paraméterek túlilleszkedést jelezhetnek, hiszen ezek a tagok mind $\nu(t)$ -t kontrollálják, a modell nehezen határozza meg a kárnagyságok pontos hatását. Másodszor, a magas t -értékek azt jelentik, hogy a paraméterek becslése instabil. Lineáris regressziónál hasonló problémákat gyakran a loglikelihood-függvénybe tett regularizációs tagok alkalmazásával orvosolják (Wooldridge 2016). Ezt a megközelítést alkalmazzuk mi is, regularizációs tagnak a paraméterek négyzetének összegét állapítjuk meg 0.01-gyel megszorozva, ezzel az alábbi táblázatot kapjuk:

Paraméter	ζ	ψ	γ	δ	ξ	β_0	β_1	κ
Becslés	0.014	15.080	5.934	19.508	1.091	0.003	0.127	1.034
Std. hiba	0.001	4.281	0.952	4.526	0.108	0.001	0.053	0.126
t -érték	13.952	3.522	6.234	4.310	10.084	3.797	2.411	8.188

2. táblázat. A modell illesztett paramétereinek becslései, sztenderd hibái és t -értékei regularizációval

Láthatjuk, hogy a háttérintenzitás változatlan, valamint a kárnagyságok eloszlásának ξ paramétere mellett κ értéke is megközelítőleg egyezik. A t -értékek mind 2 feletti, így a paraméterek becslése pontosnak mondható. Az ide tartozó loglikelihood-függvény maximális értéke -31.4. Az elkövetkezendőkben ezt a második megközelítést preferáljuk. Az SEMPP modell mellett $\delta = 0$ fixálásával illesztünk egy SEPP modellt is, melyben a jelek nagysága nincs hatással a folyamat alapintenzitás-függvényére. Ennek az illesztett paramétereit az alábbi táblázatban láthatóak:

Paraméter	ζ	ψ	γ	δ	ξ	β_0	β_1	κ
Becslés	0.014	15.093	5.295	0	1.197	0.003	0.107	1.064

3. táblázat. Az SEPP modell becsült paramétereit

Ebben az esetben is hasonló mondható el ζ , ξ és κ értékeiről mint az SEMPP

modellnél, megjegyzendő azonban, hogy az optimumban a loglikelihood-függvény -69.2-re csökkent.

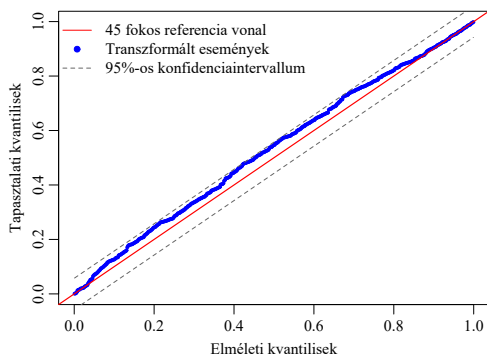
5.2. Illeszkedésvizsgálat, összehasonlítás a klasszikus modellel

A harmadik fejezetben leírtaknak megfelelően a deklaszterezett empirikus kárnagyságokra illesztünk egy hagyományos pontfolyamat modellt is, ugyanazon u küszöbérték feletti megfigyelésekre. Ezen módosított idősor megfigyelései már függetlenek, azonban így csak 309 extrém értékre illeszthetjük az eloszlást, tehát jelentős információvesztéssel kell szembenéznünk. Ezen modell skálaparamétere és háttérintenzitása időben konstans, a becsült paraméterei a 3. táblázatban láthatóak. A háttérintenzitás 0.018-as értéke valamivel nagyobb, mint a SEMPP modell esetében, ami annak ellenére, hogy az idősorban kevesebb megfigyelés található nem meglepő, hiszen a modell feltétele, valamint a deklaszterezés eredménye az u feletti megfigyelések függetlensége, mely az ATD függvény ábrázolását vizsgálva teljesül. A háttérintenzitás becsült értéke ebben az esetben egyszerűen az extrém károk száma elosztva a vizsgált időszak hosszával.

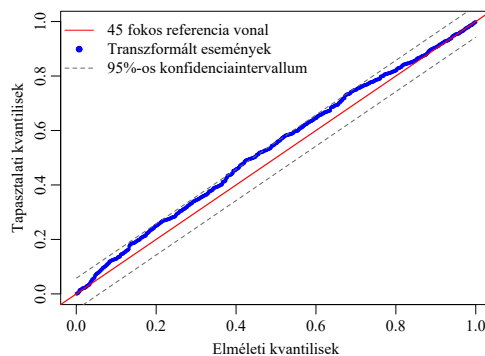
Paraméter	Háttérintenzitás	σ	ξ
Becslés	0.018	0.053	1.289
Std. hiba	-	0.004	0.043
t -érték	-	13.250	29.977

4. táblázat. A Poisson pontfolyamat modell illesztett paramétereinek becslései, szten-
derd hibái és t -értékei

Az itt taglalt modellben a jelek nagysága független a folyamat alakulásától, ξ értéke magasabb mint az SEMPP modell esetében, tehát az illesztett GPD eloszlásnak szintén nem létezik a várható értéke. Az időben konstans becsült skálaparaméter értéke 0.053, ennek majd a szimulációval foglalkozó fejezetben lesz jelentősége. Az alábbi két ábrán a SEMPP és SEPP modellek 4.3.1-es fejezetben taglalt reziduális illeszkedésvizsgálata látható.



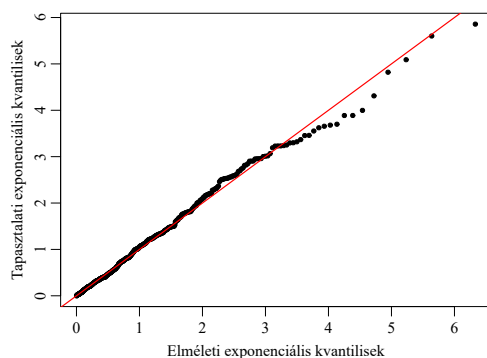
(a) SEMPP modell



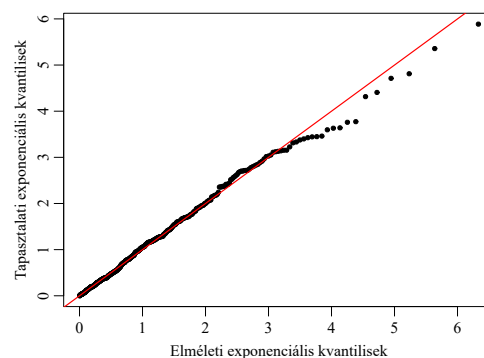
(b) SEPP modell

5. ábra. Reziduális tesztek valószínűségi ábrái: bal oldalon az SEMPP, jobb oldalon az SEPP modell transzformációval

Tökéletes illeszkedés esetén a transzformált időpontok kék vonal rásimulna a piros referencia vonalra. A két ábra igen hasonló, azonban a SEPP modell esetében a transzformált időpontok vonala egy kicsivel kilép a konfidenciaintervallum határolta sávból, így az illeszkedése 5 százalékos szignifikanciaszinten nem megfelelő. Ezt támasztja alá az ugyancsak a 4.3.1-es fejezetben említett Kolmogorov-Smirnov teszt p-értékei, melyek rendre 0.101 és 0.014, vagyis az SEPP modell nem ragadja meg kellően az extrém károk időbeli eloszlását. Ez megerősíti azt a hipotézist, hogy az extrém kárnagyságok folyamatának alakulásában az egyes jelek értékei nem hagyhatók figyelmen kívül. Ez a jelenség hasonló a földrengéseket kutatók tapasztalataihoz, miszerint egy-egy nagyobb esemény után megnövekedik a további események bekövetkezésének valószínűsége (Ogata 1988). Egy lehetséges magyarázata ennek, hogy egy ekkora területre kiterjedő portfólió esetén egy folyó valamely vízgyűjtő területén bekövetkező természeti katasztrófa jelentősen megnövelheti a vízszintet, mely újabb károkhoz vezethet a későbbiekben az adott folyó mentén. A jelek dimenziójának illeszkedésvizsgálata a következő két ábrán látható:



(a) SEMPP



(b) SEPP

6. ábra. Exponenciális kvantilis-ábrák SEMPP és SEPP modellek esetében

A két ábra közötti eltérés alig észrevehető, talán valamivel jobban illeszkedik az SEMPP modell az extrém kárnagyságokra, azonban a megfelelő Kolmogorov-Smirnov teszt mindkét modellt adekvátnak tekinti ebből a szempontból. A loglikelihood-függvények optimumbeli értékének jelentős különbsége tehát az idődimenzióbeli illeszkedés eltérésére vezethető vissza. A szimuláció során ezért az SEMPP modell paramétereit használjuk fel.

6. Szimulációs eredmények, viszontbiztosítási példa a modell alkalmazására

6.1. Szimuláció

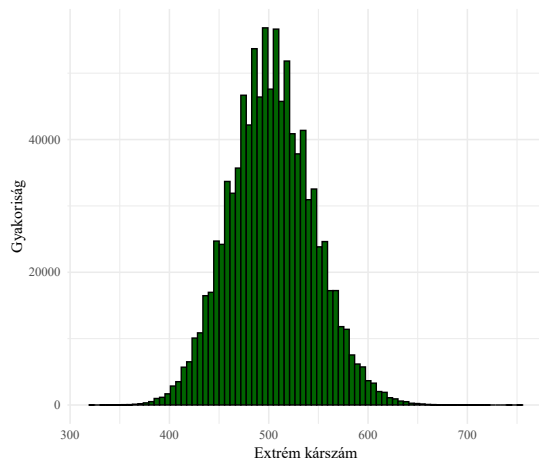
Ahogy azt az előző fejezetben bemutattuk, az illesztett modellek minden esetben végtelen várható értékű GPD eloszlást eredményeznek. Egy fix portfólió vizsgálatakor azonban nem reális az, ha egyetlen extrém káresemény esetén például 200 milliárd dolláros kárnagyságot kapunk. Ilyen események a szimuláció során igen ritkán fordulnak elő, azonban az eredményeket igen súlyosan torzíthatják. Ezen probléma

kiküszöbölése érdekében néhány feltételezéssel és egyszerűsítéssel kell élnünk. A szimulációk során a küszöbérték feletti maximális napi kárnagyságot 12 milliárd dollárra korlátozzuk, mely csaknem a Katrina hurrikán miatt felmerülő károk kétszerese. A szimulációs algoritmus a 4.3.2-es fejezetben leírtak logikáját követi. Számítási szempontból igen kedvezőtlen lenne, ha minden egyes esemény hatását végigfuttatnánk az egész vizsgált intervallumon, így az egyes hatásokkal maximum 8 napig kalkulálunk. Természetesen a generált események számának várható értéke az ablak variálásával párhuzamosan változik, azonban többszöri próba után levonható az a következtetés, hogy az ablakméret 5 feletti értékei között inszignifikáns az eltérés. Ezen megfigyelés összhangban van a becsült paraméterek értékeivel. Az alábbi táblázatból leolvasható, hogy az algoritmus milyen eredményeket produkál egy, az általunk használt adatbázis hosszával megegyező, vagyis 46.67 éves intervallumon az illesztett SEMPP modell paramétereivel:

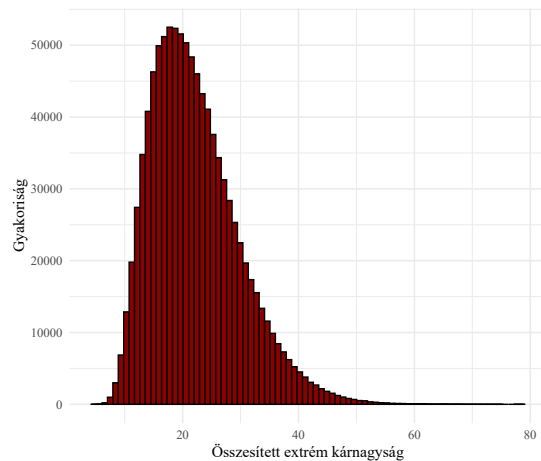
Statisztika	Események száma	Összkár	Átlagos kár
Minimum	320	4.493	0.0118
1. kvartilis	474	16.500	0.0331
Medián	502	21.026	0.0418
Átlag	503.1	22.172	0.0440
3. kvartilis	531	26.608	0.0526
Maximum	751	78.384	0.1543

5. táblázat. A szimulált események számának, az összesített extrém károknak és az átlagos u küszöb feletti napi károknak statisztikai jellemzői (1 000 000 szimuláció alapján).

Emlékezhetünk, hogy az FNIP idősorban 561 extrém esemény szerepelt, az összesített extrém kárnagyság 26.25 dollár és az átlagos küszöb feletti kár 0.046 milliárd dollár. Az alábbi ábra az események számának és a kumulált extrém károknak hisztogramjait tartalmazza.



(a) Extrém kárszám eloszlása



(b) Összesített extrém kár eloszlása

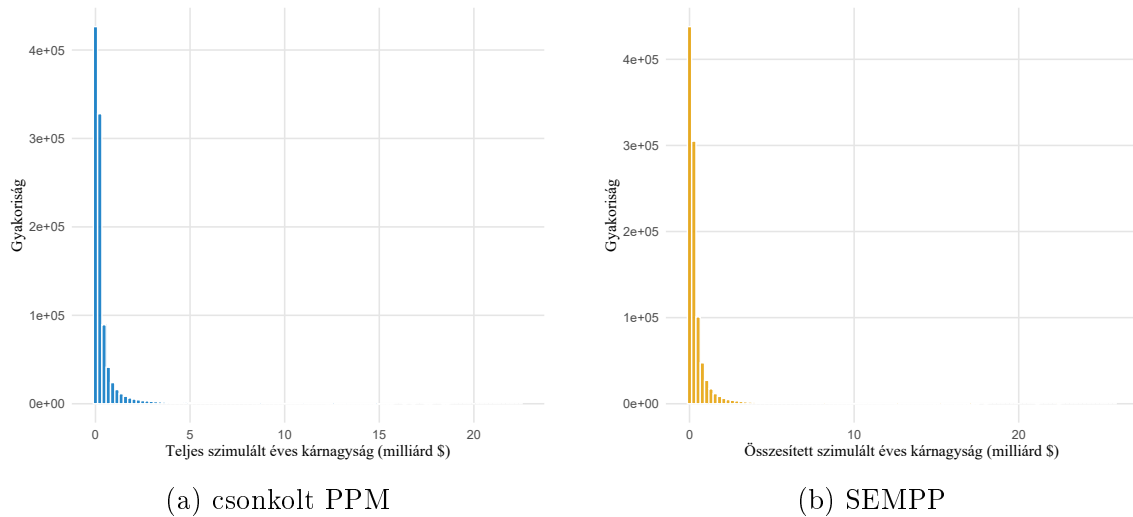
7. ábra. Hisztogramok 1 000 000 szimuláció alapján: extrém kárszám (balra) és összesített extrém kár (jobbra)

Következő lépésben az éves szimulált extrém kárnagyságok alakulására vagyunk kíváncsiak. Az összehasonlítás érdekében lefuttatjuk a szimulációt sima Poisson pontfolyamat modell esetében is, az illesztett háttérrel és GPD paraméterekkel. Ekkor azonban egy jelentősebb problémával találjuk szemben magunkat. A Poisson pontfolyamat modell becsült paraméterei túlságosan vastagfarkú eloszlást eredményeznek, a szimuláció során azt az eredményt kapjuk, hogy egy Katrina hurrikán volumenű természeti katasztrófa átlagosan 7 évente következik be. Ez a következő alfejezetben tárgyalt viszontbiztosítási szerződések árazását valamint a direkt-, és a viszontbiztosító eredményének elemzését ellehetetlenítené. Ennek orvoslására illesztünk egy csonkolt GPD eloszlású modellt 12-es felső határ mellett, melyről Albrecher és tsai. (2017) könyvében olvashatunk. A csonkolás lényege, hogy a módosított eloszlás tartója a $[0, M]$ intervallum, melyen $F_{csonk} = \frac{F(x)}{F(M)}$, így az eloszlást átsúlyozzuk az elemzéshez megfelelő tartományban. A likelihood-függvény megfelelő módosításával $\sigma = 0.0045$ és $\xi = 1.33$ adódik. Az alábbi táblázat az 5 legnagyobb extrém kárnagyságú év melletti visszatérési időket tartalmazza a 3 modell esetén.

Év	Éves kár (Mrd \$)	Visszatérési idő (év)		
		SEMP	PPM	PPM, csonkolt
2005	7.33	180.0	7.40	133.0
2012	3.68	52.9	4.20	42.9
2017	3.35	46.6	3.90	38.4
2022	2.30	28.0	2.92	24.9
2008	1.60	17.9	2.26	17.1

6. táblázat. A legnagyobb éves károk (Mrd USD) és becsült visszatérési idejük három különböző modellel

Jól látható, hogy a sima Poisson pontfolyamatra épült modell mennyivel felülbecsli az éves károk nagyságát. A másik két modell reálisabbnak tűnő visszatérési időket produkál, érdekes eredmény, hogy függetlenség feltételezése mellett, kevesebb extrém káreseményből is átlagosan nagyobb kumulált kárnagyságokat kapunk a csonkolt PPM modell esetében. Az alábbi ábrán az SEMPP és csonkolt PPM modellel szimulált éves extrém kárnagyságok szerepelnek.



8. ábra. A PPM és SEMPP modellekből származó szimulált éves káreloszlások hisztogramjai

6.2. Viszontbiztosítás

6.2.1. CatXL és ECOMOR szerződések

Egy viszontbiztosítási szerződés megkötésekor a viszontbiztosító arra vállal kötelezettséget, hogy a direkt biztosítóhoz beérkező díj egy részének ellenében átvállalja annak bizonyos arányú/nagyságú szolgáltatását valamilyen portfóliójához tartozó ügyfél vagy szegmens esetében. A viszontbiztosítást vásárló biztosítót cedensnek szokás nevezni. Ebben a fejezetben megvizsgáljuk, hogy esemény alapú kártöbblet (*Catastrophy Excess of Loss* - CatXL) és ECOMOR (*Excess Coverage with Minimum or Retention*) viszontbiztosítási szerződések mellett a csonkolt Poisson pontfolyamat és SEMPP modellel milyen pénzáramlásokat produkál a szimuláció 1 éves időhorizonton mind a cedens, mind a viszontbiztosító részéről. Először Kerényi (2011) elemzését követve bemutatjuk a viszontbiztosítási szerződések általános matematikai leírását, majd definiáljuk a vizsgált szerződéstípusokat. Egy viszontbiztosítási szerződést a viszontbiztosítót megillető P_1 díj mellett egy T mérhető függvény jellemzi; ha a fizetendő kárigény x , akkor a cedens Tx , míg a viszontbiztosító $x - Tx$ összeget fizet. Ha az ügyfeleknek fizetendő X kár eloszlása $F(x)$, TX és $X - TX$ eloszlása pedig rendre $F_0(x)$ és $F_1(x)$, a díjszámítást pedig várható érték elvvel végezzük, akkor fennállnak az alábbi egyenlőségek:

$$\int x dF(x) = P,$$

$$\int x dF_0(x) = \int Tx dF(x) = P_0,$$

$$\int x dF_1(x) = \int (x - Tx) dF(x) = P_1,$$

ahol $P_0 = P - P_1$. A cedens kockázatát így ezen jelölések mellett a $(P_0, F_0(x))$, a viszontbiztosító kockázatát pedig $(P_1, F_1(x))$ pár írja le.

A CatXL és az ECOMOR szerződési formák is a nem arányos viszontbiztosítási szerződések csoportjába tartoznak. Előbbi esetében a T transzformációt az össze-

sített éves X kárnagyság helyett $\sum X_i$ összegre kell alkalmazni, ahol X_i az egyes, adott évben bekövetkező katasztrófákhoz tartozó kárnagyság. CatXL esetén a viszontbiztosító egy adott természeti katasztrófa után akkor fizet, ha az ide tartozó összesített kár meghalad egy M küszöböt, melyet eseményenkénti megtartásnak nevezünk, vagyis

$$TX_i = \min(X_i, M).$$

A viszontbiztosító védelme érdekében gyakran állapítanak meg egy adott katasztrófánál fizetendő felső határt is. ECOMOR szerződés esetén nincs önrész, legyen $X_{n,n} \leq \dots \leq X_{1,n}$ az $X_1, \dots, X_n$ adott évben bekövetkező károk rendezett statisztikája. Ekkor ha x az éves teljes kárnagyság, egy $\text{ECOMOR}(k)$ szerződés kárkifizetési függvénye a viszontbiztosító szempontjából az alábbi:

$$x - Tx = X_{1,n} + X_{2,n} + \dots + X_{k-1,n} - (k-1)X_{k,n}.$$

Magyarul, a cedenstől átvállalják a legnagyobb $k-1$ kárt, azonban annak érdekében, hogy a viszontbiztosító védve legyen, ezt az értéket csökkentik a k -adik kár $(k-1)$ -szeresével. A szerződés díja természetesen erősen függ k megválasztásától. A viszontbiztosítási díjnál alkalmazható különböző díjelvekről jó összefoglaló található Ekheden és Ola (2014) tanulmányában.

6.2.2. Szerződések árazása

A szerződéseket az egyszerűség kedvéért várható érték elven árazzuk, P_1 -et a szimuláció segítségével kalkuláljuk. A cedens és a viszontbiztosító által felmerülő költségek érdemben nem befolyásolnák az eredményeket, így figyelmen kívül hagyjuk őket. A CatXL szerződés általában komoly jogi és díjazási problémákat vet fel, hiszen nehéz meghatározni, hogy mit kell egy adott eseményhez tartozó kárnak tekinteni. A gyakorlatban ezt általában adott természeti katasztrófa kezdetétől számított időkorláttal oldják meg, amely árvízkároknál 168 óra (Kerényi 2011). Az elemzés során azzal a feltételezéssel élünk, hogy az egyes klaszterek, melyeknek hossza igen eltérő lehet, külön természeti katasztrófát jelölnek, és tartalmazzák az adott katasztrófa-hoz tartozó

összes kárt. Emlékezhethünk, hogy a küszöbértékünk nagyjából 3.3 millió dollár. Az alábbi táblázat a szimulációból számított, különböző eseményenkénti megtartások és felső határok melletti P_1 értékeket tartalmazza:

Megtartás	Limit = 100	Limit = 200	Limit = 300	Limit = ∞
5	131.48	176.13	204.25	424.35
10	115.91	159.28	186.89	405.82
20	96.40	137.40	160.88	380.69
30	83.67	122.57	148.32	362.74
40	74.38	111.37	136.29	348.54
50	67.17	102.45	126.59	336.74

7. táblázat. Várható éves viszontbiztosítói kifizetések (millió USD-ben) eseményenkénti megtartás és különböző viszontbiztosítási limitek mellett, SEMPP modell

Ahogy azt várhattuk, az eseményenkénti megtartás növelésével csökken a viszontbiztosítónak fizetendő P_1 díj. Természetesen a felső határ esetében pont az ellenkező mondható el, ami egy érdekes eredmény, hogy 300 millió dolláros felső határ mellett kevesebb mint fele akkora P_1 értéke, mint limit nélkül. Az SEMPP modell által szimulált átlagos éves kárnagyság 452 millió dollár, ebből arra következtethetünk, hogy az éves kárnagyság relatív szórása aránylag kicsi. Ugyanezen paraméterek mellett a csonkított Poisson pontfolyamat modell eredményei az alábbiak:

Megtartás	Limit = 100	Limit = 200	Limit = 300	Limit = ∞
5	81.05	128.26	159.95	450.25
10	79.04	125.15	156.33	445.28
20	74.93	118.97	148.98	435.48
30	70.84	112.09	142.22	425.99
40	66.88	105.62	135.64	416.92
50	63.15	101.89	129.47	408.31

8. táblázat. Várható éves viszontbiztosítói kifizetések (millió USD-ben) a csonkolt Poisson–Pareto modell alapján, különböző eseményenkénti megtartás és limit esetén.

Mivel ez a modell teljes függetlenséget feltételez, azaz nem veszi figyelembe az események bekövetkezésének ideje, valamint nagyságuk közötti összefüggési struktúrát, átlagban kevesebb eseményt kapunk évente, ezentúl a folyamatnak nincs klaszterezési tendenciája. Ebből adódóan az egyes katasztrófák ritkábban érik el az eseményenkénti megtartás szintjét, így általánosságban kisebb a viszontbiztosítónak fizetendő P_1 díj. Érdekes azonban, hogy felső határ nélkül mindenhol magasabb P_1 értéke, mint az SEMPP modell szimulációjánál. Az átlagos éves kárnagyság nem sokban tér el a két modell között, a csonkolt Poisson pontfolyamat modellnél 455 millió dollár, így az eltérés az eloszlás jobboldali viselkedésének tudható be.

Az ECOMOR(k) viszontbiztosítási szerződés díjait különböző k értékek mellett az alábbi táblázat tartalmazza:

k	SEMPP modell	PPM modell
3	423.05	430.49
4	435.24	440.58
5	442.28	446.84
6	446.52	450.86

9. táblázat. Várható éves viszontbiztosítói kifizetések (millió USD-ben) ECOMOR(k) szerződések esetén, a SEMPP és a PPP modellek alapján.

A két modell által kalkulált P_1 díjak igen hasonlóak, valamivel nagyobb értékeket kapunk a csonkított Poisson pontfolyamat esetében. Ez természetesen részben betudható annak, hogy $\xi_{PPM} > \xi_{SEMPP}$, azonban az öngerjesztő jelölt pontfolyamat modell skálaparaméterének sztochasztikus volta miatt ez az eredmény nem volt garantált.

7. Összefoglalás

Dolgozatunk célja egy olyan sztochasztikus modell felépítése volt, mely az extrém károk bekövetkezésének időpontja és az egyes események nagysága interakciójának figyelembevételével képes leírni az amerikai NFIP portfóliót érő károk folyamatát. A tanulmány során bemutattuk a modell illesztésének menetét, kiemelve a Hawkes folyamaton alapuló struktúra előnyeit és hiányosságait. A hagyományos módszertannal ellentétben, az SEMPP modell alkalmasnak bizonyult a főként nagyobb természeti katasztrófákhoz kapcsolódó klasztereződési tendencia megragadására. Az illesztés az empirikus adatokra statisztikai szempontból kielégítő, valamint a szimulációs algoritmus alkalmazható különböző viszontbiztosítási struktúrák árazására, CatXL viszontbiztosítási szerződés esetén a modell a cedens szempontjából magasabb díjakat eredményez, míg az ECOMOR-nál fordított a helyzet. Levonható a konklúzió, hogy a modell adekvát a természeti katasztrófák által okozott extrém károk leírására, de természetesen számtalan módon továbbfejleszthető.

A dolgozatban használt modell komplexitásának növelésével párhuzamosan a vizsgálat tárgya és a lehetséges alkalmazási területek bővítése is kitűzhető új kutatási célként. Utóbbira kézenfekvő példák lehetnek a következők. Megvizsgálható, hogy mennyiben befolyásolja egy adott biztosító szavatolótőke szükségletét, valamint kockázati mérőszámait és a tartalékok nagyságát az, ha a kárfolyamatot öngerjesztő modellekkel írjuk le, így figyelembe véve a nagy károk klasztereződési tendenciáját. További alkalmazási terület lehet Michaletzky (2001) jegyzetében tárgyalt rizikófolyamatok vizsgálata, melyekkel egy biztosító csődvalószínűsége írható le. Az előbb említett jegyzet is foglalkozik a témával bizonyos végtelen várható értékű eloszlások mellett, azonban csak független, azonos eloszlású valószínűségi változók figyelembe-

vételével. A modell bővítése több irányban is lehetséges. Érdekes lehet megvizsgálni különböző kapocsfüggvények mellett azt, hogy a kiterjesztett általánosított Pareto eloszlás (EGPD) segítségével szimultán leírható-e a káreloszlás széleinek viselkedése, ahogy ezt teszi Naveau és tsai. (2016) csapadékintenzitás esetében. Ezentúl kézenfekvő lenne megvizsgálni, hogy a szeizmológiában gyakran alkalmazott modellek (Ogata 1998; Ogata 1988; Daley és Vere-Jones 2006) mintájára akár tűzesetek, akár árvizek vagy egyéb természeti katasztrófa által okozott extrém károk földrajzi elhelyezkedésének hozzáadásával milyen konklúziókat lehet levonni. Ez például az NFIP adatbázis esetében elősegítené a különböző régiók besorolását kockázati szintek szerint.

Hivatkozások

- Albrecher, H., J. Beirlant és J.L. Teugels (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780470772683. URL: <https://books.google.hu/books?id=qkgzDwAAQBAJ>.
- Balkema, A. A. és L. de Haan (1974. szept.). *Limit Laws For Order Statistics*. Econometric Institute Archives 272067. Erasmus University Rotterdam. DOI: 10.22004/ag.econ.272067.
- Banyár, József (2016). *Életbiztosítás*. Budapest: Budapesti Corvinus Egyetem. URL: <https://unipub.lib.uni-corvinus.hu/2587/>.
- Castillo-Mateo, Jorge, Ana C. Cebrián és Jesús Asín (2023). “RecordTest: An R Package to Analyze Non-Stationarity in the Extremes Based on Record-Breaking Events”. *Journal of Statistical Software* 106.5, 1–28. old. DOI: 10.18637/jss.v106.i05.
- Chavez-Demoulin, V., A. C. Davison és A. J. McNeil (2005). “Estimating value-at-risk: a point process approach”. *Quantitative Finance* 5.2, 227–234. old. DOI: 10.1080/14697680500039613.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer. ISBN: 9781852334598. URL: <https://books.google.hu/books?id=2nugUEaKqFEC>.
- Daley, D.J. és D. Vere-Jones (2006). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Probability and Its Applications. Springer New York. ISBN: 9780387955414. URL: <https://books.google.hu/books?id=Ev2iXQKItpUC>.
- Dennis, JE és RB Schnabel (1982). “Numerical Methods for Nonlinear Equations and Unconstrained Optimization Prentice Hall”. *Englewood Cliffs*.
- Dombrowski, Timothy, Dimuthu Ratnadiwakara és V. Carlos Slawson Jr. (2021). “The FIMA NFIP’s Redacted Policies and Redacted Claims Datasets”. *Journal of Real Estate Literature* 28.2, 190–212. old. DOI: 10.1080/09277544.2021.1876435.

- EIOPA (2020). *2020 Review of Solvency II*. URL: https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii/2020-review-solvency-ii_en.
- Ekheden, Erland és Hössjer Ola (2014). “Pricing catastrophe risk in life (re)insurance”. *Scandinavian Actuarial Journal* 2014.4, 352–367. old. DOI: 10.1080/03461238.2012.695747.
- Embrechts, P., C. Klüppelberg és T. Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg. ISBN: 9783540609315. URL: <https://books.google.hu/books?id=BX0I2pICfJUC>.
- Ferreira, Ana és Laurens de Haan (2015). “On the block maxima method in extreme value theory: PWM estimators”. *The Annals of Statistics* 43.1, 276–298. old. DOI: 10.1214/14-AOS1280.
- Ferro, Christopher A. T. és Johan Segers (2003. ápr.). “Inference for Clusters of Extreme Values”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.2, 545–556. old. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00401.
- Fisher, R. A. és L. H. C. Tippett (1928). “Limiting forms of the frequency distribution of the largest or smallest member of a sample”. *Mathematical Proceedings of the Cambridge Philosophical Society* 24.2, 180–190. old. DOI: 10.1017/S0305004100015681.
- Fréchet, Maurice (1927). “Sur la loi de probabilité de l’écart maximum”. *Ann. de la Soc. Polonaise de Math.* URL: <https://cir.nii.ac.jp/crid/1572261550191409280>.
- Gilleland, Eric és Richard W. Katz (2016). “extRemes 2.0: An Extreme Value Analysis Package in R”. *Journal of Statistical Software* 72.8, 1–39. old. DOI: 10.18637/jss.v072.i08.
- Gnedenko, B. (1943). “Sur La Distribution Limite Du Terme Maximum D’Une Série Aléatoire”. *Annals of Mathematics* 44.3, 423–453. old. ISSN: 0003486X, 19398980. URL: <http://www.jstor.org/stable/1968974>.

- Haan, L. de és A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York. ISBN: 9780387239460. URL: <https://books.google.hu/books?id=catZC117d7gC>.
- Harte, David (2010). “PtProcess: An R Package for Modelling Marked Point Processes Indexed by Time”. *Journal of Statistical Software* 35.8, 1–32. old. DOI: 10.18637/jss.v035.i08.
- Hawkes, Alan G. (1971. ápr.). “Spectra of some self-exciting and mutually exciting point processes”. *Biometrika* 58.1, 83–90. old. ISSN: 0006-3444. DOI: 10.1093/biomet/58.1.83.
- Hawkes, Alan G. és David Oakes (1974). “A cluster process representation of a self-exciting process”. *Journal of Applied Probability* 11.3, 493–503. old. DOI: 10.2307/3212693.
- Horn, Diane P és Jared T Brown (2017). *Introduction to the national flood insurance program (nfip)*. Congressional Research Service Washington, DC.
- Horn, Diane P és Baird Webel (2018). “Private flood insurance and the national flood insurance program”. *Washington, DC: Congressional Research Service*.
- Kerényi, István (2011). *Vizszontbiztosítás*. 5. köt. Aktuárius jegyzetek. Budapest: Budapesti Corvinus Egyetem Operációkutatás és Aktuáriustudományok Tanszék. URL: <https://unipub.lib.uni-corvinus.hu/388/>.
- Leadbetter, M. R., Georg Lindgren és Holger Rootzén (1983). “Exceedances of Levels and k-th Largest Maxima”. *Extremes and Related Properties of Random Sequences and Processes*. New York, NY: Springer New York, 31–48. old. ISBN: 978-1-4612-5449-2. DOI: 10.1007/978-1-4612-5449-2_2.
- Li, Xiaoting, Christian Genest és Jonathan Jalbert (2021). “A self-exciting marked point process model for drought analysis”. *Environmetrics* 32.8, e2697. DOI: <https://doi.org/10.1002/env.2697>.
- Longnaker, JL (1962). “History of Insurance Law”. *Ins. LJ*, 642. old.
- McDonald, Robert és Anna Paulson (2015. máj.). “AIG in Hindsight”. *Journal of Economic Perspectives* 29.2, 81–106. old. DOI: 10.1257/jep.29.2.81.

- Michaletzky, György (2001). *Kockázati folyamatok. Egyetemi jegyzet. (Risk processes. Lecture Notes) (in Hungarian)*. Budapest: Eötvös Kiadó, 164. old. URL: <https://eprints.sztaki.hu/2496/>.
- Mises, R. (1936). *Wahrscheinlichkeit Statistik und Wahrheit: Einführung in die Neue Wahrscheinlichkeitslehre und Ihre Anwendung*. Springer Vienna. ISBN: 9783709120675. URL: <https://books.google.hu/books?id=IfWEBwAAQBAJ>.
- Naveau, Philippe, Raphael Huser, Pierre Ribereau és Alexis Hannart (2016). “Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection”. *Water Resources Research* 52.4, 2753–2769. old. DOI: <https://doi.org/10.1002/2015WR018552>.
- Nocedal, Jorge és Stephen J. Wright (2006). “Quadratic Programming”. 448–492. old. DOI: 10.1007/978-0-387-40065-5_16.
- Ogata, Yoshihiko (1981). “On Lewis’ simulation method for point processes”. *IEEE Transactions on Information Theory* 27.1, 23–31. old. DOI: 10.1109/TIT.1981.1056305.
- (1988). “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. *Journal of the American Statistical Association* 83.401, 9–27. old. DOI: 10.1080/01621459.1988.10478560.
- (1998). “Space-time point-process models for earthquake occurrences”. *Annals of the Institute of Statistical Mathematics* 50, 379–402. old. DOI: <https://doi.org/10.1023/A:1003403601725>.
- Papastathopoulos, Ioannis és Jonathan A. Tawn (2013). “Extended generalised Pareto models for tail estimation”. *Journal of Statistical Planning and Inference* 143.1, 131–143. old. ISSN: 0378-3758. DOI: <https://doi.org/10.1016/j.jspi.2012.07.001>.
- Pickands, James (1975). “Statistical Inference Using Extreme Order Statistics”. *The Annals of Statistics* 3.1, 119–131. old. ISSN: 00905364, 21688966. URL: <http://www.jstor.org/stable/2958083> (elérés dátuma 2025. 05. 13.).
- Reinhart, Alex (2018). “A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications”. *Statistical Science* 33.3, 299–318. old. ISSN: 08834237,

21688745. URL: <https://www.jstor.org/stable/26770999> (elérés dátuma 2025. 05. 13.).
- Reiss, Rolf-Dieter és Michael Thomas (2007). “Statistical Models for Exceedance Processes”. 247–262. old. DOI: 10.1007/978-3-7643-7399-3_9.
- Resnick, S.I. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer New York. ISBN: 9780387759531. URL: <https://books.google.hu/books?id=AGwQBwAAQBAJ>.
- Vaughan, E.J. (1996). *Risk Management*. Wiley. ISBN: 9780471107590. URL: <https://books.google.hu/books?id=0UeJQgAACAAJ>.
- Wooldridge, J.M. (2016). *Introductory Econometrics: A Modern Approach*. Cengage Learning. ISBN: 9781473754393. URL: <https://books.google.hu/books?id=HveHAQAACAAJ>.

Etikai nyilatkozat

*generatív mesterséges intelligencia rendszer vagy szolgáltatás felhasználásáról
teljesítmény-értékeléshez leadott feladat elkészítése során*

NAGYBALI TAMÁS hallgató (Neptun-kód: B83LM5) a Budapesti Corvinus Egyetem Etikai Kódexén alapuló és eszerint érvényesülő etikai felelősségem tudatában kijelentem, hogy a BIZTOSÍTÁSI ÉS PÉNZTUDOMÁNYI MATEMATIKA BA/BSc/MA/MSc/Posztgraduális képzési program keretében meghirdetett DIPLOMAMUNKA SZEMINÁRIUM tantárgyhoz (tantárgykód: MM4N2SD4) kapcsolódó feladat (vizsgadolgozat/projektmunka/szakdolgozat/diplomamunka/disszertáció) teljesítéséhez generatív mesterséges intelligencia rendszert vagy szolgáltatást (a továbbiakban: GMI) az alábbiak szerint vettem igénybe.

(Aláhúzással jelölje a megfelelő elemet vagy elemeket az alábbi listában!)

- 1) Nem használtam fel GMI rendszert vagy szolgáltatást
- 2) GMI alapú rendszert használtam fel
 - a) Szakirodalom kutatásához, válogatásához
rendszer neve, verziója: _____
 - b) Adatok előzetes elemzéséhez, értékeléséhez
rendszer neve, verziója: _____
 - c) Számítógépes program kód írásához, fejlesztéséhez, ellenőrzéséhez
rendszer neve, verziója: ChatGPT 4.0
 - d) Szöveges tartalom stilisztikai és nyelvtani javításához, parafrázáláshoz
rendszer neve, verziója: _____
 - e) Képi vagy audiovizuális tartalom fejlesztéséhez, létrehozásához
rendszer neve, verziója: _____
 - f) Egyes szövegrészek fordításához (pl. magyar és idegen nyelv között)
rendszer neve, verziója: _____
 - g) Egyéb módon, és pedig: _____
rendszer neve, verziója: _____

Kijelentem, hogy GMI által előállított új tartalomelemeket az egyetemi GMI szabályozás által megengedett módon használtam fel és a leadott munkában (pl. lábjegyzetben „GMI által generált tartalom” megjegyzéssel és a terjedelem azonosításával) mindenhol megjelöltem.

Tudomással bírok arról, hogy az intézmény az általam benyújtott munkákat mesterséges intelligencia detektor alkalmazással átvizsgálhatja, s ennek eredményétől függően kezdeményezheti etikai lépések megtételét.

Budapest, 2025.05.16

Nagybali Tamás
aláírás