

EÖTVÖS LORÁND TUDOMÁNYEGYETEM  
TERMÉSZETTUDOMÁNYI KAR

VALÓSZÍNŰSÉGELMÉLETI ÉS STATISZTIKA TANSZÉK



## Nem-élet tartalékszámítás egyedi kárinformációk alapján szimulációs módszerrel

**Készítette:**

Pogonyi Péter

Biztosítási és Pénzügyi Matematika MSc

Aktuáriusi szakirány

**Témavezető:**

Pónuzs Róbert

BUDAPEST

2016

# Tartalomjegyzék

<b>1. IBNR kárszám becslése</b>	<b>6</b>
1.1. Gyakoriság előrevetítése . . . . .	6
1.2. Korrigált késési eloszlásfüggvény . . . . .	8
<b>2. Kárgyakorisági modell illesztése</b>	<b>14</b>
<b>3. Károk IBNER korrekciója</b>	<b>20</b>
3.1. Általánosított lineáris modell . . . . .	21
<b>4. Káreloszlás becslése</b>	<b>23</b>
<b>5. Monte-Carlo szimuláció</b>	<b>27</b>
<b>6. Adatelemzés, modellezés</b>	<b>30</b>

# Bevezetés

A dolgozatban a nem-élet biztosításban leggyakrabban használt háromszöges tartalékszámítási módszer helyett új tartalékolási modellt építünk fel. A háromszög alapú technikáknál sok információt veszünk azzal, hogy a károkat háromszögekbe tömörítjük. Ez kétféle módon is megjelenik. Ha például 10 000 kárunk van 8 évre tekintve, akkor ezt 36 pontba fogjuk besűríteni egy háromszögön belül, továbbá nem javít az információ mennyiségén az sem, ha ugyanezt 10 000 kár helyett 100 000-re nézzük. Tehát ha például az IBNR tartalékokról nem csak egy pontbecslés szeretnénk, hanem egy valószínűségi eloszlást, akkor az információvesztés miatt a háromszög pontjaiból nehéz lenne eloszlást illeszteni. A célom, hogy az [1] cikkben ismertetett irányelvekhez konkrét gyakorlati számítási módszereket rendeljek, továbbá az így felépített modellt implementáljam és alkalmazzam valós adatokon. Két főbb, a véletlenül alapuló kársoporttal fogunk foglalkozni, az IBNR és RBNS károkkal. Az elsőre egy sztochasztikus modellt fogunk felépíteni, a másodikra pedig egy determinisztikus. A megfelelő eloszlások, paraméterek kiválasztásához különböző eljárásokat mutatok be, továbbá az elméleti tárgyalás után egy mintavételezéssel randomizált, transformált, anonim valós adatsoron bemutatom a gyakorlati megvalósítást is. A végén pedig bemutatom az eredményeket. Az új modellhez a keretrendszert Pietro Parodi mutatta be, amely lényegében az imént említett információvesztést próbálja elkerülni azzal, hogy minden egyedi kárinformációt belevesz a modellezésbe. Ez egy kárgyakoriság-kárnagyság alapú modell (vagy összetett kockázatok modellje), tehát külön van egy eloszlás a károkra és a kárnagyságokra, majd Monte-Carlo szimuláció segítségével meghatározzuk az aggregált káreloszlást. Korábban is foglalkoztak sztochasztikus tartalékolással. 1993-ban Norberg ([11]) hasonlóan felosztotta a károkat Rendezett, IBNR, RBNS csoportokra és azt feltételezte, hogy a kárdarabszám egy nem-homogén Poisson folyamat generálja, a kárnagyságok pedig egy Általánosított Pareto eloszlásból származnak. Norberg modelljét továbbfejlesztte Haastrap és Arjas ([8]) egyedi kárinformációkat is belevette a modellbe, mint például bekövetkezés, kifizetés, lezárás, újrainvitás, stb.. Többet készítették esettanulmányt a sztochasztikus modellekről, amelyek általánosságban jobbnak bizonyultak, mint a régi háromszö-

ges módszerek. A háromszög alapú tartalékolást még olyan időszakban találták ki, amikor a számítási kapacitás elég alacsony volt így egyszerű papíron is kiszámítható módszer kellett. Azonban manapság a nagy teljesítményű számítógépekkel ki lehet számolni akár kárunként is a tartalékokat. Az újféle tartalékolás előnyeként több helyen kiemelik, hogy nagyon rugalmas és könnyen formálható, hogy minél jobb becslést adjon a tartalékokra, továbbá részletesebb eredményt kapunk. Hátránya a háromszögekkel szemben pedig a bonyolultsága, hiszen előfordulhat, hogy nem megfelelően kalibráljuk a modellünket vagy feleslegesen túlparaméterezzük.

A dolgozat első fejezetében a teljes kárdarabszámokat szeretném megbecsülni a megfigyelt időszakokra. A nehézséget azok a károk jelentik, amelyet bekövetkeztek az adott periódus alatt, de még nem jelentették be őket (IBNR). Ennek a becsléséhez egy ún. bejelentés késési eloszlásfüggvényt számolunk ki, mely megmondja hogy mennyi a valószínűsége, hogy bejelentik a kárt az időszakon belül, feltéve hogy be is következett.

A második fejezetben a kárdarabszámok alapján egy kárgyakorisági modellt illesztünk az általános  $(a, b, 0)$  eloszlások közül. Illetve pár gyakorlati diagnosztikai módszert nézünk, hogy mely eloszlást célszerű a gyakorlatban használni.

A harmadik fejezetben az kárnagyság eloszlás becsléséhez szükséges előkészületeket tesszük meg. A kárainkat egy ún. IBNeR (Incurred, But Not enough Reported) faktorról korrigáljuk. Amennyiben bejelentenek egy kárt és beállítanak rá valamilyen tételes tartalékokat (lehet), előfordulhat hogy a végleges kifizetési összeg eltér ettől. Tehát azon károkat, amelyhez tartozik még tételes tartalék kiigazítjuk ezekkel a faktorokkal. Az IBNeR faktorok becsléséhez röviden átvesszük a GLM módszert is, de több eljárással is megközelíthető a probléma.

A negyedik fejezetben történik a kárnagyság eloszlásának becslése. Először általánosan áttekintjük, hogy értelmezzük az eloszlásfüggvényt, majd konkrét módszert mutatunk be annak meghatározására. A becsléshez szükségünk lesz az extrém érték eloszlásokra is, hiszen előfordulhatnak kiugróan magas károk és ezeket is le kell tudnunk kezelni. Az eloszlásillesztéshez szükséges paraméterbecsléseket részletezve leírom.

Az ötödik fejezetben elsőként a szimulációs algoritmusokat ismertetem, amellyel meghatározzuk az aggregált káreloszlásokat. Három főbb kötelezettségi kategória van: az IBNR, RBNS és UPR. Az utóbbival csak említés szintjén foglalkozunk, mivel teljes mértékben megegyezik az IBNR szimulációjával. Az IBNR károk szimulálása a korábbi fejezetekben felépített kárgyakoriság és kárnagyság modellek alapján lesz elvégezve, míg az RBNS károknál a GLM modellből kapott IBNER faktorokat fogjuk használni.

Az utolsó fejezetben valós adatokon építem fel a modellt. Ahogy korábban is említettem, az adatok anonimitás végett transzformálva és randomizálva lettek oly módon, hogy az eredeti káradatokat ne tükrözzék, azokat ne lehessen visszafejteni. Lakossági vagyónbiztosításokhoz tartozó károkkal fogok foglalkozni. Az ágazatnak vannak sajátosságai, amit figyelembe kell venni a modell felépítésénél. Majd az eredményeket prezentálom és összehasonlítom a megvalósult értékekkel.

# 1. fejezet

## IBNR kárszám becslése

### 1.1. Gyakoriság előrevetítése

A kárinformáció hiányából adódó tartalékot két főbb elemre lehet osztani. Az egyik az IBNR, a másik pedig az IBNER. Az IBNR olyan károkat foglal magába, amelyek már megtörténtek, de még nem jelentették be őket. Az IBNER pedig a már bejelentett károk várható tartalékváltozását próbálja megfogni. Az utóbbival később fogunk foglalkozni. Az általános gyakorlatban háromszöges módszerrel határozzák meg az IBNR kárdarabszámot, viszont most más megközelítésben lesz tárgyalva.

Először az ultimate (végső) kárgyakoriságot fogjuk előrejelezni, mivel számunkra a bejelentett károk darabszáma ismert, így ha megbecsüljük az időszakra vonatkozó végső kárdarabszámot, akkor a kettő különbsége pont az IBNR kárgyakoriságot fogja adni. A modell felépítése követi [1] logikáját és jelöléseit. Matematikailag a következőképp fog kinézni.

Legyen  $t$  a mai dátum és az előbbieket alapján a teljes kárdarabszámot szeretnénk meghatározni a  $[0, t]$ -n bejelentett károokra. Jelen környezetben az időintervallum tetszőleges hosszúságú lehet, azonban elemzés során felszeleteljük évekre. Ez semmiben nem befolyásolja a modellezést. Feltételezzük, hogy a károkat egy véletlen eloszlás alapján jelentik be. A kárbejelentések késésének eloszlásfüggvényét  $F(t)$ -vel jelöljük.

Most feltételezni fogjuk, hogy a kárbejelentés késésének eloszlása ismert, azonban a 1.2 fejezetben eldobjuk ezt a feltételt és meghatározzuk az eloszlásfüggvényét.

Továbbá legyen  $\nu(t)$  a kárgyakoriság függvény, amely azt mutatja meg, hogy  $t$  időpillanatban mennyi kár következett be. Ennél a függvénynél jelenhetnek meg szezonális hatások, hiszen például a mezőgazdasági és lakossági vagyonszámítások esetében tipikusan nyáron történnek a káresemények nagy része, míg az év többi részében általában kevésbé jellemző a magas kárgyakoriság. Ez magába foglalja azt is, hogy ahol magasabb a kárgyakoriság, ott nagyobb lesz a kár bekövetkezésének va-

lőszínősége is. Jelöljük  $\mu_t$ -vel az összes kárdarabszámot a  $[0, t]$  intervallumon. Ekkor a várható értéke:

$$E(\mu_t) = \int_0^t \nu(x) dx.$$

Példaként nézzük azt az egyszerű esetet, amikor egyenletes kárgyakoriságot feltételezünk, azaz  $\nu(t) = c$  valamilyen  $c$  konstansra. Ekkor

$$E(\mu_t) = \int_0^t c dx = c \cdot t.$$

Továbbiakban  $r_t$  fogja jelölni a  $[0, t]$ -n bejelentett károk darabszámát. Nyilván  $r_t \leq \mu_t$ . A bejelentett károk várható értéke:

$$E(r_t) = \int_0^t \nu(x) \cdot F(t-x) dx. \quad (1.1)$$

Vizsgáljuk meg, hogy szemléletesen mit is jelent a (1.1) integrál. Ha most is feltételezzük az egyenletes kárgyakoriságot, akkor ahogy haladunk  $x$ -el az időben  $t$ -ig, akkor  $t-x$  tartani fog a 0-hoz, így  $F(t-x)$  pedig egyre inkább csökken. Ez pontosan azt jelenti, hogy a minél közelebb vagyunk a mai dátumhoz  $t$ -hez, annál kevésbé valószínű hogy bejelentik még az adott időpontig történt károkat.

A  $t$  időpillanatban nyilván ismerjük a bejelentett károk számát (hiszen be lettek jelentve), tehát  $r_t$  ismert és  $\mu_t$ -t kell becsülnünk valahogy. Jelölje  $\mu_t$  becslését  $\hat{\mu}_t$ . A következőképp becsülünk:

$$\hat{\mu}_t = \frac{\int_0^t \nu(x) dx}{\int_0^t \nu(x) \cdot F(t-x) dx} r_t = \frac{r_t}{E(r_t)} \cdot E(\mu_t). \quad (1.2)$$

Vegyük mindkét oldal várható értékét (feltéve, hogy  $\mu_t$  és  $r_t$  függetlenek):

$$E(\hat{\mu}_t) = E\left(\frac{r_t}{E(r_t)} \cdot E(\mu_t)\right) = \frac{E(r_t)}{E(r_t)} \cdot E(\mu_t) = E(\mu_t).$$

Tehát tényleg egy torzítatlan becslést kaptunk. Ha ismét feltesszük, hogy egyenletes kárgyakoriság van, továbbá a kárbejelentés késésének eloszlása  $\lambda$ -paraméterű exponenciális, akkor az adódik hogy:

$$\begin{aligned} \hat{\mu}_t &= \frac{c \cdot t}{\int_0^t c \cdot (1 - e^{-(t-x)\lambda}) dx} r_t = \frac{t}{\int_0^t (1 - e^{-u\lambda}) du} r_t = \frac{t}{t - \int_0^t e^{-u\lambda} du} r_t = \\ &= \frac{t}{t - \left(\frac{-e^{-t\lambda}}{\lambda} + \frac{1}{\lambda}\right)} r_t = \frac{t}{t + \frac{e^{-t\lambda}}{\lambda} - \frac{1}{\lambda}} r_t. \end{aligned}$$

Az eddig felépített modellben végig folytonos időben gondolkoztunk, de a gyakorlatban ez nem kivitelezhető, ezért szükség van a diszkrétizálásra. Legyen  $0 = x_0 \leq x_1 \leq \dots \leq x_n = t$  egy felosztása a  $[0, t]$  intervallumnak (nálunk jelen esetben napokat fognak jelenteni az  $x$ -ek). A (1.2) képlet diszkrét változata:

$$\hat{\mu}_t = \frac{\sum_{i=0}^n \nu(x_i) \cdot (x_i - x_{i-1})}{\sum_{i=0}^n \nu(x_i) \cdot F(t - x_i) \cdot (x_i - x_{i-1})} r_t.$$

A látványosabb szemlélet miatt ismét feltételezzünk egyenletes kárdarabszám eloszlást. Ekkor,

$$\hat{\mu}_t = \frac{\sum_{i=0}^n c}{\sum_{i=0}^n c \cdot F(t - x_i)} r_t.$$

Most írjuk ki az egyenlet jobb oldalán látható kifejezést:

$$\frac{c + c + \dots + c}{c \cdot F(t) + c \cdot F(t - 1) + \dots + c \cdot F(0)}. \quad (1.3)$$

A számlálóban és a nevezőben is  $t$  darab tag van. A számláló azt fejezi ki, hogy minden nap  $c$  darab kár következett be. A nevezőben pedig aszerint súlyozzuk ezeket a  $c$ -ket, hogy épp hol állunk az időben a  $[0, t]$  intervallumon. Nyilván a nevező mindig kisebb lesz a számlálónál (kivéve ha degenerált eloszlásunk van, de ez kizárható). Tekintsük az első napot: összesen  $c$  kár történt, a bejelentés késés eloszlásunk alapján a  $t$  időponting várhatóan  $F(t)$  részét jelentik be, tehát megnézzük mekkora az  $\frac{\text{össz kárszám}}{\text{várhatóan bejelentett kárszám}}$  arány és rászorozzuk a ténylegesen bejelentett kárdarabszámra.

A másik megközelítés pedig hasonló a [1]-ban leírtakhoz. Egyszerűsítsünk, azaz vegyük ki az iménti példából a  $c$ -t, mint mennyiségi faktort. Ekkor a számlálóban 1-eseket adunk össze, a nevezőben pedig valószínűségeket. Ezt úgy lehet értelmezni, mintha napokat tekintenénk konkrét kárdarabszámok helyett. Fent 1-eseket adunk össze, hiszen ott a károk 100%-ban bejelentettek, míg a nevezőben az első napra  $F(t)$  részt jelentenek be, második napra már csak  $F(t - 1)$ , és így tovább. A [1] terminológiában ezeket megszolgált napoknak nevezik. Így felírhatjuk, hogy:

$$\hat{\mu}_t = \frac{\text{Napok száma}}{\text{Megszolgált napok száma}} \cdot r_t.$$

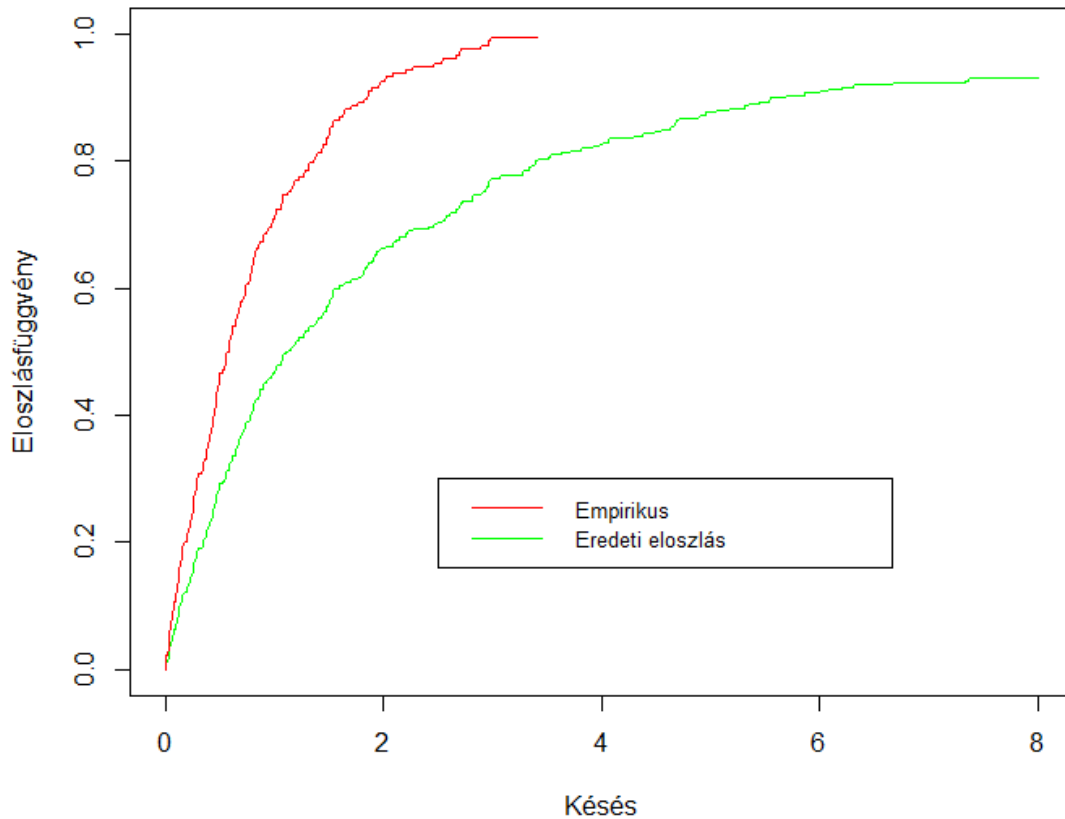
Amennyiben az  $u$  időpontig (ahol  $u < t$ ) szeretnénk előrevetíteni a kárszámot, akkor csak a szummázás és az integrálás határát kell kicserélni.

## 1.2. Korrigált késési eloszlásfüggvény

Korábban feltételeztük, hogy a kárbejelentések késésének eloszlásfüggvénye ismert. Azonban ez nem teljesen igaz, mert mi csak egy konkrét állapot szerint, egy korlátos megfigyelési ablakon keresztül látjuk a bejelentett károk késedelmi idejét. Például, ha egy adott 1 éves időszakot nézünk, akkor maximum csak 1 éves késést tudunk megfigyelni, de könnyen előfordulhat, hogy 1 évnél későbbi bejelentések is befutnak, úgy hogy a megfigyelt időszakban következett be a kár. Az 1.1-es ábra jól szemlélteti a jelenséget. A példában 300 darab kárt generáltam egyenletesen 4 éves időtávlatban, továbbá a károkhoz véletlenszerűen mintavételeztem bejelentés késéseket, melyek



(2,3)-paraméterű Pareto eloszlásból származnak. Ennek az eloszlásnak 3 a várható értéke. Azon károk, melyeknél így a bejelentés dátuma meghaladta volna a 4 évet eldobtam. A megfigyelt időszakban a legnagyobb bejelentés 3.41 év lett így. A piros vonal jelöli a megfigyelési ablakban bejelentett károk empirikus eloszlását, a zöld vonal pedig az eredeti eloszlást.



1.1. ábra. A bejelentési késési eloszlásfüggvény a megfigyelt ablakon keresztül piros vonallal, illetve az eredeti eloszlásfüggvény zöldel

Ezek előbbieik alapján korrigálnunk kell az eloszlásfüggvényünket, hogy figyelembe vegyük az adott időszakban bekövetkezett, de még be nem jelentett károkat is. Formálisan a következőképp néz ki.

Tegyük fel, hogy a megfigyelési időszakunk a  $[0, a]$  intervallum. A megfigyelt sűrűségfüggvényt jelöljük  $f_a(t)$ -vel és keressük az eredeti  $f(t)$  sűrűségfüggvényt.

Legyen  $T_0$  egy valószínűségi változó, amely a kár bekövetkezését jelöli. A  $T$  pedig jelölje a bejelentés késésének hosszát, szintén mint valószínűségi változó. Tehát ez alapján  $f_a(t)$  a  $T$  változó feltételes sűrűségfüggvényét jelöli, azzal a feltétellel, hogy

a kár bekövetkezése és annak bejelentése megtörtént  $a$  előtt. Az imént bevezetett jelöléssel a következőképpen írhatjuk fel formálisan az empirikus sűrűségfüggvényt:

$$f_a(t) = P(T = t | T + T_0 \leq a). \quad (1.4)$$

Alkalmazzuk a Bayes-tételt.

$$P(T = t | T + T_0 \leq a) = \frac{P(T + T_0 \leq a | T = t)P(T = t)}{P(T + T_0 \leq a)}.$$

A jobb oldali kifejezésben számlálójában az első tagban alkalmazhatjuk a feltételt, a második tag pedig definíció szerint  $f(t)$ . A nevezőt pedig jelöljük  $B(a)$ -nak. Így tehát azt kapjuk, hogy

$$f_a(t) = \frac{P(T + T_0 \leq a | T = t)P(T = t)}{P(T + T_0 \leq a)} = \frac{P(T_0 \leq a - t)f(t)}{B(a)}.$$

A fenti levezetés nyilván a  $[0, a]$  intervallumra igaz, azon kívül az  $f_a(t)$  nulla. Tehát,

$$f_a(t) = \begin{cases} \frac{P(T_0 \leq a - t)f(t)}{B(a)} & t < a \\ 0 & t \geq a \end{cases} \quad (1.5)$$

Átrendezve (1.5)-t  $f(t)$ -re az alsó ágánál egy ismeretlen részt kapnánk, hiszen nem garantálja semmi, hogy  $f(t)$  nulla legyen az  $a$  felett. Ezért az lesz a feltevésünk, hogy  $a$ -tól egy  $f_\tau$  farokeloszlással folytatódik a sűrűségfüggvény. Tehát,

$$f(t) = \begin{cases} \frac{B(a)f_a(t)}{P(T_0 \leq a - t)} & t < a \\ f_\tau(t) & t \geq a \end{cases} \quad (1.6)$$

A (1.5) formulával szeretnénk meghatározni  $f_a(t)$  értékét. Először tekintsük a  $P(T_0 \leq a - t)$  tagot.

A legegyszerűbb esetben feltételezzük, hogy a károk egyenletesen következnek be a  $[0, a]$  intervallumon (így  $f_a(t)$  úgy is megfogalmazható, hogy egy  $[0, a]$ -ban véletlen  $t_0$  kezdőponttal indított szakasz, amelynek  $t$  a hosszúsága mekkora valószínűséggel esik bele a  $[0, a]$  intervallumba, lásd (1.5)). Tehát tegyük, fel hogy  $T_0$  egyenletes eloszlású valószínűségi változó. Ekkor,

$$P(T_0 \leq a - t) = \begin{cases} \frac{a-t}{a} = 1 - \frac{t}{a} & 0 \leq t \leq a \\ 0 & \text{különben} \end{cases}$$

Gyakorlatban azonban ez nem egy helytálló feltevés, hiszen a károk bekövetkezésének időpontja függ az adott portfóliótól és a kockázat típusától. A második esetben

a már korábban bevezetett  $\nu(t)$ -vel írjuk fel a fenti valószínűséget:

$$P(T_0 \leq a - t) = \frac{\int_0^{a-t} \nu(x) dx}{\int_0^a \nu(x) dx}. \quad (1.7)$$

Szemléletesen (1.7)-ben a számláló a  $[0, a - t]$ -n bekövetkezett károkat számolja meg, a nevezőben pedig az egész intervallumon bekövetkezett károkat. Ha feltesszük azt, hogy azért változik a  $\nu(t)$ , mert évente különböző a biztosítási portfóliók kitétsége (tehát növekszik vagy épp csökken az állomány), akkor a következőképp lehet felírni (1.7)-t:

$$P(T_0 \leq a - t) = \frac{\int_0^{a-t} \nu(x) dx}{\int_0^a \nu(x) dx} = \frac{\sum_{i=0}^{\lfloor a-t \rfloor - 1} \epsilon_i + \epsilon_{\lfloor a-t \rfloor} \cdot (a - t - \lfloor a - t \rfloor)}{\sum_{i=0}^{\lfloor a \rfloor - 1} \epsilon_i + \epsilon_{\lfloor a \rfloor} \cdot (a - \lfloor a \rfloor)}.$$

A fenti egyenletben  $\epsilon_i$  az  $i$ . időszak kitétségét jelöli. A későbbi implementáláshoz megnézzünk kétféle módszert, amely szerint meghatározható az eloszlásfüggvény.

Először leírjuk a feltételezéseket, amelyek mindkét megközelítésben ugyanazok:

1.  $f_a$  megfigyelt sűrűségfüggvény ismert a  $[0, a]$  intervallumon.
2.  $\int_0^a f_a(t) dt = 1$ .
3. Az  $f(t)$  sűrűségfüggvény időben nem változik, tehát nem használunk más eloszlást különböző bekövetkezési dátumokra.
4.  $f(t)$  nem függ a károk nagyságától.

## 1. módszer

Az első eljárásban hasonlóan számolunk, mint amikor eloszlást illesztünk momentumok módszerével, mégpedig:

1. Felteszünk valamilyen előzetes valószínűségi eloszlást  $f(t)$ -re.
2. A megfigyelt bejelentés késések alapján várható értéket (illetve ha szükséges szórásnégyzetet is) meghatározzuk.
3. Az (1.5) alapján kiszámoljuk  $f_a(t) - t$ .
4. Az előző pontban számolt elméleti  $f_a(t)$  becslés segítségével meghatározzuk az elméleti várhatóértéket (és szórásnégyzetet is, ha kell).
5. A megfigyelt várhatóértéknek egyeznie kell a becsült várhatóértékkel, ezért az egyenletrendszer numerikusan megoldva megkapjuk az eredeti paramétereket.

6. Így megkaptuk az eloszlást, mégpedig az imént kapott paraméterekkel.

Ha  $f(t)$ -re exponenciális eloszlást feltételezünk, akkor leegyszerűsödik a számítás, hiszen csak egy paraméterünk van. Bár ez a feltevés elég nyersnek tűnhet, azonban sokszor megfelelően modellezi a gyakorlatot. Az exponenciális eset megtalálható [1]-ben.

Mi most Pareto-eloszlásra fogjuk megnézni ezt a módszert, azzal a feltétellel, hogy a károk egyenletesen következnek be. Az európai  $(c, b)$ -paraméterű Pareto eloszlású változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0 & x \leq c \\ 1 - \left(\frac{c}{x}\right)^b & x > c \end{cases}$$

A sűrűségfüggvény:

$$f(x) = \begin{cases} 0 & x \leq c \\ b \cdot \left(\frac{c^b}{x^{b+1}}\right) & x > c \end{cases}$$

A (1.5) formulában egyedül a  $B(a)$  tag maradt ismeretlen. Mivel definíció szerint két valószínűségi változó összegéről van szó, ezért meghatározhatjuk konvolúciós képlettel az együttes eloszlást (feltéve, hogy a változóink függetlenek). Tehát,

$$B(z) = P(T + T_0 \leq z) = \int_{-\infty}^{\infty} P(T \leq z - x) \cdot f_{T_0}(x) dx.$$

Az integrál első tagjánál definíció szerint csak akkor nem nulla az eloszlásfüggvény, ha  $z - x > c$ , azaz  $z - c > x$ . A második tag pedig a  $[0, a]$ -n egyenletes eloszlású változó sűrűségfüggvénye, ami  $[0, a]$ -n kívül nulla. Két eset van: amikor a  $z \leq c$ , akkor a valószínűség 0, hiszen a Pareto eloszlású változó legalább a  $c$  értéket veszi fel. A másik eset, amikor  $z > c$ , ekkor viszont az imént említett feltételek miatt az integrálás határai 0 és  $\min(z - c, a)$ . Folytatva az előző (a Pareto-eloszlás második paraméterét  $b$ -vel jelöljük) képletet:

$$B(z) = \begin{cases} 0 & z \leq c \\ \int_0^{\min(z-c, a)} \left(1 - \left(\frac{c}{z-x}\right)^b\right) \cdot \frac{1}{a} dx & z > c \end{cases}$$

Ha  $z \rightarrow \infty$ , akkor az alsó ágban a minimum  $a$  lesz, továbbá az integrálon belül a  $z$ -t tartalmazó tag eltűnik, így könnyen látható, hogy 1 lesz az eloszlásfüggvény a végtelenben. Nekünk  $B(a)$ -t kell meghatározni, tehát a második ág fog teljesülni.

$$B(a) = \int_0^{a-c} \left(1 - \left(\frac{c}{a-x}\right)^b\right) \cdot \frac{1}{a} dx = \frac{a - c + \frac{c^b}{a^{b-1}} - c}{a}.$$

Itt vigyázni kell, hiszen ha behelyettesítjük  $a$ -t, akkor az eloszlásfüggvény képletéből adódik, hogy csak akkor áll fent az egyenlőség, ha  $a > c$ .

Most már felírhatjuk a sűrűségfüggvényt:

$$f_a(t) = \frac{a \cdot b \cdot \frac{c^b}{t^{b+1}} \cdot \left(1 - \frac{t}{a}\right)}{a - c + \frac{\frac{c^b}{a^{b-1}} - c}{b-1}}.$$

A minta alapján számolt várható érték és variancia segítségével felírjuk a következő egyenleteket:

$$E(X) = \int_c^a t \cdot f_a(t) dt = \frac{b(-a^2c^b + a^b c(a(2-b) + c(b-1)))}{(2-b)(ac^b + a^b(a(b-1) - bc))}.$$

$$D^2(X) = \frac{(b-1)b(-a^3c^b + a^b c^2(a(3-b) + (b-2)c))}{(3-b)(b-2)(ac + a^b(a(b-1) - bc))} + (E(X))^2$$

Numerikusan megoldva megkapjuk az eltolts paramétereket.

## 2. módszer

Az első módszer inkább egy elméletibb megoldást nyújtott az eloszlásfüggvény elnyújtására. Ha a megfigyelt adatainkat szeretnénk használni, akkor a következőképp járhatunk el.

1. Külső információk alapján, valamilyen farokeloszlást feltételezünk az  $[a, \infty]$  intervallumra.
2. Az (1.6) egyenletből jól látszik, hogy a nevező mindig kisebb mint 1, ezért lényegében felskálázzuk  $f_a(t)$  értékeit először.
3.  $B(a)$ -ra úgy lehet tekinteni, mint egy korrekciós faktor, amivel az előző pontban kapott felskálázott értéket hozzuk vissza. Tehát a  $B(a)$  értékét úgy határozzuk meg, hogy a sűrűségfüggvény integrálja  $[0, \infty]$ -n 1 legyen.

Ez egy hasonló eredményt ad, mint az első módszer. A [1] könyv szerzője említ egy megoldást, amely majdnem megegyezik a második módszerrel, azzal a különbséggel, hogy a transzformált sűrűségfüggvény véges tartójú lesz nála, ezáltal a letolt eloszlásfüggvény szintén véges pontban éri el az 1-et. Gyakorlati szempontból ez se rossz megoldás, de matematikailag precízebb végeredményt kapunk azáltal, ha valamilyen elosztást illesztünk vagy feltételezünk egy farokeloszlást.

## 2. fejezet

# Kárgyakorisági modell illesztése

Az előző fejezetben lettek meghatározva a projektált kárdarabszámok. A következő feladat egy gyakorisági modell illesztése. Általában a nem-élet károk esetében, a kárszámra  $(a, b, 0)$  eloszlású valószínűségi változókat szoktak használni. Röviden definiáljuk a fontosabb fogalmakat.

**2.1. Definíció.** Egy  $X$  valószínűségi változó  $(a, b, 0)$  eloszlású, amennyiben teljesül hogy:

$$p_{k+1} = \left(a + \frac{b}{k}\right) p_k, \quad (2.1)$$

ahol  $p_k = P(X = k)$  és  $k = 1, 2, 3, \dots$

Bizonyítás nélkül kimondjuk a következő tételt (megtalálható [2]-ben).

**2.2. Tétel.** *Ha  $X$  valószínűségi változó  $(a, b, 0)$  eloszlású, akkor  $X$  Poisson, Negatív-Binomiális vagy Binomiális eloszlású.*

Miért ezeket az eloszlásokat szeretik használni a gyakorlatban? Kárszámláló eloszlásként is felfoghatóak, hiszen a természetes számok halmazából kerülnek ki a lehetséges értékek, továbbá általában a gyakorlati problémák jól leírhatóak ezekkel az eloszlásokkal. Mivel teljesül az  $(a, b, 0)$  tulajdonság is, ezért könnyen rekurzív mód számolhatóak a valószínűségek, ezért pl. a Panjer rekurzió esetén egy viszonylag gyorsan számolható képletet kapunk az aggregált káreloszlás kiszámítására.

A nem-élet biztosítási aggregált káreloszlásoknál két modellt szoktunk elkülöníteni. Az első az egyéni kockázat modellje, a második pedig az összetett kockázatok modellje. Az előbbi esetben feltételezzük, hogy a veszélyközösségünk nem homogén, továbbá az  $X$  változóink az egy szerződésre jutó kifizetést mutatják, ezáltal rögzítetten  $n$  darab szerződésünk van. Az összkár ( $S$ ):

$$S = X_1 + \dots + X_n.$$

Az utóbbi esetben pedig káronként nézzük a kifizetéseket. Az első esettel ellentétben itt véletlen tagszámú összegünk van, hiszen nem ismerjük a károk darabszámát (pontosan ezt akarjuk meghatározni).

Az iménti megfontolások alapján a binomiális eloszlást általában csak az egyéni kockázatok modelljénél alkalmazzák, hiszen ott egy szerződésen vagy bekövetkezik a kár, vagy nem. A dolgozatban a összetett kockázatok modelljét tárgyaljuk, ezért a Poisson és a Negatív-binomiális eloszlás lesz első sorban érdekes számunkra.

Tegyük fel, hogy megbecsültük az ultimate kár darabszámot és már csak egy eloszláshoz van szükségünk a gyakorisági modellhez. Már csak az a kérdés maradt, hogy hogyan válasszuk ki a megfelelő eloszlást az adatainkhoz és mik lesznek a paramétereink?

Az első módszer (részletesebben [3]-ban) a várható értékhez és a varianciához kapcsolódik. Jól ismert, hogy egy Poisson eloszlású változó várható értéke megegyezik a szórásnégyzetével. Könnyen belátható, hogy Negatív-Binomiális változó esetében pedig a szórásnégyzet nagyobb (binomiális esetben pedig a szórásnégyzet kisebb). Ezek alapján képezzük a  $\frac{V}{M}$  hányadost ( $\frac{\text{variancia}}{\text{átlag}}$ ). Ha ez az érték nagyobb, mint 1, akkor próbálkozzunk negatív-binomiális eloszlással. Amennyiben 1-hez közeli (vagy akár 1 alatti), akkor pedig Poisson eloszlással. Fontos megjegyezni, hogy ha szignifikánsan eltér az 1-től az érték, akkor a Poisson eloszlás helyett mást illeszteni.

A fenti eljárásnak az a hibája, hogy nem feltétlen veszi figyelembe, hogy az adataink véletlen mintából származnak. Ez azt jelenti, hogy például ha legenerálunk 10 számot Poisson eloszlás szerint, akkor a  $V \setminus M$  hányados akár 0,3 vagy 2 is lehet. A másik megoldás a statisztikai hipotézisvizsgálat (megtalálható [3]-ben). Ezt a következőképp végezhetjük el:

1. Választunk egy  $\alpha$  szignifikanciaszintet, ami alapján a hipotézisvizsgálatot elvégezzük.
2. Kiszámoljuk a  $V \setminus M$ -eket évenként.
3. Becsüljük meg a  $V \setminus M$  eloszlásfüggvényét, majd kiszámoljuk az  $[F(\frac{\alpha}{2}), F(1 - \frac{\alpha}{2})]$  intervallum végpontjait. Itt  $F(k, x)$  egy két változós függvény, hiszen függ az évtől is.
4. Ha a  $V \setminus M$  kisebb mint  $F^{-1}(k, \frac{\alpha}{2})$ , akkor legyen Binomiális. Ha nagyobb mint  $F^{-1}(k, \frac{\alpha}{2})$  és kisebb mint  $F^{-1}(k, 1 - \frac{\alpha}{2})$ , akkor legyen Poisson. Egyébként pedig Negatív-Binomiális.

Azonban ez az algoritmus nem teljesen korrekt, hiszen nem veszi figyelembe, hogyha a kitettségünk változik az évek során (például növekszik a szerződéseinknek

a száma). Továbbá, ha a  $V \setminus M$  nem szignifikánsan tér el az 1-től, akkor általánosságban mindig a Poisson modellhez jutunk, ami nem feltétlenül a legjobb illeszkedést nyújtja. A Poisson eloszlás egyik fő hibája, hogy azokban az esetekben mutat jó illeszkedést, amikor szezonális káraink vannak. Ha például tekintjük a mezőgazdasági károkat és feltesszük, hogy a kárdarabszámot 500 várható értékű Poisson változóval modellezzük, akkor a szórás nagyjából 22 lesz, míg egy szárazabb nyár esetén sokkal nagyobb lehet a károk száma, mint például 522.

A korábbi módszereket és megfigyeléseket összegezve [3] szerint a következő algoritmus ajánlott:

1. Ha a modellünk az egyéni kockázatok modelljére támaszkodik, akkor használjunk binomiálisat.
2. Az összetett kockázatok modellje esetében pedig próbáljunk első sorban NB eloszlással modellezni. Ha pedig nem kapunk értelmes paramétereket az NB eloszlásra, akkor próbáljuk meg a Poisson eloszlást.

Ezek után már csak az eloszlásillesztés kell elvégezni. Bár korábban sokszor hivatkoztunk a Poisson eloszlásra, most a szükség és érthetőség kedvéért definiáljuk.

**2.3. Definíció.** Egy  $X$  valószínűségi változót  $\lambda (> 0)$ -paraméterű Poisson eloszlásúnak nevezünk, ha

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \text{ahol } k \geq 0.$$

Tudjuk, hogy Poisson eloszlás esetében  $E(X) = \lambda = Var(X)$ . Hasonlóan itt definiáljuk a Negatív-Binomiális eloszlást is.

**2.4. Definíció.** Egy  $X$  valószínűségi változót  $(r, \beta)$ -paraméterű Negatív-Binomiális eloszlásúnak nevezünk, ha

$$P(X = k) = \binom{r+k-1}{k} \left( \frac{1}{1+\beta} \right)^r \left( \frac{\beta}{1+\beta} \right)^k \quad \text{ahol } k \geq 0 \quad (2.2)$$

Itt is nyilván megköveteljük a paraméterek pozitivitását.

Két módszer alapján illeszthetünk eloszlást. Nézzük elsőként a maximum likelihood becslést. Alapesetben tudjuk, hogy a Poisson eloszlásnál a becslés megegyezik a mintaátlaggal. Amennyiben a kitettségünk nem állandó a különböző éve során, akkor ki kell igazítanunk a kárgyakoriságokat. Tegyük fel, hogy  $N$  évre van adatunk, továbbá jelöljük az  $N+1$ . év kitettségét  $E_{N+1}$ -el. Erre azért van szükségünk, hogy a jelenlegi állapot szerinti kitettségre hozzuk a kárgyakoriságot. Ha  $X_i$  jelöli az  $i$ . évi kárgyakoriságot, akkor formálisan a következőképp néz ki a kiigazítás:

$$X_i^E = X_i \cdot \frac{N \cdot E_{N+1}}{\sum_{j=1}^N E_j}.$$



Heurisztikusan ez azt jelenti, hogy az  $i$ . év kárgyakoriságát leosztjuk egységnyi kitettségre (az összes évet beleértve) és utána visszahozzuk az új évi kitettség „átlagára”. Figyeljük meg, hogy amennyiben  $E_{N+1}$  az első  $N$  év kitettségeinek átlagához közeli szám, akkor a fenti formulában szereplő hányados 1-hez közeli érték.

Most számoljuk ki a maximum likelihood becslést. A likelihood függvény:

$$L = f(X_1|\lambda) \cdots f(X_N|\lambda) = \frac{e^{-\lambda} \lambda^{X_1^E}}{X_1^E!} \frac{e^{-\lambda} \lambda^{X_2^E}}{X_2^E!} \cdots \frac{e^{-\lambda} \lambda^{X_N^E}}{X_N^E!}.$$

Ezek alapján a loglikelihood:

$$\ln L = -N\lambda + \ln \lambda \cdot \sum_{i=1}^N X_i^E - \ln \prod_{i=1}^N X_i^E.$$

Számoljuk ki a maximum helyét  $\lambda$  szerint.

$$0 = \frac{\partial \ln f(X_1, \dots, X_N|\lambda)}{\partial \lambda} = -N + \frac{\sum_{i=1}^N X_i^E}{\lambda}.$$

Innen pedig:

$$\hat{\lambda} = \frac{\sum_{i=1}^N X_i^E}{N} = \frac{\sum_{i=1}^N X_i \cdot \frac{N \cdot E_{N+1}}{\sum_{j=1}^N E_j}}{N} = \frac{\sum_{i=1}^N X_i}{\sum_{j=1}^N E_j} \cdot E_{N+1}.$$

A momentumok módszerre Poisson eloszlásnál egybeesik a maximum likelihood becsléssel.

Vizsgáljuk meg a (2.2)-ben definiált NB eloszlás milyen eredményre vezet a momentumok módszerével. Jelöljük a mintaátlagot az előbb levezetett  $\hat{\lambda}$ -val (itt nem a Poisson eloszlás paraméterét jelöljük  $\lambda$ -val), a szórásnégyzetet pedig  $\hat{\nu}$ -vel. A várható értéke egy  $(r, \beta)$ -paraméterű NB eloszlású  $X$  változónak:

$$\hat{\lambda} = E(X) = r\beta.$$

Innen adódik, hogy

$$\hat{r} = \frac{\hat{\lambda}}{\beta}.$$

A másik egyenletünk a varianciára:

$$\hat{\nu} = D^2(X) = r\beta(1 + \beta).$$

Behelyettesítjük  $r$ -et.

$$\hat{\nu} = \frac{\hat{\lambda}}{\beta} \beta (1 + \beta).$$

Átrendezéssel kapjuk, hogy

$$\hat{\beta} = \frac{\hat{\nu}}{\hat{\lambda}} - 1.$$

Tekintsük a maximum likelihood becslést NB eloszlásra is (a számítás egy speciális esete megtalálható [10]-ban). Itt is ki szeretnénk igazítani a paramétereinket a kitettséggel, ezért egy lineáris kapcsolatot feltételezünk az  $r$  és az  $E_i$  között (szintén az  $i$ . évi kitettséget jelöli). Tehát  $r = a \cdot E$ . Ez a feltételezés helytálló, hiszen előbb láttuk, hogy egy NB eloszlású valószínűségi változó várható értéke  $r\beta$  és mivel a kitettség nő, ezáltal  $r$  is nő, így tehát nagyobb lesz a várható kárgyakoriság. A likelihood függvény ebben az esetben:

$$L = f(X_1|(a, \beta)) \cdots f(X_N|(a, \beta)) = \binom{aE_1 + X_1 - 1}{X_1} \left(\frac{1}{1 + \beta}\right)^{aE_1} \left(\frac{\beta}{1 + \beta}\right)^{X_1} \cdots \binom{aE_N + X_N - 1}{X_N} \left(\frac{1}{1 + \beta}\right)^{aE_N} \left(\frac{\beta}{1 + \beta}\right)^{X_N}.$$

Emeljünk ki egy általános tagot és vegyük a logaritmusát.

$$\ln(f(X_i|(a, \beta))) = \ln\left(\binom{aE_i + X_i - 1}{X_i}\right) + aE_i \ln\left(\frac{1}{1 + \beta}\right) + X_i \ln\left(\frac{\beta}{1 + \beta}\right).$$

Először deriváljunk  $\beta$  szerint. Így az első tag teljesen eltűnik

$$\frac{\partial \ln(f(X_i|(a, \beta)))}{\partial \beta} = -a \frac{E_i}{(1 + \beta)} + \frac{X_i}{\beta(1 + \beta)}.$$

Ahhoz, hogy  $a$  szerint lederiváljuk, először az első tagot át kell írni más alakba.

$$\ln\left(\binom{aE_i + X_i - 1}{X_i}\right) = \ln\left(\frac{(aE_i + X_i - 1) \cdots (aE_i)}{X_i!}\right) = \sum_{j=0}^{X_i-1} \ln(aE_i + X_i - 1 - j) - \ln(X_i!)$$

Cseréljük ki az index változót, mégpedig  $l = X_i - 1 - j$ . Vegyük észre, hogy ekkor  $l$  ugyanazt a tartományt futja be, azzal a kivétellel, hogy most növekednek a tagok. Így,

$$\ln\left(\binom{aE_i + X_i - 1}{X_i}\right) = \sum_{l=0}^{X_i-1} \ln(aE_i + l) - \ln(X_i!).$$

Ha  $a$ -ban deriváljuk a kifejezést, akkor a második tag eltűnik. Tehát a derivált,

$$\frac{\partial \ln(f(X_i|(a, \beta)))}{\partial a} = \sum_{l=0}^{X_i-1} \frac{E_i}{aE_i + l} - E_i \ln(1 + \beta).$$

A logaritmus miatt összegezzük a kapott eredményeket mindegyik mintaelemre.

$$0 = \sum_{i=1}^N E_i \left( \sum_{l=0}^{X_i-1} \frac{1}{aE_i + l} - \ln(1 + \beta) \right) \quad (2.3)$$

$$0 = \sum_{i=1}^N -a \frac{E_i}{(1 + \beta)} + \frac{X_i}{\beta(1 + \beta)} \quad (2.4)$$

$$r\beta = \frac{\sum_{i=1}^N E_i}{N} a\beta = \frac{\sum_{i=1}^N X_i}{N}. \quad (2.5)$$

A harmadik egyenlet a változó várható értékéből, illetve a kitétségre tett linearitási feltételből következik. Az egyenleteket numerikusan megoldva megkapjuk a paraméterbecsléseket. Fontos megjegyezni, hogy az  $r$ -nél szintén az  $N + 1$ . kitétség adatot kell használni.

Amennyiben van egy kárgyakoriság becslésünk az ultimate kárdarabra, akkor látható, hogy az IBNR kárgyakoriság is ugyanolyan eloszlásból származik, amelynek a várható értéke az IBNR kárdarabszám, a  $V \setminus M$  pedig az ultimate becslés  $V \setminus M$ -je.

Így az IBNR becslés szórása nem más, mint  $\sqrt{V \setminus M \cdot (\hat{\mu}_t - r_t)}$ . A  $\hat{\mu}_t - r_t$  tag az IBNR kárgyakoriság várható értéke, ezt felszorozzuk a  $V \setminus M$  aránnyal, ami az IBNR kárgyakoriság varianciája lesz így.

Poisson esetben is teljesen analóg módon ki lehet számolni a szórást.

## 3. fejezet

# Károk IBNER korrekciója

Eddig a kárgyakoriságra próbáltunk valamilyen modellt felállítani. A kárnagyságokra is szeretnénk hasonlóan eloszlást illeszteni. Mielőtt ezt megtennénk, elő kell készíteni az adatainkat.

Ha egy kár bekövetkezik, akkor a biztosító beállítja a kárhoz tartozó tételes tartalékot, amelyet a jövőbeli kárkifizetések fedezetére fog használni. Ahogy haladunk előre az időben, előfordulhat hogy a biztosító már kifizetett a kárra valamilyen mértékű összeget, viszont a tételes tartalék nagysága is változhatott a kifizetésen kívül. A kár lezárásig összegyűlt kifizetések összege (tehát a káron már nincs tételes tartalék, illetve nem lesznek további kifizetések) és az adott állapot szerinti kifizetések és tételes tartalék összegének különbségét nevezzük IBNER mennyiségnek (Incurred, but not enough reported). Tehát tudjuk egy állapot szerint mennyit fizettünk ki és még mennyit kell, viszont a végső összeg el fog térni ennek az összegétől. Ha nem konkrétan az IBNER mennyiségét szeretnénk megkapni, hanem hogy mivel kellene rászorozni az adott állapot szerinti tartaléokra, hogy megkapjuk a végső kifizetési összeget, akkor IBNER faktorokról beszélünk. Egy rövid példán keresztül bemutatjuk az IBNER faktort és mennyiséget.

Év	Kifizetés	Tartalék	IBNER	Végső összeg	IBNeR faktor
2006	0	1500	800	2300	1.53
2007	300	1300	700	2300	1.5384
2008	1000	1200	(-)100	2100	0.916
2009	2000	200	100	2300	1.5
2010	2300	0	0	2300	0

A táblázatban egy 2006-ban bekövetkezett kár szerepel, amely 2010-ben záródik le. A végső kifizetett összeg 2300. Az IBNER faktorokat a  $\frac{\text{Végső összeg-Kifizetés}}{\text{Tartalék}}$  képletel számoljuk. A 2008-as évben már kifizettünk 1000-et és képeztünk 1200 tartalékot, tehát túl-tartalékoltunk, a többi évben pedig alul-tartalékoltunk. Az imént tett megfigyelésből adódik, hogy túl-tartalékolás esetén az IBNER faktor kisebb, mint 1, ellenkező esetben pedig nagyobb, mint 1. Akkor mondhatjuk, hogy pontos a tartalékbecslésünk, ha ez az érték átlagosan 1 körül ingadozik.

Milyen módszerek léteznek az IBNER kezelésére?

1. **Nem számolunk IBNER-al.** Ha nincs megfelelő információ káronként a tartalékokról, illetve a kifizetésekről.
2. **Csak lezárt károkat tekintünk.** Ebben az esetben torzítani fognak az IBNER értékeink, hiszen a nagy károk esetén több mint valószínű, hogy a kár nincs lezárva még, tehát nem vesszük be az elemzésbe. Így [3] szerint ez csak akkor járható út, ha az elemzés során használunk olyan eljárást, ami torzítatlanná teszi a mintánkat.
3. **Háromszöges módszer.** Ha évenkénti információnk van minden kárra, akkor háromszögekkel is becsülhetjük az IBNER faktorokat. Megtalálható [3]-ban.
4. **Általánosított lineáris modell.** Az IBNER faktorok modellezése GLM-el. Ezzel fogunk most foglalkozni részletesebben.

### 3.1. Általánosított lineáris modell

Az IBNER faktorok több változótól is függhetnek. A hagyományos lineáris regresszióval helyett azonban egy sokkal általánosabb módszerrel fogjuk modellezni a faktorokat. Röviden tárgyaljuk, hogy hogyan épül fel az általánosított lineáris modell (GLM). A GLM egyik fő előnye az általános lineáris modell és lineáris regresszióhoz képest, hogy a magyarázni kívánt változó nem feltétlenül normális eloszlású, hanem egy sokkal univerzálisabb ún. exponenciális családból származik. Továbbá az imént említett modelleknél a magyarázó változóinknak additív hatása volt a magyarázandó változóra, míg ez nem feltétlenül van így és a GLM segítségével multiplikatív hatást is beleépítünk a modellünkbe. A modellt elsősorban [4] és [5] alapján tárgyaljuk. Az GLM három fő részből tevődik össze:

1. **Véletlen elem:**  $\underline{Y}$ , ahol az  $\underline{Y}$  elemei függetlenek, továbbá exponenciális családból származnak (ezt lejjebb definiáljuk).

2. **Lineáris prediktor:**  $\eta = X\underline{\beta}$ , ahol  $X$  egy  $n \times n$ -es mátrix.
3. **Link függvény:**  $g$  differenciálható és monoton függvény, amelyre  $E[\underline{Y}] = \underline{\mu} = g^{-1}(\underline{\eta}) = g^{-1}(X\underline{\beta})$ .

**3.1. Definíció.** Egy  $X$  változó exponenciális eloszláscsaládból származik, ha a sűrűségfüggvénye felírható, mint

$$f(X, \theta_i; \phi) = \exp \left\{ \frac{Y_i \theta_i - b(\theta_i)}{a(\phi)} + c(Y_i; \phi) \right\}.$$

A definícióban szereplő függvények az eloszlástól függenek. A  $\phi$ -t nevezzük skála-paraméternek, a  $\theta_i$ -t pedig kanonikus paraméternek. Az utóbbinak a várható értékre van hatása. Az  $a$  függvényt nevezik diszperziós függvénynek, amelynek az alakja  $\frac{\phi}{\omega_i}$ , ahol  $\omega_i$  a súlyt jelöli (ami általában a kitettség). A  $b(\theta_i)$  a kumuláns függvény, amely nem függ  $Y_i$ -től és kétszeresen differenciálható. A  $c$  függvény nincsen különösebb szerepe. Példák exponenciális eloszláscsaládba tartozó eloszlásokra: Normális, Gamma, Exponenciális, Poission, Tweedie. . . .

Az egyik legfontosabb tulajdonsága ezeknek az eloszlásoknak, hogy az eloszlás szórása és várható értéke egyértelműen meghatározza magát az eloszlást. Továbbá a variancia függ a várható értéktől. Ezt nevezzük *varianciafüggvénynek*. Jelöljük  $V(x)$ -el a varianciafüggvényt, ekkor a következő egyenleteket írhatjuk fel:

$$E(Y_i) = \mu_i = b'(\theta_i)$$

$$D^2(Y_i) = b''(\theta_i)a(\phi) = \frac{\phi V(\mu_i)}{\omega_i}.$$

Gyakorlati szempontból, tehát ismerjük a magyarázó változóink értékeit, továbbá a hozzájuk tartozó magyarázandó változóértékeket. Keressük azokat a  $\beta$  együtthatókat, amelyekkel a fenti formátumban tudjuk modellezni a magyarázandó mennyiséget.

Mi az IBNER faktorokat szeretnénk modellezni, ahol a magyarázó változók között lesznek a függőkár hányad, tartalék nagysága, kárnagyság, bejelentés és bekövetkezés éve, stb.. Különböző szelekciós kritérium (mint pl. az AIC) alapján kiszűrjük majd a nem szignifikáns erővel rendelkező változókat.

## 4. fejezet

# Káreloszlás becslése

Miután korrigáltuk a károkat a megfelelő IBNER faktorokkal, elkezdhetjük becsülni a káreloszlásunkat. Tehát adottak a historikus kárösszegeink, amelyek különböző évekből származhatnak. Ha például egy kár 2004-ben következett be, de 2012-ben jelentették be, akkor számolnunk kell a pénz időértékével is, hiszen a 2012-es átlagkárok valószínűleg nagyobbak, mint a 2004-esek. Továbbá számolni lehet azzal is, ha valamilyen oknál fogva megváltozik a károk típusának összetétele, például egy természeti katasztrófa miatt növekedett a kárnagyság, kárgyakoriság vagy egy biztonsági intézkedés miatt épp hogy csökkent. Az előbb említett kockázatváltozás nem lesz említve külön, hiszen ilyen esetben mindig el lehet szeparálni az adott kártípust és önállóan elvégezni a káreloszlás becslését.

A feladatunk, hogy meghatározzuk a sima és késési inflációval korrigált káreloszlást. Ezáltal az eloszlásfüggvényünk függeni fog a kár bekövetkezési és késési idejétől. A [1] jelölést alkalmazva, jelölje  $F_X(x)$  annak a valószínűségét, hogy egy  $X$  kár kisebb vagy egyenlő mint  $x$ . A korrigált eloszlásunkat pedig jelölje  $F_X^{t_0, \delta}(x)$ , ami azt adja meg, hogy mennyi a valószínűsége annak, hogy egy  $X$  kár kisebb mint  $x$ , feltéve hogy a kár  $t_0$ -ban következett be és  $\delta$  idő után jelentik be. Formálisan:

$$F_X^{t_0, \delta}(x) = P(X \leq x | t_0\text{-ban következett be és } \delta \text{ idő után jelentik be}). \quad (4.1)$$

A következő egyszerű, de gyakorlatban jól működő feltevésünk lesz. Tudjuk, hogy az  $F_X(x)$  alapeloszlás ismert (lejjebb lesz a kiszámítása tárgyalva), ezért azt tesszük fel, hogy a (4.1) részben definiált eloszlásfüggvény az alapeloszlás egy átskálázása a két inflációs rátával. Jelölje  $r$  a károk inflációs rátáját és  $s$  pedig a késésből adódó inflációt. Ha a vizsgálati időpont  $t$ , akkor a (4.1) átírva a következő:

$$F_X^{t_0, \delta}(x) = F_X \left( \frac{(1+r)^{t-t_0} \cdot x}{(1+s)^\delta} \right). \quad (4.2)$$

A számlálóban lényegében a mai értékére hozzuk fel a károkat. A nevezőben pedig annak a hatását kezeljük le, hogy a később bejelentett károk általában nagyobbak. Ebben az esetben  $s$  pozitív, míg ha a korán bejelentett károk a nagyobbak, akkor negatív.

Ezek után térjünk rá arra, hogy hogyan is kell meghatározni az alap kárösszeg eloszlást. Itt egy újabb témakört érintünk, mégpedig az extrém érték eloszlásokat (részletesebb leírást adnak a [7], [6] könyvek). A káraink között valószínűleg van jó néhány olyan eset, amikor túlságosan nagy a kárösszeg és emellett ritkán fordulnak elő. Ezért megbontjuk a kárainkat úgymond kis károokra és nagy károokra. Azt a határt, ami elválasztja a két csoportot *küszöbnek* (jelölése  $\mu$ ) fogjuk nevezni. A továbbiakhoz szükségünk lesz a következő definícióra:

**4.1. Definíció.** Azt mondjuk, hogy egy  $X$  valószínűségi változó *általánosított Pareto eloszlású* (GPD, General Pareto Distribution)  $(\mu, \xi, \sigma)$  paraméterekkel, ha a  $G_X(x)$  eloszlásfüggvénye:

$$G_X(x) = \begin{cases} 1 - (1 + \xi \cdot \frac{x-\mu}{\sigma}) & \xi \neq 0 \\ 1 - e^{-\frac{x-\mu}{\sigma}} & \xi = 0 \end{cases}$$

Az alábbi tétel mutatja meg számunkra miért hasznos a GPD eloszlás.

**4.2. Tétel** (Pickand-Balkema-de Haan). *Legyenek  $X_1, X_2, \dots$  független, azonos eloszlású valószínűségi változók. Továbbá jelöljük  $F_u$ -val az  $u$  küszöbvre vett feltételes eloszlást, azaz*

$$F_u(x) = P(X - u \leq x | X > u)$$

*Ekkor  $F_u(x) \xrightarrow{D} G(x)$ , ha  $u \rightarrow \infty$ , ahol  $G(x)$  GPD eloszlású.*

Tehát a tétel azt mondja, hogy a küszöbérték alatti része eloszlásától függetlenül a farokeloszlás modellezhető GPD-vel. Ezek alapján az kárnagyág eloszlásunk (jelölje  $F_X(x)$  a küszöb alatti károk eloszlását):

$$F(x) = \begin{cases} F_X(x) & 0 \leq x \leq \mu \\ F_X(\mu) + (1 - F_X(\mu))G(x) & x > \mu \end{cases}$$

ahol  $F(x)$  a kis károokra illesztett eloszlást jelöli. Azonban ahhoz, hogy el tudjuk végezni az eloszlásillesztést, meg kell adnunk egy  $\mu$  küszöbértéket, illetve a GPD eloszlás paramétereit.

## Paraméterek becslése

Először a küszöbindexet fogjuk meghatározni. Definiáljuk az ún. *Várható küszöb függvényt* (angolul mean excess value function, tehát a "küszöb feletti közép függ-



vény"), aminek az angol neve jól leírja a lényegét, tehát az küszöbnél nagyobb értékek küszöb feletti részének várható értékét nézzük. Ha  $e(\mu)$ -vel jelöljük a függvényt, akkor:

$$e(\mu) = E(X - \mu | X > \mu) = \frac{\sigma + \xi\mu}{1 - \xi}.$$

Az utóbbi egyenlőség bizonyítása megtalálható [6]-ben. Jól látható, hogy ez a függvény lineáris a küszöbértékben. Tehát ha ábrázoljuk az  $e(\mu)$  függvényt, akkor úgy kell megválasztani a küszöbértéket, hogy a grafikonunk lineáris trendet mutassunk. Fontos megjegyezni, hogy matematikailag ez nem egy precíz eljárás, de a leggyakrabban alkalmazott a gyakorlatban és jól közelíti a valóságot. Konkrét alkalmazása a szimulációs fejezetben történik.

Két megközelítésben megnézzünk egy másik eljárást is, amellyel meghatározhatjuk a küszöbértéket. Definiáljunk előtte a Kolmogorov-Szmirnov távolságot.

**4.3. Definíció.** Legyen  $X_1, \dots, X_n$  véletlen minta és  $F_{n,\mu}(x) = \frac{\sum_{i=1}^n 1_{\{\mu \leq X_i \leq x\}}}{\sum_{i=1}^n 1_{\{X_i > \mu\}}}$  empirikus eloszlása az  $F_\mu(x) = P(X \leq x | X > \mu)$  (tehát  $F_\mu$  a farokeloszlás  $\mu$  paramétertől tekintve) eloszlásfüggvénynek. Ekkor  $F_{n,\mu}(x)$  és  $G(x)$  Kolmogorov-Szmirnov távolsága:

$$d^\mu(F_{n,\mu}(x), G(x)) = \sup_{x \in \mathcal{R}} |F_{n,\mu}(x) - G(x)|$$

Továbbá normalizált KS távolságról hívjuk a  $d_{NORM}^\mu(F_{n,\mu}(x), G(x))$ -ot, amelyet a következőképp számolunk:

$$d_{NORM}^\mu(F_{n,\mu}(x), G(x)) = d^\mu(F_{n,\mu}(x), G(x)) \cdot \sqrt{\sum_{i=1}^n 1_{\{X_i > \mu\}}}.$$

Az alábbi algoritmussal a küszöbindexet számolja ki a KS távolság segítségével.

1. Beállítunk egy kezdeti  $\mu_k$  küszöbindexet ( $k = 0$ ).
2. Elvégezzük a paraméterbecslést a GPD eloszlásra.
3. Meghatározzuk a  $F_{n,\mu}(x)$ -et.
4. Meghatározzuk a GPD eloszlás mintaelemeken felvett helyét.
5. Kiszámoljuk a normalizált KS távolságot.
6.  $k$ -t növelve 1-el futtatjuk a ciklust  $n$ -ig.
7. Kiválasztjuk azt a küszöbindexet, amelyre a KS távolság minimális.

Célszerű a küszöbindex beosztást úgy választani, hogy ne legyen túlságosan számításigényes, de ne túl nagy léptéket határozzunk meg a pontok között. Belátható, hogy ez a küszöbindex becslés konzisztens lesz.

A másik megközelítés is hasonlóan működik. A lényege, hogy mind a  $\xi$  paramétert, mind a normalizált KS távolságot ábrázoljuk  $\mu$  függvényeként. A grafikonon azt a küszöbindexet választjuk, ahol a  $\xi$  paraméter relatíve stabilizálódik, illetve a normalizált KS távolság kicsi. Az illesztett modell illeszkedésvizsgálatáról az adatelemzéses résznél beszélek röviden.

A GPD eloszlás paramétereinek a becslése maximum likelihood módszerrel vagy Hill-becsléssel történik. Az olvasó részletesebb olvashat róla [7] vagy [9].

## 5. fejezet

# Monte-Carlo szimuláció

Az aggregált káreloszlás többféle módszerrel is kiszámolható. A három leggyakrabban használt eljárás a Monte Carlo szimuláció, a Gyors Fourier Transzformáció és a Panjer-rekurzió. Az utóbbi két módszer esetében szükséges diszkretizálni az eloszlásfüggvényeinket, illetve a két főbb különbség a módszerek között, hogy míg a Panjer-rekurzió egy relatíve könnyebben implementálható és számolható formula, ezzel együtt a futási ideje lassabb a Fourier-transzformációhoz képest.

A legelterjedtebb eljárás azonban a Monte-Carlo szimuláció. A módszer lényege, hogy egy rögzített szimulációs korlát (legyen  $N$ ) mellett először legenerálunk  $N$  darab kárarabszámot. Majd az összes scenárió minden egyes kárához generálunk egy kárösszeget is. Így scenárióként összeadjuk a kárösszegeket. Ha egy becslést szeretnénk kapni a tartalékra, akkor kiátlagoljuk a scenáriókat. Ha pedig egy biztonsági szint szerint szeretnénk értéket kapni a tartalékra, akkor kiszámítjuk a hozzá tartozó percentilist.

Miután meghatároztuk a kárgyakorisági és kárnagyság eloszlást, ismertetjük a szimulációs algoritmusokat (megtalálhatóak kicsit másként [1]-ben).

A célunk, hogy meghatározzuk az összes várható pénzbeli kötelezettségünket a már bekövetkezett károkra. A kötelezettségeinknek két fő alkotó eleme van.

1. **IBNR kötelezettség:** A korábbiak alapján ide tartoznak azon várható kifizetések, amelyek a bekövetkezett, de nem bejelentett károkból fakadnak. Ehhez építettük a modellünket elsősorban.
2. **RBNS kötelezettség:** Ide tartoznak a nyitott károk, tehát már bejelentették őket, de várhatóan változni fog a végső kifizetési összeg a jelenlegi tartalékhoz képest, ezért korrigálni kell az IBNER faktorokkal.

Tehát összefoglalva:

$$\text{Összes kötelezettség a bekövetkezett károkra} = \text{IBNR} + \text{RBNS}$$

**5.1. Megjegyzés.** A szimulációba be lehet venni azokra a károkra vonatkozó kifizetéseket is, amelyek a jövőben következnek be, de a szerződéses kötelezettség, amelyből a kár fakad, már fenn áll. Ez teljes mértékben megegyezik az IBNR modellel. Annyi a minimális különbség, hogy az IBNR kárszámokat rávetítjük a becslni kívánt időszak kitettségére.

## IBNR szimuláció

A szimulációs lépések a következőképp fognak kinézni:

---

### Algoritmus 1 IBNR aggregált káreloszlás kiszámítása

---

**Bemenet:**

- IBNR eloszlások paraméterei (attól függően, hogy összeadjuk-e az összes évet vagy pedig külön becsüljük):  $\lambda$  és  $V$ . Mintaelemek jelölése:  $n$ .
- Kárnagyság eloszlás paraméterei: empirikus +  $(\xi, \sigma, \mu)$  és a mintaelemek, melyeknek jelölése:  $X$ .

–  $N$  : Szimulációk száma

–  $Y$  : évek száma

**for**  $j = 1, \dots, Y$  **do**

**for**  $i = 1, \dots, N$  **do**

    Generáljunk le  $n_i$  darab IBNR kárt. Minden kárhoz generáljunk egy bekövetkezési dátumot.

**for**  $l = 1, \dots, n_i$  **do**

      Generáljunk minden egy  $X_l^i$  kárösszeget a kárnagyság eloszlásból, majd skálázzuk át bekövetkezési dátum szerint ((4.2)).

**end for**

**end for**

  Számoljuk ki scenáriónként az aggregált kárösszegeket:  $S_i = \sum_{k=1}^{n_i} \hat{X}_k^i$ ,  $i = 1, \dots, N$

**end for**

Rendezzük sorba az aggregált kárösszegeket és számoljuk ki az empirikus eloszlásfüggvényt.

**Kimenet:** Aggregált IBNR káreloszlás.

$F_S(x) = \frac{\max(i | S_i^* \leq x)}{N}$  ahol \* a rendezett mintát jelöli.

$(\alpha_1, \dots, \alpha_u)$  biztonsági szint melletti  $x$  percentilisek:

$x_{alpha_i} = F_S^{-1}(\alpha_i) \quad \forall i \in \{1, \dots, u\}$

---

Előfordulhat, hogy az utolsó évek növekvő bizonytalansága miatt az éveket egy adatként összesűrítjük. Ekkor egyértelműen csak egy évre kell szimulálni, továbbá csak egy kárgyakoriság eloszlásunk van.

## RBNS szimuláció

Az RBNS tartalékok igazából már a kárnagyság eloszlásának becslésénél meghatároztuk. A [1]-ben található egy kifinomultabb algoritmus, ahol az RBNS tartalék is sztochasztikus.

---

### Algoritmus 2 RBNS károk eloszlása

---

**Bemenet:**

- $X_i$  nyitott károk tartaléka, illetve a GLM modellhez szükséges bemeneti paraméterek.
- IBNER modell együtthatói.

**for**  $\forall X_i$  nyitott kárra **do**

$$RBNS_{X_i} = X_i \cdot IBNER(x, r, \dots)$$

**end for**

Ezután összegezzük az RBNS értékeket.

$$RBNS = \sum_{\forall i} RBNS_{X_i}$$

**Kimenet:**  $RBNS$

---

Amennyiben sztochasztikus modellünk van az IBNER faktorokra, akkor minden nyitott kárhoz generálunk egy dátumot, amikor a kárt lerendezik és egy mintával IBNER faktor eloszlásból, a végső kárösszegre szorozzuk fel a nyitott kárt.

## 6. fejezet

# Adatelemzés, modellezés

A szakdolgozatban tárgyalt modellt alkalmazzuk valós adatokon. Az anonimitás érdekében átskáláztam és transzformáltam az adatokat oly módon, hogy az eredeti káradatokat ne tükrözzék, azokat ne lehessen visszafejteni. A kárgyakoriság modellezéséhez főként Microsoft Excel lett használva. Az IBNER faktorok becslése Emblem nevű aktuáriusi szoftverrel, illetve R-ben történt. A kárnagyság eloszlás becslése pedig R-ben lett végrehajtva.

- Adataink lakossági vagyonszociális károkat tartalmaznak. A károk bekövetkezési és bejelentési időszakának a modellben a 2006.01.01-től 2009.12.31-ig tartó időszakot választottam, azonban a valós adatok kárdátumai eltérnek.
- A vizsgált időszak: 2006.01.01-2009.12.31.

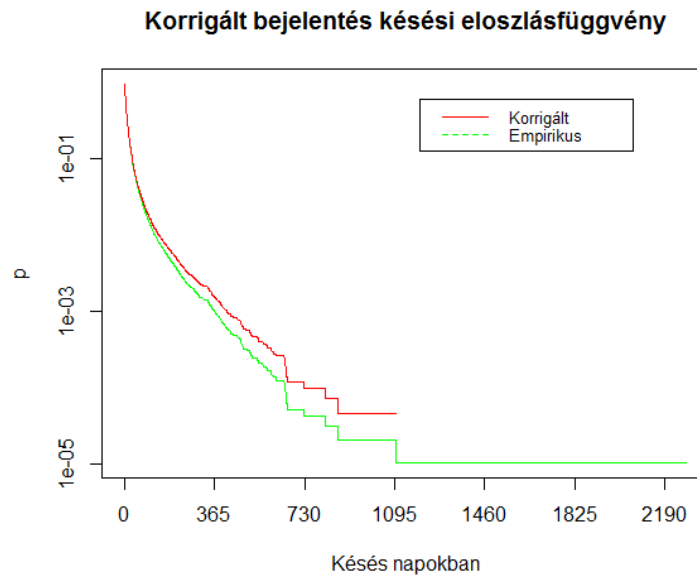
Mik a céljaink?

- RBNS: IBNER faktorok becslése és ezáltal RBNS károk meghatározása.
- Meghatározni az IBNR káreloszlást. Ide tartoznak a 2006.01.01 és 2009.12.31 között bekövetkezett, de nem bejelentett károk.
- Az RBNS és IBNER károk tartalékának becslése.

## Kárgyakorisági modell

Az első fejezet alapján először megkonstruáljuk a korrigált késési eloszlásfüggvényt. A bejelentett károk ki lettek igazítva az adott évi kitettséggel. Az empirikus, illetve az illesztéses módszer közül az előbbi jobb eredményeket adott. Az ágazatra jellemző a nagyon rövid kárbejelentési idő (pár nap), ezért nem nagy mértékben tolódott el

az alap eloszlásfüggvény. Mindösszesen 0,7 napot sikerült korrigálni a bejelentés késések várható értékén. Az különbség kimutatásához logaritmikus skálán ábrázoltam  $(1 - F(t))$  függvényt. A korrigált eloszlásfüggvény magasabban van, mert ha egy kár bekövetkezik 0-ban, akkor egy nagyobb a valószínűsége, hogy  $t$ -ig nem jelentik be, hiszen elnyújtottuk kicsit a bejelentések várható értékét.



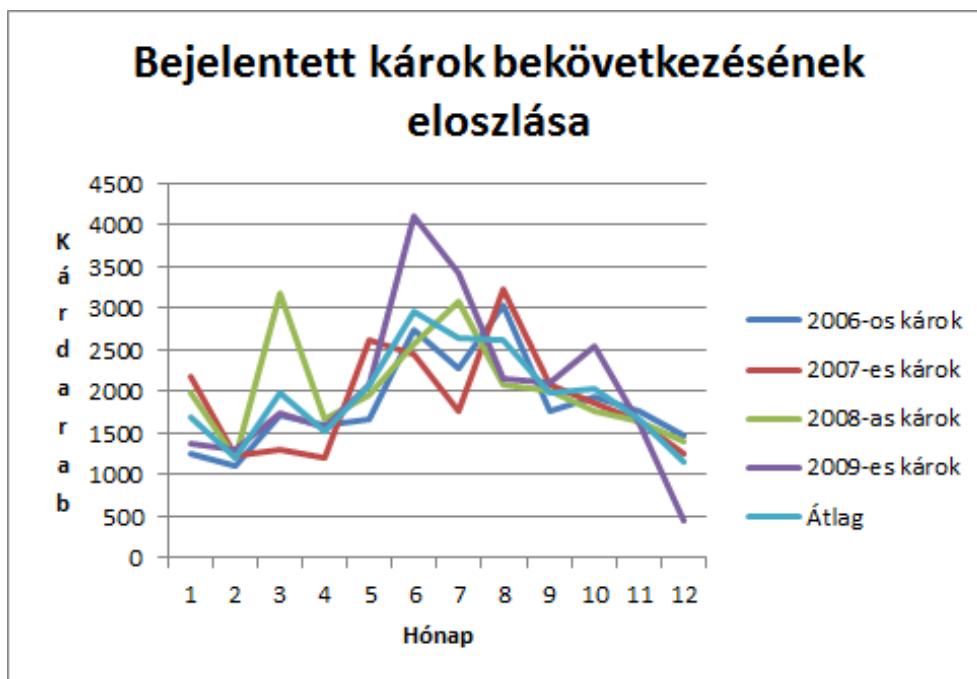
6.1. ábra. Az empirikus (zöld) és korrigált (piros) késési eloszlásfüggvény.

A bejelentett károknál csak a nem-zéró károkat szabad figyelembe venni, tehát azokat amelyeken történt kifizetés (vagy van rajta tartalék). Amennyiben az összes kárt tekintjük, akkor előzetesen meg kell vizsgálni, hogy a károk hány százalék nulázódik ki évenként és korrigálni kell az összes kárszámot.

Jelen esetben a károk körülbelül 90%-a lesz nem-zéró kár:

Év	Nem-zéró károk aránya
2006	89,84%
2007	90,68%
2008	89,86%
2009	90,2%

Itt az elemzés során csak azok a károk lettek felhasználva, ahol történt kifizetés. Többféle kárgyakorisági függvényt  $(\nu(t))$  megnéztem. Első körben a feltevésem az egyenletes kárgyakoriság volt. Ekkor szinte majdnem megegyező eredményt kap-



6.2. ábra. A bejelentett károk bekövetkezési kárgyakorisága napi bontásban különböző években

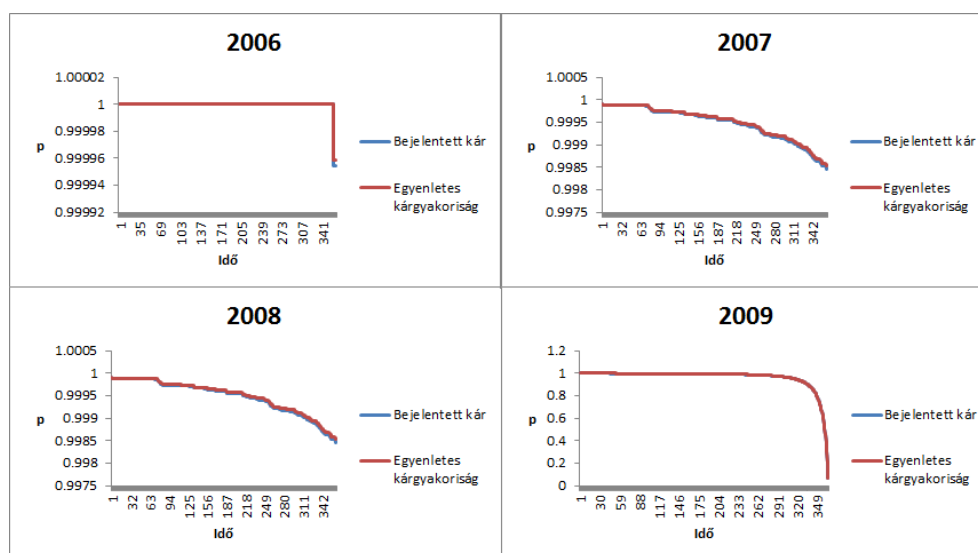
tam az IBNR károk várható értékre a lánclétra becsléshez képest. Amennyiben a kárgyakoriságot a bejelentett károk bekövetkezési időpontja szerinti kárgyakorisággént definiáljuk, akkor kevesebb IBNR kárt becsült, mint az egyenletes. A harmadik feltevés, hogy minden évben a kitéttiséggel kiigazított átlagos bejelentett károk bekövetkezési időpontja szerinti kárgyakorisága egyezik meg a kárgyakorisági függvénnyel. A különböző modellek hatása az IBNR kárszámra megtalálható a 6.1 táblázatban is. A variancia növekedett az IBNR kárszámok növekedésével együtt. Nyilvánvalóan akkor kapnánk a legnagyobb IBNR kárszámot, ha azt feltételeznénk, hogy a károk év elején történnek (extremális esetben az összes kár első nap). A 6.2 ábra alapján jól látható, hogy a bejelentett károk nem egyenletesen következtek be.

A különböző kárgyakorisági függvényeknek nem volt szignifikáns hatása a korrigált késési eloszlásfüggvényre (6.3 ábra). A piros vonalnál egyenletes kárgyakoriságot feltételeztünk, a kéknél pedig a bejelentett károk eloszlása alapján néztük a kárgyakoriságot.



6.1. táblázat. Különböző kárgyakorisági függvények hatása az IBNR darabszámra

Év	Lánclétra	Súlyozott átlag	Egyenletes	Bejelentett károk bekövetkezése szerint
2006	0	0	0	0
2007	2	1	1	1
2008	14	12	12	13
2009	915	742	949	636



6.3. ábra. A kárgyakoriság függvény választásának hatása a korrigált késési eloszlásfüggvényre ( $1 - F(t)$ )-ként ábrázolva.

A korrigált bejelentés késési eloszlásfüggvény és a bejelentett károk darabszáma alapján meghatároztam az ultimate és IBNR kárdarabszámokat. Az 6.2 táblázatban szerepelnek az eredmények. Az ultimate faktorok, tehát a 4. és 3. sorban szereplő kifejezések hányadosa. Jól láthatóan minél előrébb megyünk az időben, annál nagyobb az IBNR kárdarabszám (tehát bekövetkezési éveként a bejelentett károk aránya csökken az összes bekövetkezett kárhoz képest) és a bizonytalanság is növekszik.

6.2. táblázat. Ultimate és IBNR kárdarab becslés

1	Év	2006	2007	2008	2009
2	Bejelentett kárdarabszám	22348	22849	24555	24520
3	Várható kárgyakoriság (kitettséggel kiigazítva)	21844	22726	24629	25510
4	Várható bejelentett kárszám	21844	22724	24616	24514
5	Ultimate faktorok	100,00%	100,01%	100,05%	102,59%
6	Ultimate kárdarabszám	22348	22850	24568	25156
7	Ultimate kárdarabszám (kitettséggel kiigazítva)	23394	24050	25617	25716
8	IBNR kárdarabszám	0	1	13	636
9	Standard hiba	1	6	22	155

Az  $V \setminus M$  érték jóval meghaladja az 1-et, ezért negatív-binomiális eloszlást illeszttem az adatokra (jobb illeszkedést mutatott, mint a Poisson). Ha összevonjuk az éveket egy időszakra, akkor a 6.3 táblázatot kapjuk.

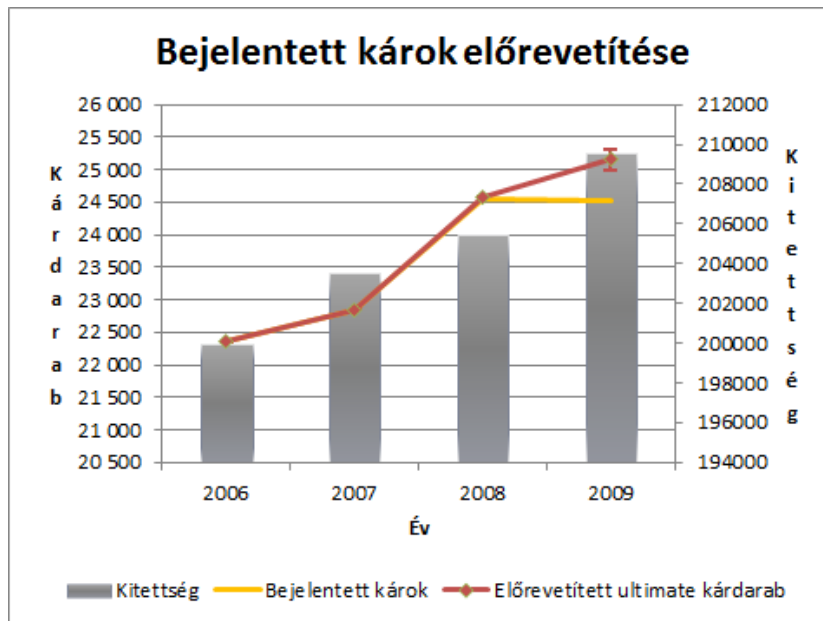
6.3. táblázat. Összevont kárdarab becslési táblázat

1	Év	2006-2009
2	Bejelentett kárdarabszám	94 272
3	Várható kárgyakoriság (kitettséggel kiigazítva)	94 348
4	Várható bejelentett kárszám	93 698
5	Ultimate faktorok	100,69%
6	Ultimate kárdarabszám	94926
7	Ultimate kárdarabszám (kitettséggel kiigazítva)	94926
8	IBNR kárdarabszám	654

Az eloszlásillesztés momentumok módszerével történt, mert a károk darabszáma miatt a maximum likelihood becslés nagyon lassú volt.

A 6.2 táblázat ábrázolva diagramon látható a 6.4 ábrán. A piros és sárga vonal közötti különbség adja meg az IBNR kárszámot. A hibavonalak pedig jelölik a becslés

szórását. Jól megfigyelhető az ultimate kárdarab növekedése a kitettség párhuzamos lineáris növekedésével.



6.4. ábra. Az ultimate kárdarabszám és a bejelentett károk ábrázolva a becslés standard hibájával.

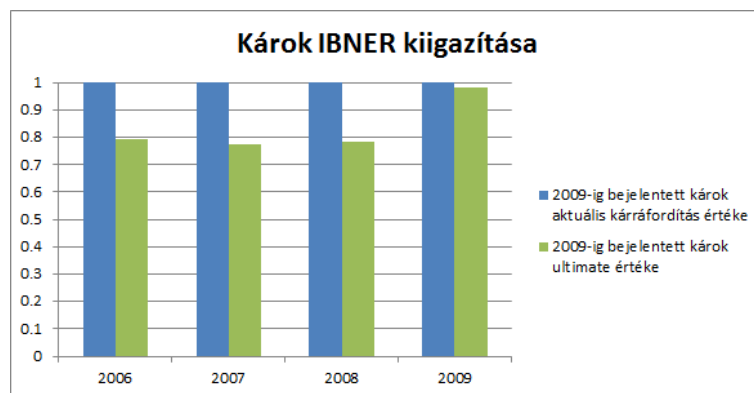
## Kárnagyság eloszlás

Ahogy korábban is megemlítettük, az IBNER faktorokat Általánosított Lineáris Modellel határozzuk meg. A modellben logaritmusos kapcsoló függvény lett választva, illetve Gamma eloszlású zaj. Ez a modell illeszkedett legjobban az adatokra.

Az elemzésben először három magyarázó változót tekintettem: kárnagyság, kárkifizési idő évben, illetve az függőkár hányadot (mennyi a még tételes tartalék az eddig kifizetett összeg és a tartalék összegének arányában).

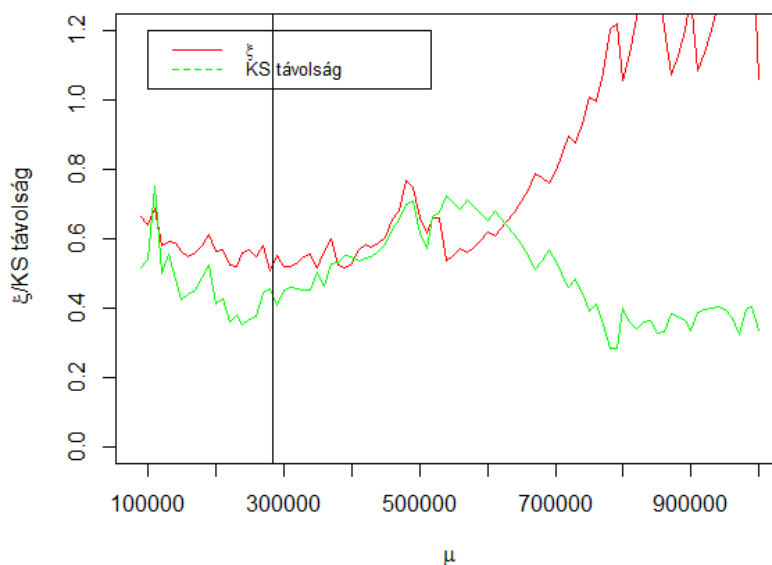
Lakossági vagyonszociális tartozó károk 99%-a első évben kifut, tehát lezárják a kárt, ezáltal a kárkifizési változónak nem volt erős magyarázó hatása. A kárkifizéssel párhuzamban, a függőkár százalék is első évben általában 100%, viszont második évre kifizetik a károk nagy részét. A 4 éves vizsgált kártörténeti periódus alatt nem volt olyan kár, amelyet nem rendeztek volna le maximum 3 év alatt.

Végül az illesztett modellben a magyarázó változók: a kárnagyság, tételes tartalék és a függőkár hányad. A 6.5 ábrán a kék oszlopok jelölik az IBNER nélküli kárátfordításokat, a zöld oszlopok pedig az IBNER utáni korrekciót. Jól láthatóan az IBNER modell azt mondja, hogy kevesebbet fogunk kifizetni mint az eddigi kifizetések és a tartalékok összege, továbbá a régebbi éveknél ez a hatás jelentősebb (2006-ra több mint 20%-ot csökkent, míg 2009-re csak néhány százalékot).



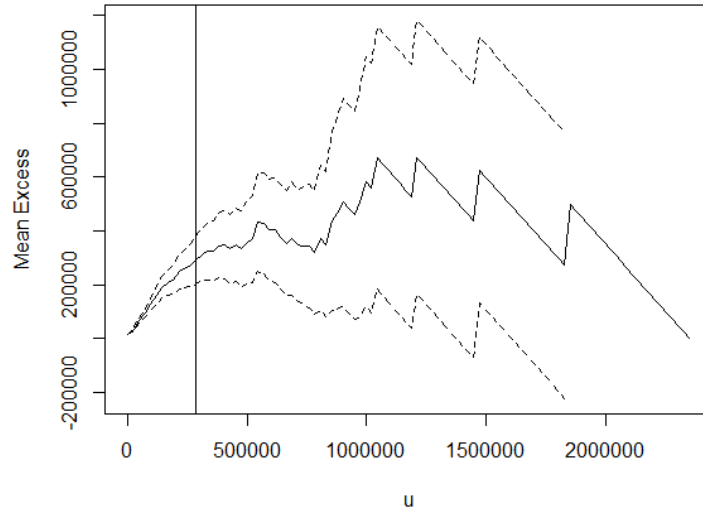
6.5. ábra. Az aktuális és IBNER korrekció utáni kárráfordítás

Az ágazat gyors kárkifutása miatt nincs szükség megkülönböztetni a kárnagyság eloszlást bejelentés éve szerint is, illetve feltételezzük, hogy a vizsgált periódus hossza miatt bekövetkezési évenként is azonos kárnagyság eloszlás áll fenn. Az IBNER faktorokkal kiigazított károkra egy empirikus/GPD modell lett illesztve, mivel a farokeloszlás előtti részre nem illeszkedett semmilyen megszokott eloszlás. A küszöbértéket többféle módszerrel is vizsgáltam. Az 6.6 ábrán jól látható, hogy a  $\xi$  értéke a 285000-nél behúzott vonalnál kezd el stabilizálódni, illetve a normalizált  $KS$  távolság is alacsony. Vizsgáltam további modelleket a kiszemelt küszöbérték 50000-es környezetében, azonban nem mutattak szignifikáns eltérést.



6.6. ábra. A Kolmogorov-Szmirnov távolságok és  $\xi$ -k különböző küszöbértékekre.

Ha megnézzük az empirikus várható küszöbfüggvényt:



6.7. ábra. Az empirikus várható küszöbérték függvény különböző küszöbértékekre.

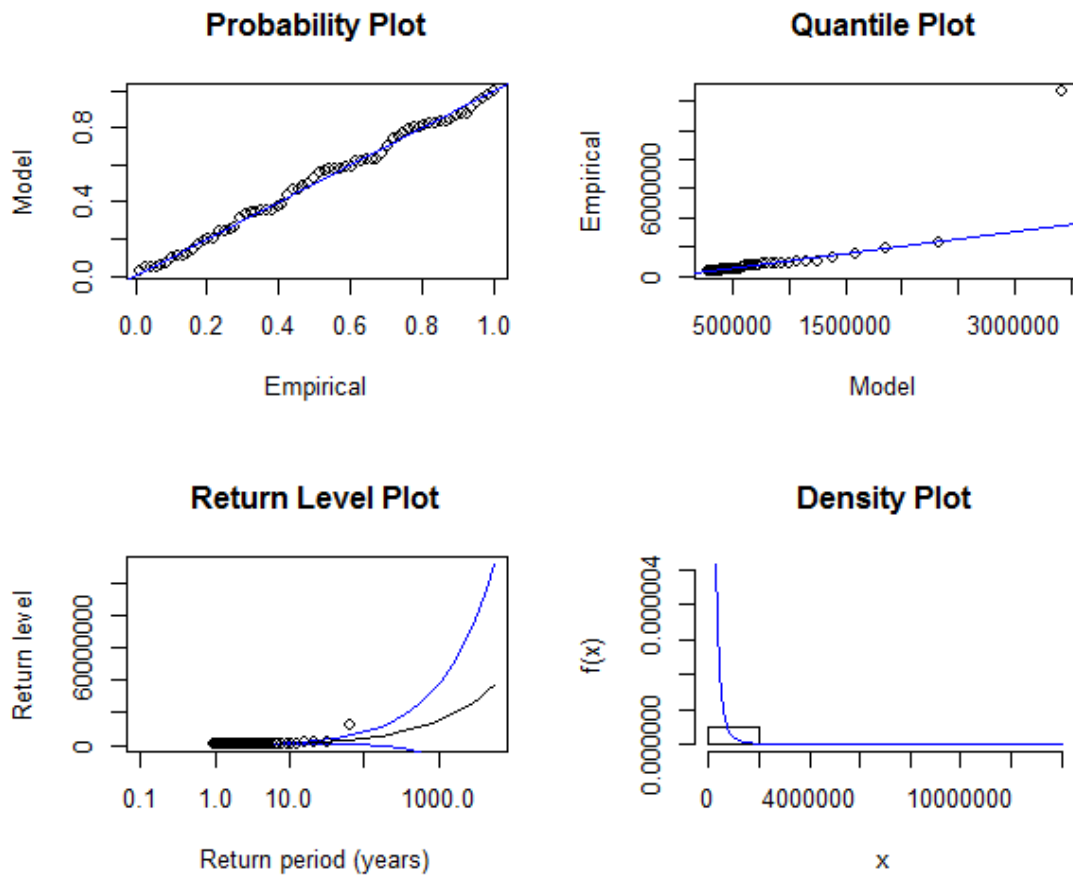
Jól láthatóan szintén körülbelül 285000-nél kezdődik a lineáris trend (a végén túl bizonytalanok a megfigyelések, az elején pedig túl alacsony a küszöbérték).

Ezek alapján az illesztett GPD eloszlás paraméterei:  $\mu = 285000$ ,  $\xi = 0.5362$  és  $\sigma = 193044.745$ .

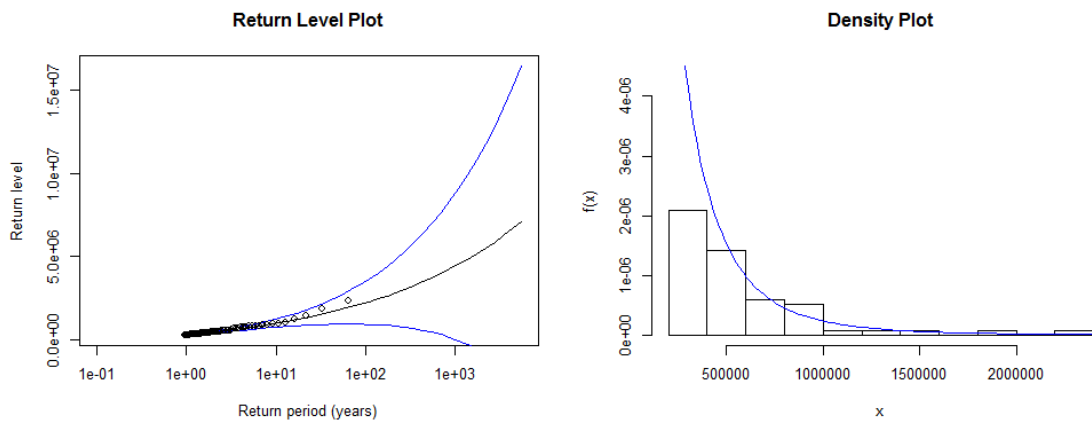
Nézzük meg a GPD modell illeszkedését.

A 6.8 ábrából látszódik, hogy az egyetlen extrémális érték eltorzítja a képet, de alapvetően a modellünk jól illeszkedik az adatsorra (ez látszik a első két ábrából, hiszen nagyjából a 45 fokos egyenesen helyezkednek el a pontok). Amennyiben kivesszük azt az egy értéket, akkor szebb ábrákat kapunk.

A bal alsó ábrán azt látjuk, hogy nagyjából milyen visszatérési idővel következnek be az adott megfigyelésünk a modell szerint és hogy benne van-e a késsel jelzett konfidenciaintervallumban. Az eloszlás kvantiliseit visszatérési szinteknek nevezzük. Legyen  $z$  a  $p$ -kvantilis. Ekkor  $z_p$  visszatérési ideje  $1 - p$  reciproka lesz. Például 99,5%-nál ez a 200 év. Így tehát  $z_p$  kvantilis lesz azaz érték, amelyet átlagosan  $1/(1 - p)$  évente haladunk meg. A kilógó értéknek kicsit kevesebb, mint 100 év a visszatérési ideje. A 6.9 ábrán látható hogy mi a különbség az ábrák között, ha kivesszük az extrémális értéket.



6.8. ábra. Modellilleszkedés vizsgálata grafikai eszközökkel. A jobb alsó ábrán pedig a becsült sűrűségfüggvény



6.9. ábra. Az extrémális érték kivétele után korrigálódik az ábra

A szimulációt lefuttattam végrehajtottam külön évekre, illetve összevontan is. Percentilisekkel felírva az alábbi táblázatot kaptam:

6.4. táblázat. Összkár percentilisek és kvantilisek éves bontásban

Percentilis/Év	2006	2007	2008	2009	2006-2009
10%	0	0	0	6 935 898	7 179 822
20%	0	0	0	7 902 606	8 169 033
30%	0	0	3 132	8 649 849	8 951 408
40%	0	0	19 277	9 358 571	9 652 649
50%	0	0	47 541	10 070 528	10 361 222
60%	0	0	97 543	10 795 494	11 103 858
70%	0	0	174 452	11 601 668	11 957 228
80%	0	0	320 149	12 593 265	13 015 152
90%	0	11 760	624 689	14 206 426	14 571 710
95%	0	84 324	993 104	15 611 613	16 170 973
98%	0	310 027	1 513 314	17 429 240	18 230 597
99%	0	604 108	1 967 737	18 938 142	19 852 577
99,5%	0	933 627	2 455 536	20 929 716	22 044 300
99,9%	0	1 923 029	3 963 652	30 847 010	39 721 112

6.5. táblázat. Statisztikai adatok a scenáriókról

Év	2006	2007	2008	2009
Átlag	282	23 065	214 515	10 418 577
Minimum	0	0	0	3 339 521
Maximum	951 426	4 020 632	5 677 070	164 216 902

Az európai biztosítókra 2016.január 1.-től hatályba lépett Szolvencia 2 irányelv szerint a biztosítóknak úgy kell megképezni a szavatoló tőkéjüket (szavatoló tőkeszükséglet, SCR), hogy az elkövetkezendő 1 évre 99,5%-os valószínűséggel fedezni tudják kötelezettségeiket. Ez pont azt jelenti, hogy a biztosító csődje legfeljebb 0,05%-os, ami a 200 éves eseménynek felel meg 1 éves időtávra. Továbbá a minimális szavatoló tőkeszükséglet nagyságát is meghatározza (ami alá ha beesik a biztosító, akkor a felügyelet közbelép), amely a 85%-os percentilisnek felel meg és az SCR 25 és 45 százaléka közé kell hogy essen.

A 6.4 táblázat szerint, ha 99,5%-os biztonság mellett szeretnénk tartalékolni a 2006 és 2009 bekövetkezett, de nem bejelentett károkra, akkor nagyjából 24-25 millió

Ft körüli összeget kellene tartalékba helyeznünk. A scenáriók átlaga egy pontbecslést ad a tartalékra. A 6.5 táblázatból látható, hogy szintén a 2006 és 2009-es IBNR károokra 10,6 milliós tartalékot kellene képeznünk. A különbség a kettő között abból adódik, hogy a biztosítók best-estimate módszerrel nem sztochasztikusan tartalékolnak, így tehát a tartalékot úgy képzik meg, hogy várható értékben fedezni tudják a kifizetéseket. Ahogy haladunk előre az években, annál kisebb percentilisnek felel meg ez a tartalék, ami logikus hiszen egyre jobban nő a bizonytalanság. Az első évben szinte tökéletesen fedve vagyunk, második évben a 90-95 százalék között, harmadik évben már 70-80 százalék között és végül az utolsó évben már csak az 50-60 százalékos szintet érjük el.

Ahogy már korábban is említettem, mivel az RBNS modell determinisztikus, a kombinált (IBNR + RBNS) káreloszlás olyan értelemben egyszerűsödik, hogy az IBNR eloszláshoz csak hozzá kell adni az RBNS tartalékot.

Ahhoz hogy összetudjam hasonlítani a megvalósult adatokkal, először ezeket az adatokat is ugyanolyan módon transzformálnom kellett, mint a bemeneti adatokat. Ezek után vettem az IBNR kárnagyság scenáriók átlagát és összegeztem az RBNS tartalékkal. Itt feltételezzük, hogy a már nem lesz változás a kárráfordításban, ami jogos feltevés, mivel ahogy korábban is említettem a lakossági vagyonszámításnál előforduló károk hamar kifutnak. Az így kapott eredmény:

6.6. táblázat. A becsült tartalék és a megvalósult adatok közötti különbség százalékosan

<b>Év</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>
<b>Eltérés</b>	0.63%	0.05%	0.86%	0.55%

A 6.6 táblázatból az látható, hogy elég pontosan sikerült megközelíteni a vizsgált években történt ultimate kárkifizetést.

Összegezve, jól láthatóan kellően bonyolultabb egy ilyen modell felépítése, mint a háromszöges módszerek. Azonban megfelelő paraméter kalibrálással, eloszlásillesztésekkel, illetve szakmai tapasztalatok által alátámasztott döntésekkel egy viszonylag erős és pontos modellt lehet felépíteni a tartalékok becslésére (a háromszöges modellezéssel párhuzamosan). Továbbá a jobb eredmény érdekében, több szempontból is lehet bonyolítani a modellt. Hiszen sok esetben a különböző éveknek más a bejelentés késési eloszlásfüggvényünk vagy vannak olyan ágazatok, amelyeknek teljesen más a káreloszláslása is bekövetkezési évenként. Éles környezetben célszerű lenne felbontani a károkat nagy és átlagos károokra is. Ha a sztochasztikus modellezés nem célunk, attól függetlenül a modell másik nagy előnye, hogy nem csak portfólió szinten



tudunk tartalékot mondani, hanem káronként, így díjazás szempontjából is hasznos lehet a modell.

# Függelék

Az első programkód a küszöbérték választására vonatkozik. A második pedig a szimulációra.

---

```
#Káradatok beolvasása
library("MASS")
data=read.csv("karok.csv",header=FALSE,sep=";",dec=",")

#Beolvasott károk mátrix alakja

madat=as.matrix(data,nc=1)

#Kezdeti küszöbértékek definiálása 10000-es léptékkal a legnagyobb kárig

mu<-seq(0,max(madat),10000)

#Empirikus eloszlásfüggvény definiálása
emp_dist<-function(x){
  emp_data1<-sort(emp_data[emp_data<=x])
  p<-length(emp_data1)/length(emp_data)
  return(p)
}

#GPD paramétereinek és a KS távolságok vektorának definiálása

ks<-numeric(length(mu))
xi<-numeric(length(mu))
sig<-numeric(length(mu))

#Kolmogorov-Szmirnov távolságok és xi-k kiszámolása különböző
küszöbértékekre
for(i in (1:length(mu))){
```

```

emp_data<-sort(madat[madat>mu[i]])

empiric<-numeric(length=length(emp_data))

gpd<-numeric(length(emp_data))

f<-gpd.fit(emp_data,threshold=mu[i])

xi[i]<-f$mle[2]

sig[i]<-f$mle[1]

for(j in (1:length(emp_data))){

gpd[j]<-1-(1+xi[i]*(emp_data[j]-mu[i])/sig[i])^(-1/xi[i])

empiric[j]<-emp_dist(emp_data[j])
}
ks[i]<-sqrt(length(emp_data))*max(abs(empiric-gpd))
}
mrl.plot(madat)
#####

#NB eloszlás generáláshoz használt programcsomag
library("MASS")

#Szimulációk száma
N<-10000

#Évek száma
Y<-4

#VM érték
VM<-38

#Bejelentett és Ultimate kárdarabszám beolvasása különböző évekre

data_ibnr<-read.csv("reported.csv" , sep=";" , header=FALSE , dec=",")

```

```

IBNR<-as.matrix(data_ibnr,nc=2)

#Kárgyakorisági szimulációs mátrix létrehozása
frequency<-matrix(nrow=Y,ncol=N)

#Kárgyakoriságok szimulálása
for (i in 1:Y){

    #ultimate - reported
    mu<-IBNR[i,2]-IBNR[i,1]

    theta<-mu/(VM-1)

    frequency[i,]<-rnegbin(N,mu,theta)

}

#Károk beolvasása

threshold<-280000

gpd_model<-gpd.fit(madat,threshold)

#GPD modell vizsgálata
gpd.diag(gpd_model)

#GPD paraméterek
xi<-gpd_model$mle[2]
sig<-gpd_model$mle[1]

#Empirikus eloszlásfüggvény
ecdf<-function(x){

    madat1<-madat[madat<=x]

    p<-length(madat1)/length(madat)

    return(p)
}

```

```

omadat<-sort(madat)

#Kárnagyság eloszlásfüggvény

loss_dist<-function(x,threshold,xi,sig){

  if(x<=threshold) {result<-ecdf(x)}

  if(x>threshold) {result<-(ecdf(threshold))+
  (1-ecdf(threshold))*(1-(1+xi*(x-threshold)/sig)^(-1/xi))}

return(result)
}

#Inverz kárnagyság eloszlásfüggvény
inverse_ld<-function(x){

  if(x<ecdf(omadat[1])) {return(0)}

  if (x==1) {return(max(omadat))}

  if(x<loss_dist(threshold,threshold,xi,sig)){
    index<-length(omadat[omadat<=threshold])*
    x/loss_dist(threshold,threshold,xi,sig)
    return(omadat[omadat<=threshold][index])
  }
else {
return((((x-1)/(ecdf(threshold)-1))^(-xi)-1)*sig/xi+threshold)
}
}

loss<-matrix(nrow=Y,ncol=N)

#Kárösszegek szimulálása

for(j in (1:Y)){
  for(i in (1:N)){
    if(freq[j,i]==0){loss[j,i]<-0}
  }
}

```

```

        else
            {
                loss[j,i]<-sum(sapply(runif(frequency[j,i],0,1),inverse_1d))
            }
        }
    }

sort_loss<-matrix(nrow=Y,ncol=N)

for(i in 1:Y){
    sort_loss[i,]<-sort(loss[i,])
}

#Aggregált káreloszlás(kárnagyság és év függvényében)
agg_loss_dist<-function(x,y){

    return(length(sort_loss[y,sort_loss[y,]<=x])/length(sort_loss[y,]))
}

#Percentilis táblázat létrehozása
percentiles<-matrix(nrow=10,ncol=Y+1)
percentiles[1,1]<-0.1
percentiles[1,1]<-0.5
percentiles[2,1]<-0.6
percentiles[3,1]<-0.7
percentiles[4,1]<-0.8
percentiles[5,1]<-0.9
percentiles[6,1]<-0.95
percentiles[7,1]<-0.98
percentiles[8,1]<-0.99
percentiles[9,1]<-0.995
percentiles[10,1]<-0.999

for (j in 1:Y){
    for (i in 1:10){
        percentiles[i,j+1]<-
            sort_loss[j,percentiles[i,1]*length(sort_loss[j,])]
    }
}

```

---

# Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Pónuzs Róbertnek, aki szakmai tudásával, hasznos tanácsaival és észrevételeivel segített a dolgozat megírásában. Szeretném megköszönni Szüleimnek, akiktől sok támogatást és biztatást kaptam a tanulmányaimhoz. Köszönettel tartozom Nagy Zoltánnak, aki felhívta figyelmemet az esetleges hibákra. Végül szeretném megköszönni Pietro Parodinak, aki levelezésen keresztül segítőkészen válaszolt kérdéseimre a témával kapcsolatban.

# Irodalomjegyzék

- [1] Pietro Parodi, *Triangle-free reserving: a non-traditional framework for estimating reserves and reserve uncertainty*, British Actuarial Journal, 19, pp 219-233. doi:10.1017/S1357321713000354, London, 2013
- [2] Arató Miklós, *Nem-életbiztosítási matematika*, ELTE Eötvös Kiadó, 2001
- [3] Pietro Parodi, *Pricing in General Insurance*, CRC Press, 2014
- [4] Annette J. Dobson, *An Introduction to Generalized Linear Models*, Springer, 1990
- [5] Tóth András, *Általánosított lineáris modellek a biztosításban*, ELTE Diplomamunka, 2013
- [6] P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events*, Springer-Verlag, 1997
- [7] Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001
- [8] S. Haastруп and E. Arjas, *Claims reserving in continuous time: a nonparametric Bayesian approach*, ASTIN Bulletin, 26(2):139-164, 1996
- [9] Jesús Gonzalo-José Olmo, *Which Extreme Values are Really Extremes?*, Montreal, 2003
- [10] Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot, *Loss Models: From Data to Decisions*, Third Edition, Wiley, Hoboken, NJ.
- [11] Norberg Ragnar, *Prediction of outstanding liabilities in non-life insurance*, ASTIN BULLETIN, Vol. 23, No. 1, 1993