

EÖTVÖS LORÁND TUDOMÁNYEGYETEM  
TERMÉSZETTUDOMÁNYI KAR  
BUDAPESTI CORVINUS EGYETEM  
KÖZGAZDASÁGTUDOMÁNYI KAR

---

# ÖNRÉSZMODELLEK VIZSGÁLATA

— Szakdolgozat —

Témavezető:

ARATÓ MIKLÓS

Valószínűségelméleti és Statisztika  
Tanszék

Készítette:

MARXER ZALÁN BENEDEK

Biztosítási és pénzügyi matematika  
MSc.  
Aktuárius szakirány



Budapest, 2022

# NYILATKOZAT

**Név:** Marxer Zalán Benedek

**ELTE Természettudományi Kar, szak:** Biztosítási és pénzügyi matematika MSc

**NEPTUN azonosító:** LMP7N2

**Diplomamunka címe:**  
Önrészmodellek vizsgálata

A **diplomamunka** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022. 05. 27.

*Marxer Zalán Benedek*

*a hallgató aláírása*

# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>2</b>
<b>2. Önrészmodellek</b>	<b>4</b>
2.1. Önrészes biztosítások fajtái . . . . .	4
2.2. Elméleti alapok . . . . .	9
2.3. Önrész "mérése" . . . . .	15
2.4. Az önrész hatása a becslésre . . . . .	18
<b>3. Modellek</b>	<b>26</b>
3.1. Az adatok bemutatása . . . . .	26
3.2. Önrész nélkül . . . . .	28
3.3. Önrészmodellek . . . . .	31
3.4. Eloszlások illeszkedése . . . . .	34
<b>4. Összefoglalás</b>	<b>38</b>

# 1. fejezet

## Bevezetés

A biztosításban az önrész alkalmazása egy régi, jól bevált módszer. Számos termékkel találkozhatunk, amik az önrészek különböző fajtáit tartalmazzák. Ez a módszer alapvetően mindkét fél számára számos előnnyel jár. A biztosító esetében csökken a kockázat, csökken a csalás valószínűsége, hiszen a szerződőre is hárul a felmerülő kárból. A szerződő esetében pedig alacsonyabb díjról beszélhetünk. Az önrész bevezetésével csökken a kockázat, ami a díjban is megmutatkozik.

Habár az önrész megjelenése a biztosításban nem ma történt, a kárszámításra, kármodellekre gyakorolt hatása nem a legkutatottabb terület. Természetesen így is született számos érdekes tanulmány. Douglas G. Chapman [1] például már 1956-ban vizsgálta a csonkított gamma eloszlást, és a paraméterbecslésre gyakorolt hatást. A dolgozatomban az önrészes adatok esetében szintén csonkított mintavételezést alkalmaztam.

Munkám során külön vizsgáltam az önrész kárgyakoriságra és kárnagyságra gyakorolt hatását. Klugman, Panjer és Willmot könyvében [3] külön fejezet van az önrésznek szentelve, amiben egyszerűbb példákkal is szemléltetik a változásokat.

A dolgozat elején bemutatom az alapvető kárnagyság eloszlásokat, majd különböző önrész típusok esetében a megváltozott eloszlás- és sűrűségfüggvényeket. Utána bemutatom az összetett kármodellt, és hozzá kapcsolódóan bemutatok pár mutatószámot, amivel az önrész hatását lehet vizsgálni. Az

elméleti fejezet végén ismertetem a pszeudo maximum likelihood (rövidítve: pszeudo-MLE) [2] becslést, ami az önrészes biztosítások esetén egy hasznos eszköz.

Az elméleti bevezető után általam generált adatokon és valós adatokon összehasonlítom a különböző kárszám- és kárnagyság modellek eredményeit. Végül egyfajta kitekintésképp megvizsgálom, hogy az általam generált adatok esetében - ahol ismert a háttéreloszlás - mi történik "félremodellezés" esetén.

## 2. fejezet

# Önrészmodellek

Az önrész a biztosítási szerződés fontos eleme. Hatással van nemcsak a károk számára és nagyságára, de a szerződő viselkedését is tükrözheti. Az ilyen típusú biztosításokhoz kapcsolódó megfigyeléseink cenzorálásnak és csonkításnak lehetnek kitéve. Ezt mindenképpen figyelembe kell venni, amikor egy ilyen termék árát szeretnénk meghatározni. Az alábbi fejezet Gee Y. Lee [4] cikkén alapul és bemutatja az önrészes biztosításokhoz kapcsolódó alapfogalmakat.

### 2.1. Önrészes biztosítások fajtái

Amikor önrészes biztosításról beszélünk, akkor többféle konstrukcióval is találkozhatunk. A továbbiakban a levonásos önrésszel fogunk alaposabban foglalkozni, de a teljesség igénye miatt röviden összefoglalom a piacon is fellelhető típusokat.

Levonásos önrész esetében adott  $d$  konstans és  $Y$  kárnagyság. A biztosító akkor fizet, ha a kár meghaladja az önrész ( $d$ ) nagyságát, és ekkor a biztosító a kár és önrész különbségét téríti meg. A biztosító szemszögéből a következőképpen írhatjuk fel:

$$g(Y; d) = \begin{cases} 0 & , Y \leq d \\ Y - d & , d < Y < \infty \end{cases}$$

Mivel a levonásos önrész az egyik legelterjedtebb típus, ezért vizsgáljuk meg kicsit alaposabban is. Nézzük meg, hogy milyen alakot ölt az eloszlásfüggvény ebben az esetben és a leggyakoribb eloszlás-típusok esetében. Természetesen mivel biztosításról beszélünk, így az eloszlás- és sűrűségfüggvények mind a nemnegatív félegyenesen értelmezettek. A levezetésben kihasználtuk, hogy a feltételezéseink szerint a káreloszlások folytonosak.

$$F_{Y_*}(x) = \mathbb{P}(Y - d < x | Y > d) = \frac{\mathbb{P}(d < Y < x + d)}{\mathbb{P}(Y > d)} = \frac{F_Y(x + d) - F_Y(d)}{1 - F_Y(d)}$$

Exponenciális-eloszlás:  $Y \sim \text{Exp}(\lambda)$

$$F_{Y_*}(x) = \frac{(1 - e^{-\lambda(x+d)}) - (1 - e^{-\lambda d})}{1 - (1 - e^{-\lambda d})} = \frac{e^{-\lambda d} - e^{-\lambda(x+d)}}{e^{-\lambda d}} = 1 - e^{-\lambda x}$$

$$f_{Y_*}(x) = \lambda e^{-\lambda x}$$

Az exponenciális-eloszlás esetében az új eloszlás is exponenciális ugyanazzal a paraméterrel. Ez az eloszlás "örökifjú" tulajdonságának köszönhető.

Gamma-eloszlás:  $Y \sim \Gamma(\alpha, \beta)$

$$F_{Y_*}(x) = \frac{\frac{1}{\Gamma(\alpha)} \cdot \gamma(\alpha, \beta(x+d)) - \frac{1}{\Gamma(\alpha)} \cdot \gamma(\alpha, \beta d)}{\frac{1}{\Gamma(\alpha)} \cdot \gamma(\alpha, \beta d)} = \frac{\gamma(\alpha, \beta(x+d)) - \gamma(\alpha, \beta d)}{\Gamma(\alpha) - \gamma(\alpha, \beta d)} = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha) - \gamma(\alpha, \beta d)}$$

$$f_{Y_*}(x) = \frac{e^{-\beta x} \beta^\alpha x^{\alpha-1}}{\Gamma(\alpha) - \gamma(\alpha, \beta d)}$$

Lognormális-eloszlás:  $\log(Y) \sim \mathcal{N}(\mu, \sigma^2)$

$$F_{Y_*}(x) = \frac{\Phi\left(\frac{\log(x+d)-\mu}{\sigma}\right) - \Phi\left(\frac{\log(d)-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(d)-\mu}{\sigma}\right)}$$

$$f_{Y_*} = \frac{1}{(1 - \Phi\left(\frac{\log(d)-\mu}{\sigma}\right))(x+d)\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\log(x+d)-\mu)^2}{2\sigma^2}}$$

Pareto-eloszlás:  $Y \sim Par(\alpha, \beta)$

$$F_{Y_*} = \frac{1 - (\frac{\beta}{\beta+x+d})^\alpha - (1 - (\frac{\beta}{\beta+d})^\alpha)}{1 - (1 - (\frac{\beta}{\beta+d})^\alpha)} = \frac{(\frac{\beta}{\beta+d})^\alpha - (\frac{\beta}{\beta+x+d})^\alpha}{(\frac{\beta}{\beta+d})^\alpha} = 1 - (\frac{\beta+d}{\beta+d+x})^\alpha$$

$$f_{Y_*}(x) = \frac{\alpha(\frac{\beta+d}{\beta+d+x})^\alpha}{\beta+d+x}$$

A Pareto-eloszlás esetében  $Y_*$  eloszlása is ugyanúgy Pareto, csak a  $\beta$  paraméterhez hozzáadódik az önrész.

Weibull-eloszlás:  $Y \sim Weibull(\alpha, \lambda)$

$$F_{Y_*} = \frac{(1 - e^{-\lambda(x+d)^\alpha}) - (1 - e^{-\lambda d^\alpha})}{1 - (1 - e^{-\lambda d^\alpha})} = \frac{e^{-\lambda d^\alpha} - e^{-\lambda(x+d)^\alpha}}{e^{-\lambda d^\alpha}} = 1 - e^{-\lambda((x+d)^\alpha - d^\alpha)}$$

$$f_{Y_*}(x) = \alpha\lambda(d+x)^{\alpha-1}e^{-\lambda((x+d)^\alpha - d^\alpha)}$$

A Weibull-eloszlás esetében is az eredetihez nagyon hasonló eloszlás- és sűrűségfüggvényeket kapunk.

Százalékos önrésről beszélünk, amikor a kár egy bizonyos százalékát téríti meg a biztosító az ügyfél részére, másképpen megfogalmazva a kár egy adott  $p$  százaléka a szerződőre hárul. Ekkor a biztosító részére a következő függvény adja meg a kifizetést:

$$g(Y; p) = (1 - p) \cdot Y$$

A fent említett két önrésztípus kombinációja talán a leggyakoribb ma a piacon. Ha megvizsgáljuk a legnagyobb hazai biztosítók kínálatát például CASCO termékből, akkor a legtöbb esetben egy ilyen kevert önrésszel találkozunk. Vagyis az önrész értéke alapvetően a kár egy adott  $p$  százaléka, de van egy minimum  $d$  érték, ami mindenképpen a szerződőre hárul. Képlettel megfogalmazva:

$$g(Y; p; d) = \begin{cases} 0 & , Y \leq d \\ \min((1 - p) \cdot Y; Y - d) & , d < Y < \infty \end{cases}$$



A kombinált önrész esetében is írjuk fel az eloszlásfüggvényt és nézzük meg a nevezetes eloszlástípusokat is.

$$F_{Y_*}(x) = \mathbb{P}(\min((1-p)Y; Y-d) < x | Y > d) = \quad (2.1)$$

$$= \frac{\mathbb{P}(\min((1-p)Y; Y-d) < x, Y > d)}{\mathbb{P}(Y > d)} = \quad (2.2)$$

$$= \frac{\mathbb{P}(Y > d) - \mathbb{P}(\min((1-p)Y; Y-d) \geq x)}{\mathbb{P}(Y > d)} = \quad (2.3)$$

$$= \begin{cases} \frac{F_Y(x+d) - F_Y(d)}{1 - F_Y(d)} & , 0 < x \leq d \frac{1-p}{p} \\ \frac{F_Y(\frac{x}{1-p}) - F_Y(\frac{d}{p})}{1 - F_Y(d)} & , d \frac{1-p}{p} < x \end{cases} \quad (2.4)$$

Hasonlóan a levonásos önrészhez, itt is nézzük meg a gyakoribb eloszlástípusokat. A levonásos önrész nagysága legyen  $d$ , míg a százalékos önrész esetében legyen  $p$  a szerződőre eső rész.

Exponenciális-eloszlás:  $Y \sim Exp(\lambda)$

$$F_{Y_*}(x) = \begin{cases} 1 - e^{-\lambda x} & , 0 < x \leq d \frac{1-p}{p} \\ 1 - e^{-\lambda(\frac{x}{1-p} - \frac{d}{p})} & , d \frac{1-p}{p} < x \end{cases}$$

$$f_{Y_*}(x) = \begin{cases} \lambda e^{-\lambda x} & , 0 < x \leq d \frac{1-p}{p} \\ \frac{\lambda}{1-p} e^{-\lambda(\frac{x}{1-p} - \frac{d}{p})} & , d \frac{1-p}{p} < x \end{cases}$$

Az exponenciális-eloszlás esetében a levonásos önréssel ellentétben itt már nem beszélhetünk exponenciális eloszlásról.

Gamma-eloszlás:  $Y \sim \Gamma(\alpha, \beta)$

$$F_{Y_*}(x) = \begin{cases} \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha) - \gamma(\alpha, \beta d)} & , 0 < x \leq d \frac{1-p}{p} \\ \frac{\gamma(\alpha, \beta(\frac{x}{1-p} - \frac{d}{p}))}{\Gamma(\alpha) - \gamma(\alpha, \beta d)} & , d \frac{1-p}{p} < x \end{cases}$$

$$f_{Y_*}(x) = \begin{cases} \frac{e^{-\beta x} \beta^\alpha x^{\alpha-1}}{\Gamma(\alpha) - \gamma(\alpha, \beta d)} & , 0 < x \leq d \frac{1-p}{p} \\ \frac{e^{-\beta(\frac{x}{1-p} - \frac{d}{p})} (\beta(\frac{x}{1-p} - \frac{d}{p}))^\alpha}{\frac{d(p-1)+x}{\Gamma(\alpha) - \gamma(\alpha, \beta d)}} & , d \frac{1-p}{p} < x \end{cases}$$

Lognormális-eloszlás:  $\log(Y) \sim \mathcal{N}(\mu, \sigma^2)$

$$F_{Y_*}(x) = \begin{cases} \frac{\Phi(\frac{\log(x+d)-\mu}{\sigma}) - \Phi(\frac{\log(d)-\mu}{\sigma})}{1 - \Phi(\frac{\log(d)-\mu}{\sigma})} & , 0 < x \leq d^{\frac{1-p}{p}} \\ \frac{\Phi(\frac{\log(\frac{x}{1-p})-\mu}{\sigma}) - \Phi(\frac{\log(\frac{d}{p})-\mu}{\sigma})}{1 - \Phi(\frac{\log(d)-\mu}{\sigma})} & , d^{\frac{1-p}{p}} < x \end{cases}$$

$$f_{Y_*} = \begin{cases} \frac{1}{(1 - \Phi(\frac{\log(d)-\mu}{\sigma})) (x+d) \sigma \sqrt{2\pi}} \cdot e^{-\frac{(\log(x+d)-\mu)^2}{2\sigma^2}} & , 0 < x \leq d^{\frac{1-p}{p}} \\ \frac{1}{(1 - \Phi(\frac{\log(d)-\mu}{\sigma})) x \sigma \sqrt{2\pi}} \cdot e^{-\frac{(\log(\frac{x}{1-p})-\mu)^2}{2\sigma^2}} & , d^{\frac{1-p}{p}} < x \end{cases}$$

Pareto-eloszlás:  $Y \sim Par(\alpha, \beta)$

$$F_{Y_*} = \begin{cases} 1 - (\frac{\beta+d}{\beta+d+x})^\alpha & , 0 < x \leq d^{\frac{1-p}{p}} \\ 1 - (\frac{\beta+\frac{d}{p}}{\beta+\frac{x}{1-p}})^\alpha & , d^{\frac{1-p}{p}} < x \end{cases}$$

$$f_{Y_*}(x) = \begin{cases} \frac{\alpha(\frac{\beta+d}{\beta+d+x})^\alpha}{\beta+d+x} & , 0 < x \leq d^{\frac{1-p}{p}} \\ \frac{\alpha(\frac{(p-1)(\beta+\frac{d}{p})}{\beta(p-1)-x})^\alpha}{\beta(1-p)+x} & , d^{\frac{1-p}{p}} < x \end{cases}$$

Weibull-eloszlás:  $Y \sim Weibull(\alpha, \lambda)$

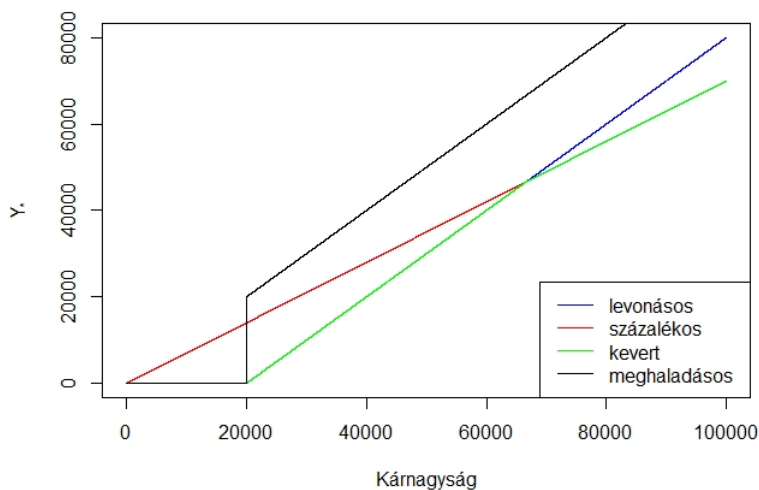
$$F_{Y_*} = \begin{cases} 1 - e^{-\lambda((x+d)^\alpha - d^\alpha)} & , 0 < x \leq d^{\frac{1-p}{p}} \\ 1 - e^{-\lambda((\frac{x}{1-p})^\alpha - (\frac{d}{p})^\alpha)} & , d^{\frac{1-p}{p}} < x \end{cases}$$

$$f_{Y_*}(x) = \begin{cases} \alpha\lambda(d+x)^{\alpha-1} e^{-\lambda((x+d)^\alpha - d^\alpha)} & , 0 < x \leq d^{\frac{1-p}{p}} \\ \frac{\alpha\lambda(\frac{x}{1-p})^{\alpha-1} e^{-\lambda((\frac{x}{1-p})^\alpha - (\frac{d}{p})^\alpha)}}{1-p} & , d^{\frac{1-p}{p}} < x \end{cases}$$

A fentiekén túl beszélhetünk még meghaladásos önrészről is. Ez nem túl elterjedt típusa az önrészeknek és kizárólag akkor élnek vele a biztosítók, ha a csalás lehetősége teljesen kizárható. A meghaladásos önrész lényege, hogy a biztosító akkor fizet, ha a kár meghaladja az önrész értékét, és ekkor a teljes kár összegét megtéríti. A következő alakban írhatjuk fel:

$$g(Y; d) = \begin{cases} 0 & , Y \leq d \\ Y & , d < Y < \infty \end{cases}$$

A négy típus viselkedését a kárnagyság függvényében egy ábra illusztrálja. A fenti paramétereket  $d = 20000$ , illetve  $p = 0.3$ -nak választottam. Jól kivehető, hogy a kevert önrész a levonásos és százalékos önrészből adódik. A meghaladásos önrész okozza a biztosítónak a nagyobb kifizetéseket.



2.1. ábra. Különböző önrésztípusok a kárnagyság függvényében

## 2.2. Elméleti alapok

Legyen  $N$  a károk száma és  $Y_j$  a kárnagyság ( $j = 1, 2, \dots, N$ ). Ezekről feltesszük, hogy függetlenek és utóbbiak azonos eloszlásúak is. Jelöljük  $d$ -vel az önrész nagyságát, így a következő függvényeket definiálhatjuk a kockázat megosztás szempontjából:

$$g(Y_j; d) = \begin{cases} 0 & Y_j \leq d \\ Y_j - d & d < Y_j < \infty \end{cases}$$

Legyen a kárszám az önrész figyelembevételével:

$$N_g(d) = \sum_{j=1}^N I(d < Y_j)$$

A kárnagyság a cenzorált megfigyelésekre:

$$Y_{g,j}(d) = \begin{cases} 0 & Y_j \leq d \\ Y_j - d & d < Y_j < \infty \end{cases}$$

És a "csonkított" megfigyelésekre:

$$Y_{*,j}(d) = Y_j - d \mid d < Y_j$$

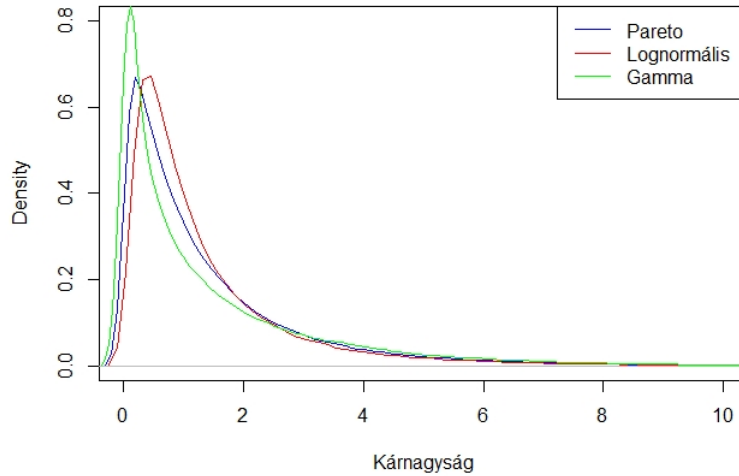
A továbbiakban "csonkított" megfigyelésekkel fogunk dolgozni. A gyakorlatban is az a jellemzőbb, hogy ezek az alacsony összegű károk nem jutnak a biztosító tudomására.

A fentiek alapján az aggregált kárra a következő képletet írhatjuk fel:

$$S_g(d) = \sum_{j=1}^N Y_{g,j}(d)$$

Szeretném kihangsúlyozni ezen a ponton, hogy a vizsgált eloszlás és az önrész nagyságának választása közel sem lényegtelen kérdés. Ennek bemutatására generáltam 100 000 elemű mintákat lognormális, gamma és (amerikai típusú) Pareto eloszlásokból, melyek várható értéke és szórásnégyzete megegyezik. A következő ábrák és táblázatok alapján jól látszik, hogy az önrész növelésével mennyire különböző az egyes eloszlások viselkedése. Mind az ábrák és mind a táblázatok értékei a "csonkított" mintából lettek számítva.

Ahogy a 2.2 ábrán is láthatjuk, mind három eloszlás jobbra ferde, de a káreloszlásoktól pont ezt is várjuk el. Mivel azonos az első és a második momentumuk, így nagyon hasonló alakot vesznek fel a tapasztalati sűrűség-

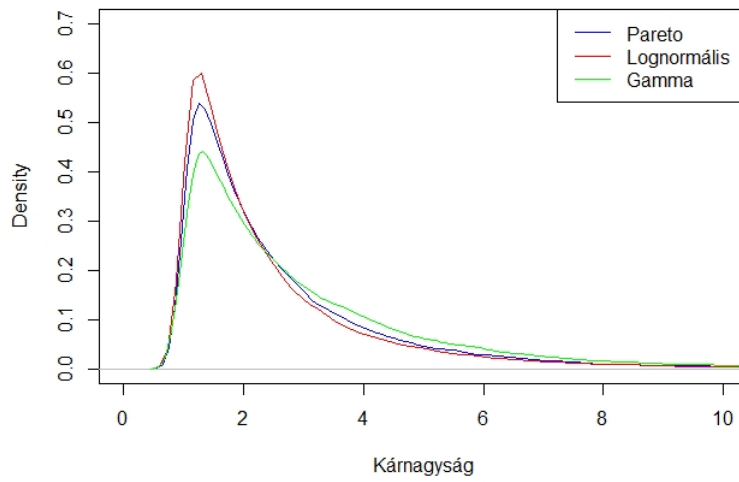


2.2. ábra. Tapasztalati sűrűségfüggvények, ha nincs önrész

függvényeik. Az első két momentummal ellentétben a harmadik és negyedik momentum, az úgynevezett ferdeség és csúcsosság esetében már szemmel is jól látható különbségeket tapasztalhatunk. A csúcsosság tekintetében a legszembetűnőbb ez a különbség, a gamma eloszlásból generált minta lényegesen különbözik a másik kettőtől, sokkal csúcsosabb náluk.

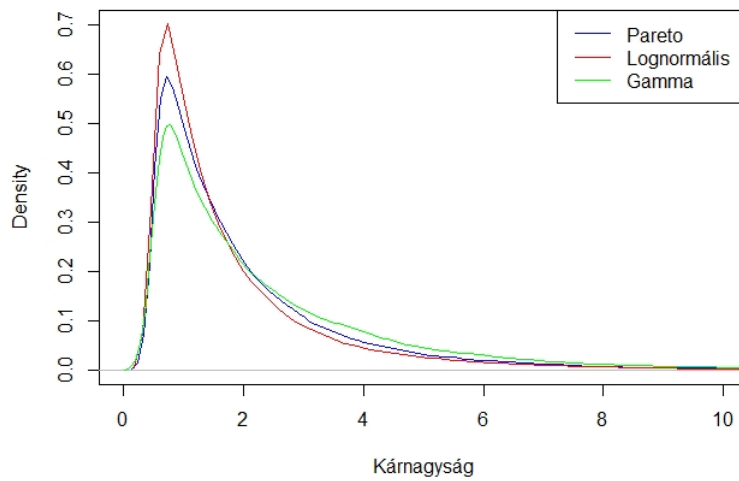
Ha bevezetünk egy a várható értéktől nem sokkal kisebb önrészt, akkor a következtetések levonásában a 2.3 ábra lesz a segítségünkre.

A jobbra ferdeség természetesen megmaradt, azonban a fenti állapothoz képest  $x = 1$ -be "tolódott" az ábra. Ez az önrésznek köszönhető. A csúcsosság esetében történt a legtöbb és legszembetűnőbb változás. Az eloszlások tulajdonképpen sorba rendeződtek, a csúcspontjaikat gyakorlatilag azonos helyen veszik fel, viszont a kiindulási állapottól eltérően ezúttal a gamma eloszlás csúcsosságának értéke a legkisebb.



2.3. ábra. Tapasztalati sűrűségfüggvények önréssel ( $d = 1$ )

Ha az önrész értéke a várható érték harmadával egyezik meg, akkor a 2.4 ábrához hasonlókat tapasztalunk:



2.4. ábra. Tapasztalati sűrűségfüggvények önréssel ( $d = 0.5$ )

Kisebb önrész esetében csúcsosabbak a vizsgált eloszlások. A várható érték és variancia változását az önrész függvényében a következő táblázatban láthatjuk.

	$d = 0$		$d = 0.5$		$d = 1$	
	$\mathbb{E}(Y_*)$	$Var(Y_*)$	$\mathbb{E}(Y_*)$	$Var(Y_*)$	$\mathbb{E}(Y_*)$	$Var(Y_*)$
Gamma	1.491	3.939	2.415	4.612	3.042	4.870
Lognormális	1.497	3.888	1.969	4.624	2.673	5.865
Pareto	1.502	3.902	2.143	4.631	2.784	5.408

2.1. táblázat. Különböző (kár)eloszlások önrésztől való függése (tapasztalati értékek)

A számok is jól mutatják, hogy míg önrész nélküli esetben a várható érték és a variancia közel megegyezik mindhárom esetben (szándékosan így választottuk a paramétereket), addig az önrész bevezetésével már jelentős különbségek alakulnak ki. A legnagyobb eltérést a gamma eloszlás esetében láthatjuk, a lognormális és a Pareto hasonlóan viselkedik. A ferdeség és csúcsosság mérőszámai már az önrész nélküli esetben is nagyon különbözőek, de az önrész bevezetése okozta változásokat a következő táblázat segítségével nyomon követhetjük.

	$d = 0$		$d = 0.5$		$d = 1$	
	$\tilde{\mu}_3(Y_*)$	$\kappa(Y_*)$	$\tilde{\mu}_3(Y_*)$	$\kappa(Y_*)$	$\tilde{\mu}_3(Y_*)$	$\kappa(Y_*)$
Gamma	2.611	12.951	2.303	10.852	2.206	10.225
Lognormális	5.504	68.818	5.318	62.832	5.016	54.804
Pareto	4.195	36.883	4.067	33.920	3.943	31.288

2.2. táblázat. Különböző (kár)eloszlások önrésztől való függése (tapasztalati értékek)

Az eloszlások ferdesége és "vastag farkúsága" a táblázatban látható adatokon is nyomon követhető. A lognormális eloszlás ezen paraméterek mellett a legferdebb, legcsúcsosabb és ez is közelít a leggyorsabban a nullához. A gamma várható értéke lesz a legnagyobb, miközben a varianciája messze elmarad a másik kettőtől.

Az aggregált kár azonban nem csak a kárnagyságon múlik. Keveset beszéltünk eddig magáról a kárszámról. A munkám során alapvetően az egy-

szerűség kedvéért Poisson-eloszlást tételezek fel, de természetesen egyéb lehetőségek is vannak. A Poisson-eloszlás feltételezését annak tulajdonságai indokolják. Amikor önrészes biztosításokat vizsgálunk, és "csonkított" mintákkal dolgozunk, akkor a kezdeti feltételezéseinket sokszor újra kell gondolni.

A Poisson-eloszlásnak van egy olyan szép tulajdonsága, hogy önrész bevezetése mellett is Poisson marad az eloszlás, csak a paraméter változik meg. Ennek levezetését a generátorfüggvény segítségével mutatom meg, ugyanis a generátorfüggvény és az eloszlás kölcsönösen meghatározzák egymást. Első lépésben belátjuk, hogy  $G_{N_g(d)}(z) = G_N(G_{I(d<Y)}(z))$ . A levezetés során kihasználjuk, hogy  $Y_1, Y_2, \dots$  *iid* és  $N$ -től is függetlenek.

$$\begin{aligned} G_{N_g(d)}(z) &= \mathbb{E}z^{\sum_{j=1}^N I(d<Y_j)} = \sum_{i=1}^{\infty} \mathbb{E}(z^{\sum_{j=1}^N I(d<Y_j)} | N = i) \cdot \mathbb{P}(N = i) = \\ &= \sum_{i=1}^{\infty} \mathbb{E}(z^{\sum_{j=1}^i I(d<Y_j)}) \cdot \mathbb{P}(N = i) = \sum_{i=1}^{\infty} (\mathbb{E}z^{I(d<Y)})^i \cdot \mathbb{P}(N = i) = \\ &= G_N(G_{I(d<Y)}(z)) \end{aligned}$$

A következő lépés  $G_{I(d<Y)}(z)$  meghatározása. Legyen  $p = \mathbb{P}(Y > d)$ .

$$G_{I(d<Y)}(z) = zp + 1 \cdot (1 - p)$$

Végül a fenti két képletből és a Poisson-eloszlás generátorfüggvényéből ( $G_N(z) = e^{\lambda(z-1)}$ ) már közvetlenül adódik:

$$G_{N_g(d)}(z) = G_N(G_{I(d<Y)}(z)) = G_N(zp + (1 - p)) = e^{\lambda(zp + (1-p) - 1)} = e^{\lambda p(z-1)}$$

Tehát az önrész bevezetése után  $N_g(d) \sim Poi(\lambda p)$ .

Hasonló levezetéssel élhetünk a negatív binomiális eloszlás esetében. Ahogy a Nem-élet biztosításmatematika című könyvben [5] is láthatjuk, a negatív binomiális eloszlást elő tudjuk állítani gamma és Poisson eloszlások keverékeként. Legyen  $v$  gamma eloszlású  $(\lambda, \alpha)$  paraméterekkel, és legyen  $\eta$  rögzített  $v$  mellett Poisson. Vagyis  $\eta|v \sim Poi(vt)$ . Ekkor  $\eta$  negatív binomiális eloszlású  $(\alpha, \frac{t}{t+\lambda})$  paraméterekkel. Az önrész bevezetése itt sem változtatja meg



az eloszlás típusát. Ahogy a könyvben is olvashatjuk, ilyen feltételek mellett az (önrész nélküli) eloszlás generátorfüggvénye a következő alakot ölti:  $G_\eta(z) = (\frac{\lambda}{\lambda+t(1-z)})^\alpha$ . Ezt felhasználva:

$$\begin{aligned} G_{N_g(d)}(z) &= G_N(G_{I(d<Y)}(z)) = G_N(zp + (1-p)) = \\ &= (\frac{\lambda}{\lambda+t(1-(zp+1-p))})^\alpha = (\frac{\lambda}{\lambda+tp(1-z)})^\alpha \end{aligned}$$

Ez pedig a negatív binomiális eloszlás generátorfüggvénye  $(\alpha, \frac{tp}{\lambda+tp})$  paraméterekkel.

## 2.3. Önrész "mérése"

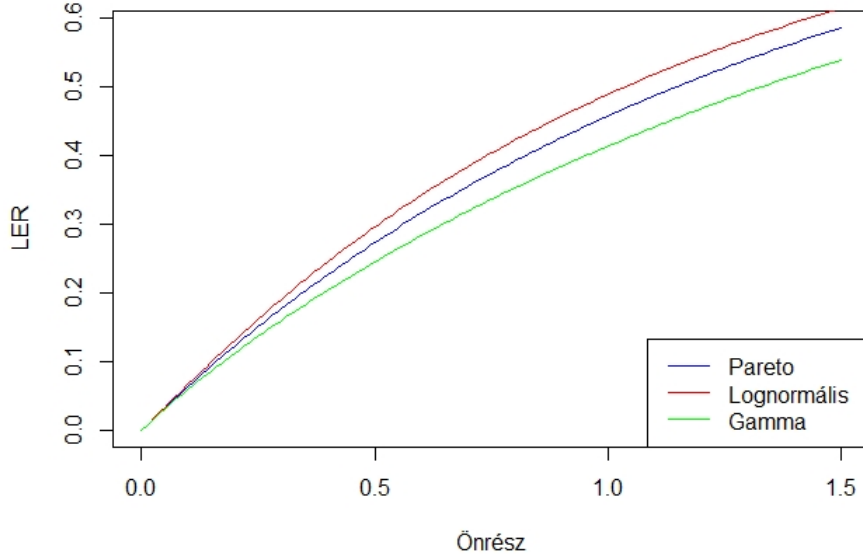
A következő mutatók, mérőszámok szintén a már fent is említett Gee Y. Lee [4] cikkén alapulnak.

Különböző mérőszámokat használhatunk, ha önrészes biztosításokkal dolgozunk. A *LER* (*loss elimination ratio*) megmutatja, hogy a várható kár mekkora hányada hárul a szerződőre. Ha a káreloszlás sűrűségfüggvénye  $f_Y(y)$ , akkor a *LER* értéke az önrész függvényében:

$$LER(d) = \frac{\int_0^d y f_Y(y) dy + d \int_d^\infty f_Y(y) dy}{\int_0^\infty y f_Y(y) dy}$$

Az előző alfejezetben már bemutatott eloszlások és a belőlük generált minta segítségével nézzük meg, hogy miként függ a *LER* értéke az önrész nagyságától és az eloszlás típusától.

A három különböző eloszlás nagyon hasonlóan viselkedik, a *LER* értéke szigorúan monoton növekvő függvénye az önrésznek a vizsgált intervallumban. A fő különbséget itt is az eloszlások ferdesége okozza. A lognormális eloszlás adja a legmagasabb értékeket, és ez tulajdonképpen az előző alfejezetben tett megfigyelések közvetlen következménye. Ehhez nyújt segítséget a 2.5 ábra.



2.5. ábra.  $LER$  értéke az önrész függvényében

Ehhez hasonló az úgynevezett indicated deductible relativity, ami a cenzorált és a normál megfigyelések várható értékének hányadosa (vagyis a károk mekkora része hárul a biztosítóra). Ez az arány a fenti  $LER$  segítségével is kifejezhető:

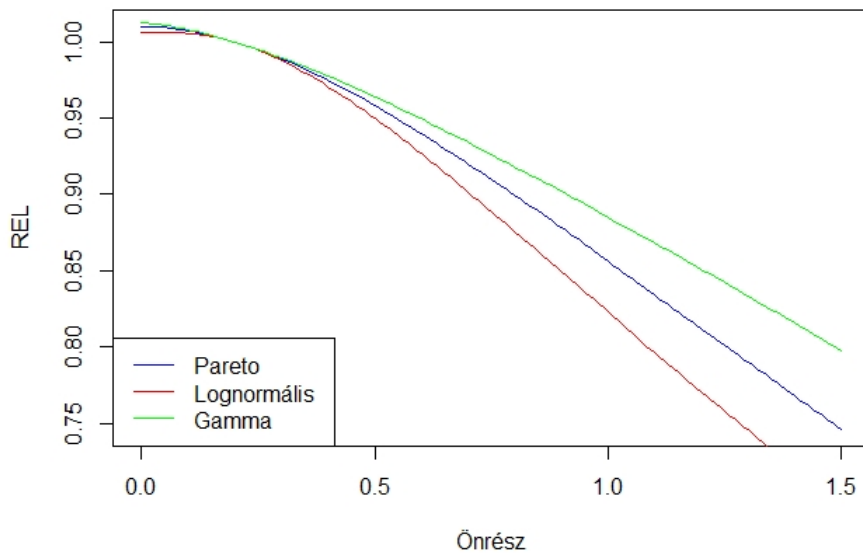
$$\text{indicated deductible relativity} = \frac{E(Y_*(d))}{E(Y(d))} = 1 - LER(d)$$

Az aggregált károkra definiáljuk a  $REL$  (*relativity of aggregate loss*) mutatót, ami megmutatja, hogy az önrész értékének megváltoztatása hogyan befolyásolja az aggregált kárt az úgynevezett bázis önrészhez ( $d_0$ ) képest:

$$REL(d_0, d) = \frac{E[S_g(d)]}{E[S_g(d_0)]} = \frac{E[N] \int_d^\infty (1 - F_Y(y)) dy}{E[N] \int_{d_0}^\infty (1 - F_Y(y)) dy} = \frac{\int_d^\infty (1 - F_Y(y)) dy}{\int_{d_0}^\infty (1 - F_Y(y)) dy} \quad (2.5)$$

A 2.6 ábra szemlélteti a  $REL$  értékének változását az önrész függvényében. A bázis önrész értékének  $d_0 = 0.2$ -t választottam, ezen a helyen lesz

1 a függvény értéke. Minél magasabb az önrész a bázis önrészhez képest, az aggregált kár értéke annál alacsonyabb lesz. A három vizsgált eloszlás esetében a csökkenés mértéke különböző. Ez a már fentebb is említett okokra vezethető vissza.



2.6. ábra.  $REL$  értéke az önrész függvényében

Mind a kárszám, mind a kárnagyság esetében GLM modellel dolgozunk. Ha a biztosítás tartalmaz önrészt, akkor az önrész nagysága is magyarázó változó lesz a modellben. Legyen  $Y_*$  a megfigyelt kár,  $\mathbf{x}_i$  a magyarázó változók és  $\boldsymbol{\beta}$  a modell segítségével kapott együtthatók vektora. A modellben vegyük az önrész logaritmusát, mint magyarázó változót. Ekkor:

$$E(Y_*(d_i)) = \exp(\mathbf{x}_i \boldsymbol{\beta}) = \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_d \log(d_i)),$$

ahol  $\beta_d$  az önrész, mint magyarázó változó együtthatója. Hasonló egyenletet tudunk felírni a kárszám esetében is:

$$E(N_g(d_i)) = \exp(\mathbf{x}_i \boldsymbol{\gamma}) = \exp(\mathbf{x}'_i \boldsymbol{\gamma} + \gamma_d \log(d_i)),$$

ahol  $\boldsymbol{\gamma}$  a magyarázó változókhoz tartozó együtthatók vektora, és  $\gamma_d$  pedig az önrészhez, mint magyarázó változóhoz tartozó együttható.

A 2.5 képletet vizsgáljuk meg ebben az esetben is. A kárnagyságot és kárszámot a fent látható módon felhasználhatjuk, és így adódik a következő:

$$REL(d_0, d) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta} + \beta_d \log(d)) \exp(\mathbf{x}'\boldsymbol{\gamma} + \gamma_d \log(d))}{\exp(\mathbf{x}'\boldsymbol{\beta} + \beta_d \log(d_0)) \exp(\mathbf{x}'\boldsymbol{\gamma} + \gamma_d \log(d_0))} = \left(\frac{d}{d_0}\right)^{\beta_d + \gamma_d}, \quad (2.6)$$

ahol feltesszük, hogy  $\beta_d + \gamma_d < 0$ .

## 2.4. Az önrész hatása a becslésre

Az aktuárius irodalomban nem kap elég figyelmet az önrész bevezetésének hatása. Jostein Paulsen és Knut Stubø [6] tanulmányukban erre világítanak rá. Mint írják, egy teljes maximum likelihood becslésben az összes paramétert kell egyidejűleg megbecsülni. Ehhez képest a pseudo-maximum likelihood becslésben először megbecsüljük a kárnagyságot, számításba vesszük az önrészt, és utána az így kapott becült paraméterek segítségével vizsgálhatjuk a kárszámot is. Ez a módszer kevésbé hatékony, de ha figyelembe vesszük a feladat komplexitását, időigényét, akkor a pseudo-maximum likelihood becslés egy jó alternatív megoldás lehet.

A fent említett cikkben az önrészt is valószínűségi változóként vizsgálták. A munkám során én ettől eltekintettem, de a teljesség igénye miatt szeretném bemutatni, hogy ebben az esetben mi a teendő. Fontos különbség még, hogy a tanulmányban meghaladásos önrésszel dolgoznak.

Röviden szeretném összefoglalni, hogy a fenti feltételek mellett hogyan alakul a pseudo-maximum likelihood becslés. Legyen  $D_i$  az  $i$ . szerződéshez tartozó önrész, mint valószínűségi változó. Ekkor a kárnagyságra meghaladásos önrész mellett adódik:

$$F_{Y_*}(x) = \begin{cases} 0 & , Y \leq D_i \\ \frac{F_Y(x; \boldsymbol{\theta}) - F_Y(D_i; \boldsymbol{\theta})}{1 - F_Y(D_i; \boldsymbol{\theta})} & , D_i < Y < \infty \end{cases}$$

A kárszám eloszlása legyen:

$$\mathbb{P}(N_i = m | D_i) = p_i(m; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i),$$

ahol  $\boldsymbol{\theta}$  egy ismeretlen  $s$  dimenziós paramétervektor és  $\boldsymbol{\lambda}$  egy  $t$  dimenziós paraméter. Legyen  $F_i$  a kárnagysághoz tartozó eloszlásfüggvény, és  $G_i$  az önrészhez tartozó eloszlásfüggvény, és a sűrűségfüggvények legyenek ennek megfelelően  $f_i$  és  $g_i$ . Ekkor a likelihood függvény a következő alakot ölti az eddig ismert feltételek mellett:

$$L_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^n \left( p_i(m; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i) \prod_{j=1}^{N_i} \frac{f_i(Y_{ij}; \boldsymbol{\theta})}{1 - F_i(D_i; \boldsymbol{\theta})} \right) \cdot \prod_{i=1}^n g_i(D_i) \quad (2.7)$$

Ez a likelihood függvény könnyen átvezethető az általam vizsgált esetre, ahol az önrész konstans és meghaladásos helyett levonásos önrészt alkalmazunk:

$$L_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^n \left( p_i(m; (\boldsymbol{\theta}, \boldsymbol{\lambda}), d) \prod_{j=1}^{N_i} \frac{f_i(Y_{ij} + d; \boldsymbol{\theta})}{1 - F_i(d; \boldsymbol{\theta})} \right) \quad (2.8)$$

A következő lépésben induljunk ki a kárnagyságból és tegyük fel, hogy a kárszám és az önrész ismertek. Ekkor a likelihood a következő alakra egyszerűsödik:

$$\tilde{L}_n(\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{j=1}^{N_i} \frac{f(Y_{ij}; \boldsymbol{\theta})}{1 - F_i(D_i; \boldsymbol{\theta})} \quad (2.9)$$

Legyen  $\tilde{\boldsymbol{\theta}}_n$ , ami maximalizálja (2.9)-et. Térjünk vissza az eredeti likelihood függvényre, és keressük  $\tilde{\boldsymbol{\lambda}}_n$ -t, ami maximalizálja (2.7)-et. A feladat a következő alakra egyszerűsödik:

$$L_{0(n)}(\tilde{\boldsymbol{\theta}}_n; \boldsymbol{\lambda}) = \prod_{i=1}^n p_i(N_i; (\tilde{\boldsymbol{\theta}}_n; \boldsymbol{\lambda}), D_i) \quad (2.10)$$

Az így kapott  $(\tilde{\boldsymbol{\theta}}_n; \tilde{\boldsymbol{\lambda}})$  becslést hívjuk pszeudo-maximum likelihood becslésnek.

A becslés aszimptotikus viselkedésének vizsgálatához tekintsük az *ARE* (asymptotic relative efficiency) mutatót, amely a következő alakban írható fel:

$$ARE(h(\tilde{\boldsymbol{\theta}}; \tilde{\boldsymbol{\lambda}})) = \frac{\hat{\sigma}^2}{\tilde{\sigma}^2},$$

ahol  $h(\tilde{\boldsymbol{\theta}}; \tilde{\boldsymbol{\lambda}})$  függvénye és feltesszük továbbá, hogy:

$$c_n(h(\tilde{\boldsymbol{\theta}}; \tilde{\boldsymbol{\lambda}}) - h(\boldsymbol{\theta}; \boldsymbol{\lambda})) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2), \text{ ha } n \rightarrow \infty$$

$$c_n(h(\hat{\boldsymbol{\theta}}; \hat{\boldsymbol{\lambda}}) - h(\boldsymbol{\theta}; \boldsymbol{\lambda})) \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}^2), \text{ ha } n \rightarrow \infty$$

A pszeudo becslést hasznosnak tekinthetjük, ha az *ARE* érték megközelíti az egyet. A már említett tanulmányban láthatjuk, hogy konkrét eloszlások esetén hogyan számolható ki ez az érték. Legyen az önrész nagysága  $D_i \sim \text{Exp}(\alpha)$ , a kár nagyság  $Y \sim \text{Exp}(\theta)$  és a kárszám pedig  $N \sim \text{Poi}(\lambda)$  eloszlású. Ekkor [6] alapján:

$$ARE(\theta) = 1 - \frac{1}{1 + \left(1 + \frac{\mathbb{E}(Y_*)}{\mathbb{E}(D)}\right)^2} \quad (2.11)$$

$$ARE(\lambda) = 1 - \frac{1}{\left(1 + \left(1 + \frac{\mathbb{E}(Y_*)}{\mathbb{E}(D)}\right)^2\right)^2} = 1 - (1 - ARE(\theta))^2 \quad (2.12)$$

$$ARE(m_d(\theta, \lambda)) = 1 - \frac{\left(\frac{1}{1 + \frac{\mathbb{E}(Y_*)}{\mathbb{E}(D)}} + \frac{d}{\mathbb{E}(Y_*)}\right)^2}{\left(1 + \left(1 + \frac{\mathbb{E}(Y_*)}{\mathbb{E}(D)}\right)^2\right) \left(1 + \left(\frac{1}{1 + \frac{\mathbb{E}(Y_*)}{\mathbb{E}(D)}} + \frac{d}{\mathbb{E}(Y_*)}\right)^2\right)} \quad (2.13)$$

Ahol  $m_d(\theta, \lambda) = \mathbb{E}(S_g(d))$  a levonásos önrész alapján számolt várható nettó kára a biztosítónak.

Ugyanezen feltételezések mellett a pseudo-MLE és az MLE becslést fogom bemutatni. Először nézzük a pseudo esetét, ehhez induljunk ki (2.9)-ből:

$$\tilde{L}_n(\theta) = \prod_{i=1}^n \prod_{j=1}^{N_i^*} \frac{f(Y_{ij}; \theta)}{1 - F_i(D_i; \theta)}$$

Ebből a loglikelihood egyszerűen megkapható:

$$\begin{aligned} \tilde{l}_n(\theta) &= \sum_{i=1}^n \sum_{j=1}^{N_i^*} (\log(f(Y_{ij}; \theta)) - \log(1 - F_i(D_i; \theta))) = \\ &= \sum_{i=1}^n \left( \sum_{j=1}^{N_i^*} \log(f(Y_{ij}; \theta)) \right) - N_i^* \log(1 - F_i(D_i; \theta)) = \\ &= \sum_{i=1}^n \left( \sum_{j=1}^{N_i^*} \log(\theta e^{-\theta Y_{ij}}) \right) - N_i^* \log(e^{-\theta D_i}) = \\ &= \sum_{i=1}^n \left( \sum_{j=1}^{N_i^*} \log(\theta) - \theta Y_{ij} \right) + N_i^* \theta D_i \end{aligned}$$

Ezután a szokásos módon nézzük a  $\theta$  szerinti deriváltját:

$$\begin{aligned} \frac{\partial \tilde{l}}{\partial \theta} &= \sum_{i=1}^n \left( \sum_{j=1}^{N_i^*} \frac{1}{\theta} - Y_{ij} \right) + N_i^* D_i = \\ &= \sum_{i=1}^n \frac{N_i^*}{\theta} - \sum_{i=1}^n \sum_{j=1}^{N_i^*} Y_{ij} + \sum_{i=1}^n N_i^* D_i \end{aligned}$$

Végül, ha a kapott eredményt egyenlővé tesszük nullával, akkor  $\tilde{\theta}$  becsléssel készen vagyunk, és a következő eredményt kapjuk:

$$\tilde{\theta} = \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n \sum_{j=1}^{N_i^*} Y_{ij} - \sum_{i=1}^n N_i^* D_i}$$

Miután az első paraméterre felírtuk a becslést, térjünk át a másik paraméterre, ehhez (2.10)-et használjuk (emlékeztetőül, Poisson eloszlás esetében az önrész megjelenése után az eloszlás Poisson marad, a paraméter változik):

$$\begin{aligned} L_{0(n)}(\tilde{\theta}_n; \lambda) &= \prod_{i=1}^n p_i(N_i; (\tilde{\theta}_n; \lambda), D_i) = \prod_{i=1}^n \frac{(\lambda(1 - F(D_i; \tilde{\theta})))^{N_i^*}}{N_i^*!} e^{-\lambda(1 - F(D_i; \tilde{\theta}))} = \\ &= \prod_{i=1}^n \frac{(\lambda e^{-\tilde{\theta} D_i})^{N_i^*}}{N_i^*!} e^{-\lambda e^{-\tilde{\theta} D_i}} \end{aligned}$$

Ebből kifejezve a loglikelihood:

$$l_{0(n)}(\tilde{\theta}_n; \lambda) = \sum_{i=1}^n \log\left(\frac{(\lambda e^{-\tilde{\theta} D_i})^{N_i^*}}{N_i^*!} e^{-\lambda e^{-\tilde{\theta} D_i}}\right) = \sum_{i=1}^n N_i^* (\log(\lambda) - \tilde{\theta} D_i) - \log(N_i^*!) - \lambda e^{-\tilde{\theta} D_i}$$

A  $\lambda$  szerinti deriváltat véve jelentősen egyszerűsödik a képlet:

$$\frac{\partial \tilde{l}_0}{\partial \lambda} = \sum_{i=1}^n \frac{N_i^*}{\lambda} - e^{-\tilde{\theta} D_i}$$

A már megszokott módon a fenti egyenletet nullával egyenlővé téve, a következő megoldást kapjuk  $\tilde{\lambda}$  becslésre:

$$\tilde{\lambda} = \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\tilde{\theta} D_i}} = \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\left(\frac{\sum_{k=1}^n N_k^*}{\sum_{k=1}^n \sum_{j=1}^{N_k^*} Y_{kj} - \sum_{k=1}^n N_k^* D_k}\right) D_i}}$$

Az MLE becsléshez vegyük a (2.7) likelihood függvényt, és helyettesítsük be a megfelelő eloszlás-, illetve sűrűségfüggvényeket. Ekkor a likelihood függvény az alábbi módon fog kinézni:



$$\hat{L}_n(\theta; \lambda) = \prod_{i=1}^n \left( \frac{(\lambda e^{-\theta D_i})^{N_i^*}}{N_i^*!} e^{-\lambda e^{-\theta D_i}} \prod_{j=1}^{N_i^*} \frac{\theta e^{-\theta Y_{ij}}}{e^{-\theta D_i}} \right) \prod_{i=1}^n \alpha e^{-\alpha D_i}$$

Ebból a loglikelihood:

$$\begin{aligned} \hat{l}_n(\theta; \lambda) &= \sum_{i=1}^n \left( N_i^* (\log(\lambda) - \theta D_i) - \log(N_i^*!) - \lambda e^{-\theta D_i} + \sum_{j=1}^{N_i^*} \log(\theta) - \theta Y_{ij} + \theta D_i \right) + \\ &\quad + \sum_{i=1}^n \log(\alpha) - \alpha D_i = \\ &= \sum_{i=1}^n \left( N_i^* \log(\lambda) - \log(N_i^*!) - \lambda e^{-\theta D_i} + \sum_{j=1}^{N_i^*} \log(\theta) - \theta Y_{ij} \right) + \\ &\quad + \sum_{i=1}^n \log(\alpha) - \alpha D_i \end{aligned}$$

A kapott loglikelihoodnak vegyük a  $\lambda$  és  $\theta$  szerinti parciális deriváltjait.

$$\frac{\partial \hat{l}}{\partial \lambda} = \frac{\sum_{i=1}^n N_i^*}{\lambda} - \sum_{i=1}^n e^{-\theta D_i}$$

Ebből már kapunk egy részeredményt  $\hat{\lambda}$ -ra:

$$\hat{\lambda} = \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\theta D_i}} \quad (2.14)$$

Most nézzük a  $\theta$  szerinti deriváltat, és használjuk a kapott (2.14) eredményt.

$$\frac{\partial \hat{l}}{\partial \theta} = \sum_{i=1}^n \hat{\lambda} D_i e^{-\theta D_i} + \sum_{j=1}^{N_i^*} \frac{1}{\theta} - Y_{ij} = \quad (2.15)$$

$$= \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\theta D_i}} \sum_{i=1}^n D_i e^{-\theta D_i} + \frac{\sum_{i=1}^n N_i^*}{\theta} - \sum_{i=1}^n \sum_{j=1}^{N_i^*} Y_{ij} \quad (2.16)$$

Összefoglalva az eredményeket,  $\hat{\theta}$  becslésre az alábbi egyenletet kell megoldani

dani  $\hat{\theta}$  -ra:

$$\frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\theta D_i}} \sum_{i=1}^n D_i e^{-\theta D_i} + \frac{\sum_{i=1}^n N_i^*}{\theta} - \sum_{i=1}^n \sum_{j=1}^{N_i^*} Y_{ij} = 0 \quad (2.17)$$

Majd az így kapott  $\hat{\theta}$  segítségével már felírható  $\hat{\lambda}$  is:

$$\hat{\lambda} = \frac{\sum_{i=1}^n N_i^*}{\sum_{i=1}^n e^{-\hat{\theta} D_i}} \quad (2.18)$$

Az itt bemutatott eredmények prezentálásához generáltam különböző méretű véletlen mintákat. A kárszám esetében a Poisson eloszlás paraméterét 0.05 -nek, az exponenciális eloszlások esetében pedig a kárnagysághoz tartozó paramétert 0.1 -nek, míg az önrészét 1 -nek választottam. Ezen feltételek mellett a következő eredményeket kaptam:

	$\tilde{\theta}$	$\tilde{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$	$ARE(\theta)$	$ARE(\lambda)$
$n = 500$	0.0941	0.0459	0.0915	0.0458	0.9941	1.0000
$n = 1000$	0.1212	0.0347	0.1238	0.0348	0.9903	0.9999
$n = 5000$	0.1028	0.0535	0.1037	0.0535	0.9924	0.9999
$n = 10000$	0.1004	0.0489	0.1007	0.0489	0.9929	0.9999

2.3. táblázat. Különböző mintaelemszámra a pseudo-MLE és MLE különbsége

A  $\lambda$  paraméter esetében a becslések lényegében azonosak a mintaelemszámtól függetlenül. A  $\theta$  paraméter esetében figyelhető meg eltérés, de az elemszám növekedésével ez a különbség megszűnik. Mivel egyparaméteres eloszlásokból generáltam mintát, ezért a pseudo-MLE és MLE közti különbség nem jelenik meg élesen. Minél komplexebb eloszlásokat vizsgálunk, annál jobban megjelenik a különbség a két becslés között. Ez egy érdekes kutatási feladat lehetne, ha nem csak ezekre az úgymond egyszerűbb eloszlástípusokra vizsgálnánk a becsléseket, hanem többparaméteres eloszlásokat vennénk mind a kárszám, mind a kárnagyságra.

Most nézzük meg, hogy mi történik abban az esetben, ha a mintaelem-szám azonos, de különböző paraméterű exponenciális eloszlásokat vizsgálunk.

	$\tilde{\theta}$	$\tilde{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$	$ARE(\theta)$	$ARE(\lambda)$
$\theta = 0.5$	0.4809	0.1054	0.4782	0.1052	0.9326	0.9955
$\theta = 0.3$	0.3281	0.0981	0.3233	0.0977	0.9585	0.9983
$\theta = 0.2$	0.2059	0.0926	0.2058	0.0926	0.9790	0.9996
$\theta = 0.1$	0.0995	0.0997	0.1000	0.0998	0.9931	1.0000

2.4. táblázat. Különböző paraméterű kárnagyság-eloszlások esetén a pseudo-MLE és MLE különbsége

Az önrész paraméterének 1-et választottam, és jól kivehető, hogy minél közelebb van az önrész és a kárnagyság paramétere egymáshoz, a pseudo-MLE becslés annál gyengébb hatásfokú. Itt azonban még szintén apró eltérésről beszélhetünk, hiszen az  $ARE$  értéke minden esetben 0.93 felett van.

## 3. fejezet

# Modellek

### 3.1. Az adatok bemutatása

Az előző fejezetben ismertetett elméleti alapokat vizsgáljuk meg szimulált adatok segítségével. A kárszámot generáljuk Poisson eloszlásból és az eloszlás paramétere legyen alapesetben 0.07, azonban minden egyes szerződés esetében ezt az értéket befolyásoljuk különböző változók figyelembevételével. Magyarázó változónak a jármű színét és teljesítményét, valamint a szerződő korát és lakóhelyét (lakóhelyének típusát) választottam. A lakóhely esetében 4 csoportot állapítottam meg, és a legkisebb sorszámút tekintetem a legkockázatosabbnak (pl. Budapest). A teljesítménnyel egyenes arányban növeltem a kockázatosságot. Az életkor esetében a fiatalabb és az idősebb korosztályt tekintetem kockázatosnak. Négyféle szint választottam magyarázó változónak, melyek pozitív vagy negatív hatással voltak a kockázatosságra.

A kárnagyság esetében a már fent ismertetett módszert ismételttem meg. Három különböző eloszlásból generáltam a mintámat, melyek paramétereit úgy választottam meg, hogy a magyarázó változók befolyása nélkül azonos várható értéket és szórást vegyenek fel. A három vizsgált kárnagyság eloszlás pedig a gamma, a lognormális és a Pareto.

Végül az így kapott mintákból létrehoztam "csonkított" megfigyeléseket, kiválasztottam különböző nagyságú önrészeket és az egyes önrész nagyságokhoz rendeltem a "csonkított" megfigyeléseket.

A következő táblázatokban bemutatom, hogy a különböző eloszlásokból generált minták esetén hogyan néznek ki a minta alapstatisztikái a károkozókra szűkítve (250000 nagyságú minta esetén).

	$\mathbb{E}(Y)$	$Var(Y)$	$\tilde{\mu}_3(Y)$	$\kappa(Y)$	$N$
Gamma	51426.7	$1.03 \cdot 10^{10}$	3.81	23.60	19519
Lognormális	62632.7	$1.52 \cdot 10^{10}$	9.23	165.22	19519
Pareto	45952.1	$8.48 \cdot 10^9$	10.62	200.77	19519

3.1. táblázat. A különböző eloszlásokból generált kárnagyságok néhány jellemzője

És ugyanez a táblázat az önrész bevezetése után:

	$\mathbb{E}(Y_*)$	$Var(Y_*)$	$\tilde{\mu}_3(Y_*)$	$\kappa(Y_*)$	$N$
Gamma	121437.0	$1.73 \cdot 10^{10}$	2.67	13.29	8077
Lognormális	97430.5	$2.23 \cdot 10^{10}$	7.98	120.33	11946
Pareto	76906.6	$1.35 \cdot 10^{10}$	8.94	135.69	10826

3.2. táblázat. A különböző eloszlásokból generált kárnagyságok néhány jellemzője önrész mellett

A legszembevetőbb változás az önrész bevezetésével az átlagkárban figyelhető meg, ezen belül is a legnagyobb változás a Gamma-eloszlás esetében látható. A kárszám jelentős mértékben csökken, miközben az átlagkár több, mint kétszeresére nő. Ez jól illeszkedik az elméleti bevezetőben látott 2.2 ábrára is, ahol a Gamma-eloszlás volt a legcsúcsosabb, és a leginkább 0-ra koncentrált. Ennek következtében az önrész bevezetése ezt érintette a leginkább érzékenyen, ahogy a károk darabszámának csökkenésében is látszik. A szórásnégyzet esetében mindegyik eloszlásra növekedést látunk, míg a ferdeség és csúcsosság különböző mértékben, de mindenhol csökken.

Az RStudio *insuranceData*<sup>1</sup> elnevezésű csomagjából a *dataOhlsson* adatfájlt választottam elemzésre. Kellő nagyságú a minta - 64548 megfigyelést tartalmaz - és nincs önrész a károkon, vagyis nem csonkított a minta. Röviden bemutatom a tiszta adatokat, majd egy általam bevezetett önrész után a csonkított megfigyeléseket is.

<sup>1</sup><https://cran.r-project.org/web/packages/insuranceData/insuranceData.pdf>

Az adat kilenc változót tartalmaz, melyek közül a kárszám és a kárnagyság a célváltozó, a maradék hét pedig a magyarázó változók. A tulajdonos életkora szerepel számként és faktorként is, mi csak a faktorral fogunk dolgozni. Ezen kívül szerepel még a jármű életkora, területi besorolás, bónusz-málusz besorolás, a szerződés tartama, és egy úgynevezett *MCClass* elnevezésű változó, amit a jármű teljesítményéből és tömegéből számoltak. Az önrész nagyságának négy különböző lehetőségből generáltam véletlenszerűen mintát, amelyet hozzárendeltem minden szerződéshez. A kárnagyságok alapstatisztikái a következőképpen néznek (a károkozókra megszorítva). Az önrészes esetben az a táblázat értékeit a levonás előtt számoltuk.

	$\mathbb{E}(Y)$	$Var(Y)$	$\tilde{\mu}_3(Y)$	$\kappa(Y)$	$N$
Önrész nélkül	24450.2	$1.49 \cdot 10^9$	3.10	17.57	697
Önrésszel	40928.8	$1.68 \cdot 10^9$	1.96	8.02	383

3.3. táblázat. A kárnagyság néhány jellemzője önrésszel és önrész nélkül

Leginkább a gamma eloszlású kárnagysághoz lehetne hasonlítani, hiszen az átlagkár majdnem a kétszeresére nőtt (a gamma eloszlás esetében még nagyobb volt a növekedés), a kárszám pedig eközben majdnem a felére csökkent. A szórásnégyzet kis mértékben nőtt, a ferdeség és csúcsosság pedig csökkent, ahogy az alacsony károkat az önrész bevezetésével kivettük a rendszerből.

## 3.2. Önrész nélkül

Az önrész bevezetése előtt készítettem az első modellt, hogy az önrész okozta problémák könnyebben követhetőek/felismerhetőek legyenek. Munkám során az általánosított lineáris modellt használtam mind a kárszám, mind a kárnagyság modellezésére.

Először bemutatom a kárszám modellezésére kapott eredményeket. Mivel Poisson eloszlásból generáltam a mintát, így logikusan a modellben is ezt állítottam be háttéreloszlásnak. A modellezés során tízszeres keresztvalidációt alkalmaztam, és a mintaelemszám hetven százalékát vettem tanító halmaznak. Az egyes validációk során kétféle értékkel követtem a modellek "jóságát". Egyrészt feljegyeztem minden validációra az abszolút eltérést a

prediktált és valós érték között (*absolut error* =  $AE$ ), másrészt vettem az eltérések négyzetösszegének egyfajta "normálását". Képlettel kifejezve:

$$AE = \left| \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{y}_i \right| \quad (3.1)$$

$$b = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{(\sum_{i=1}^n y_i)^2} \quad (3.2)$$

Ezen felül a `r`-ben számolt modell megadja az  $RMSE$  és  $MAE$  értéket is. Ezen mutatószámok változását a mintanagyság függvényében láthatjuk a következő táblázatban:

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
$n = 1000$	0.2595	0.1236	4.3855	0.0469
$n = 10000$	0.2750	0.1406	9.5284	0.0043
$n = 25000$	0.2887	0.1509	29.5260	0.0016
$n = 50000$	0.2790	0.1416	16.2110	0.0009
$n = 100000$	0.2839	0.1466	68.1606	0.0004
$n = 250000$	0.2801	0.1446	97.0074	0.0002

3.4. táblázat. A kárszám modellezése különböző mintanagyságok mellett

A mintanagyság növelésével azt figyelhetjük meg, hogy az  $RMSE$  és  $MAE$  értékek tulajdonképpen konvergálnak. Az  $\overline{AE}$  a mintanagysággal párhuzamosan nő, de  $n = 50000$  esetén kiugróan alacsony értéket vesz fel. Ez a növekedés teljesen természetes, hiszen ez a mutató az abszolút eltérést adja meg, ami nagyobb elemszám esetén magától értetődően nagyobb hibát eredményez. Végül a  $\bar{b}$  által mutatott értékekről azt mondhatjuk el, hogy a mintaelemszám növelésével ezek az értékek folyamatosan csökkennek. A későbbiekben ezt a mutatót fogjuk elsősorban figyelemmel kísérni az önrész bevezetése után, mert ez ad egy jó összehasonlítási alapot a különböző modellek esetén.

A Gamma-eloszlásból készült modellek esetében alacsony minatelemszámra nem kaptunk eredményt. Az eredményeket a kárszámhoz hasonló táblá-

zatban foglaltam össze.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 25000$	107791.4	66490.96	$3.64 \cdot 10^8$	0.0297
$n = 50000$	108471.4	66460.41	$7.85 \cdot 10^8$	0.0162
$n = 100000$	106981.2	64985.03	$1.48 \cdot 10^9$	0.0069
$n = 250000$	101708.7	62035.06	$3.58 \cdot 10^9$	0.0029

3.5. táblázat. A kárnagyság modellezése különböző mintanagyságok mellett (Gamma-eloszlás)

Itt is hasonló észrevételeket tehetünk, mint a kárszám esetében. Az egyes mutatók nagyon hasonlóan viselkednek, viszont más nagyságrendekben mozgunk. Itt is az utolsó oszlopot, a  $\bar{b}$  értéket érdemes figyelni, ami a számítás módja (a "normálás") miatt már hasonló nagyságrendben mozog, mint a kárszám. A kárnagyság esetében is ugyanaz mondható el, hogy a mintaelemszám növelésével párhuzamosan, ez a mutató folyamatosan csökken.

A Lognormális-eloszlásból készült modellek esetében az alábbi eredményeket kaptam az önrész bevezetése előtt.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 1000$	338978.3	200462.7	$4.92 \cdot 10^7$	3569.3250
$n = 10000$	132289.1	69633.8	$1.83 \cdot 10^8$	0.0801
$n = 25000$	147139.7	63375.61	$4.03 \cdot 10^8$	0.0308
$n = 50000$	113101.9	59334.48	$7.86 \cdot 10^8$	0.0142
$n = 100000$	113839.2	58408.41	$1.59 \cdot 10^9$	0.0064
$n = 250000$	117303.1	58891.16	$3.98 \cdot 10^9$	0.0027

3.6. táblázat. A kárnagyság modellezése különböző mintanagyságok mellett (Lognormális-eloszlás)

Lognormális-eloszlás esetén mindegyik mintaelemszámra kaptunk eredményt, azonban  $n = 1000$  esetében elég kirívó értékeket kaptunk, túl kicsi az elemszám ahhoz, hogy jó modellt lehessen készíteni. A többi esetben tulajdonképpen nagyon hasonló eredményeket kaptunk a Gamma-eloszlás megfelelő eredményeihez.

Térjünk át a *dataOhlsson* adatra, és először nézzük meg, hogy a modellek milyen eredményt adnak az önrész bevezetése előtt. Itt össze tudjuk hason-



lítani, hogy a gamma vagy a lognormális eloszláson alapuló modell ad-e jobb eredményt.

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
Kárszám				
Poisson	0.1055	0.0212	10.4792	0.0052
Kárnagyság				
Gamma	31340.2	22601.6	$3.80 \cdot 10^8$	0.3528
Lognormális	32347.5	21575.4	$2.74 \cdot 10^8$	0.2697

3.7. táblázat. Kárszám- és kárnagyság modellek az önrész bevezetése előtt

Az eredmények alapján a kárnagyság modellezésére egyértelműen jobb választásnak tűnik a lognormális eloszlás. Bár az  $RMSE$  értékek nagyon hasonlóak, de a többi mutatóban egyértelműen jobban néz ki a lognormális modell.

### 3.3. Önrészmodellek

Az előző alfejezet eredményeit vessük össze azzal, hogy mi történik az önrész megjelenésével. Az adatok bemutatásánál már láttuk, hogy kárnagyság eloszlástól függően az "adatvesztés" különböző mértékben érinti az adatokat. A Gamma-eloszlás esetében "veszítettük" a legtöbbet, a legnagyobb mintaelemszám esetében már majdnem másfélszeres a különbség a kárszámok esetében.

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
$n = 1000$	0.1583	0.0558	4.0729	0.1338
$n = 10000$	0.1927	0.0656	12.6491	0.0107
$n = 25000$	0.1991	0.0700	20.0646	0.0042
$n = 50000$	0.1978	0.0691	20.6040	0.0022
$n = 100000$	0.1987	0.0687	31.3466	0.0011
$n = 250000$	0.1919	0.0646	72.0454	0.0005

3.8. táblázat. A kárszám modellezése különböző mintanagyságok mellett az önrész bevezetése után (Gamma eloszlású kárnagyság)

Lognormális-eloszlású károk mellett nagyon hasonló értékeket látunk, a

legnagyobb eltérések az  $RMSE$  értékekben vannak. Ennek magyarázata, hogy a gamma-eloszlás esetében az önrész bevezetésével lényegesen több adatot "veszítünk", emiatt kisebb értékeket látunk az  $RMSE$  oszlopban (kevesebb kárszámot kell prediktálni, emiatt kisebb a hibalehetőség)

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
$n = 1000$	0.2629	0.0956	3.5114	0.1046
$n = 10000$	0.2371	0.0959	12.8416	0.0067
$n = 25000$	0.2342	0.0950	17.2355	0.0029
$n = 50000$	0.2281	0.0926	28.1035	0.0015
$n = 100000$	0.2317	0.0950	38.8642	0.0007
$n = 250000$	0.2295	0.0933	55.6214	0.0003

3.9. táblázat. A kárszám modellezése különböző mintanagyságok mellett az önrész bevezetése után (Lognormális eloszlású kárnagyság)

Ha összevetjük a fenti táblázatok számait az önrész nélküli esettel, akkor egészen hasonló eredményeket láthatunk. Ennek elsődleges oka az elméleti részben már bemutatott viselkedése a Poisson-eloszlásnak, mégpedig hogy az önrész megjelenésével is megmarad az eloszlás típusa, csak a paraméter változik. Emiatt a modell kitűnően működik a kárszám esetében, hiszen nagyon egyszerűen fogalmazva csak kevesebb kárdarabszámot kell modellezni, de minden más ugyanúgy működik, mint az önrész bevezetése előtt.

A kárnagyság ebből a szempontból egy sokkal izgalmasabb kérdés, hiszen itt nem mondható el ugyanez, így itt már eltérő eredményeket várunk az önrész nélküli esethez viszonyítva. A gamma eloszlásból generált minta itt sem adott eredményt a kisebb mintanagyságokra.

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
$n = 25000$	135667.8	93831.0	$8.78 \cdot 10^8$	0.1488
$n = 50000$	134685.7	91943.1	$1.74 \cdot 10^9$	0.0626
$n = 100000$	133757.0	90898.3	$3.52 \cdot 10^9$	0.0361
$n = 250000$	131449.5	89502.1	$8.63 \cdot 10^9$	0.0152

3.10. táblázat. A kárnagyság modellezése különböző mintanagyságok mellett az önrész bevezetése után (Gamma-eloszlás)

Ha összevetjük ezeket az értékeket az önrész nélküli eset értékeivel, akkor elmondhatjuk, hogy itt már látható az eltérés. Az  $RMSE$  mutató esetében is

szembetűnő a növekedés, de a  $\bar{b}$  értékek gyakorlatilag az ötszörösére nőttek. A  $MAE$  növekedés háttérében az átlagkár növekedést lehet megemlíteni, amit az adatok bemutatásánál láttunk, hogy a gamma eloszlás esetében kimagasló volt.

A lognormális eloszlás esetén kis mintaelemszámra ( $n = 1000$ ) szintén nagyon eltérő eredményt ad, a modell tanításához itt sem volt elegendő megfigyelés, mint ahogy az önrész nélküli esetben is láttuk. Ezen kívül még az  $n = 25000$  mintaelemszám adott eltérő eredményeket, elsősorban az  $RMSE$  volt kiugróan magas.

	$RMSE$	$MAE$	$\overline{AE}$	$\bar{b}$
$n = 1000$	513329.7	359662.8	$1.74 \cdot 10^9$	4840405.0
$n = 10000$	160597.1	90621.9	$2.78 \cdot 10^8$	0.2596
$n = 25000$	283939.6	109580.5	$6.39 \cdot 10^8$	0.0790
$n = 50000$	123770.0	69726.7	$1.22 \cdot 10^9$	0.0309
$n = 100000$	128693.9	72613.9	$2.54 \cdot 10^9$	0.0154
$n = 250000$	147877.5	76243.5	$6.47 \cdot 10^9$	0.0067

3.11. táblázat. A kárnagyság modellezése különböző mintanagyságok mellett az önrész bevezetése után (Lognormális-eloszlás)

A gamma eloszlású modell eredményeivel összevetve azt mondhatjuk, hogy nagyobb elemszámok esetében alapvetően a lognormális modell produkál kisebb értékeket mind az  $RMSE$ , mind pedig a  $\bar{b}$  értékek esetében. Az eltérés nem túl nagy, de ebből arra következtethetünk, hogy egy lognormális mintára illesztett lognormális modell megbízhatóbbnak tűnik, mint egy gamma eloszlású mintára illesztett gamma modell, de ezt majd bővebben megvizsgáljuk a fejezet végén.

Összehasonlítva az önrész nélküli esettel itt is elmondhatjuk, hogy alapvetően magasabb értékeket kaptunk, de nagyságrendileg nem akkora az eltérés. Az adatok bemutatásánál és az elméleti bevezetőben láttuk, hogy a lognormális eloszlást kevésbé érinti érzékenyen az önrész bevezetése, mint a gamma eloszlást. Itt a  $\bar{b}$  érték esetében két-háromszoros eltérésekről beszélhetünk, és az  $RMSE$  értékek is többnyire közelebb vannak egymáshoz.

Hasonlóan az önrész nélküli esethez nézzük meg a *dataOhlsson* mintára az önrész megjelenése utáni eredményeket.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
Kárszám				
Poisson	0.0804	0.0110	8.9403	0.0100
Kárnagyság				
Gamma	38858.6	27997.1	$4.87 \cdot 10^8$	1.2190
Lognormális	36771.7	24913.6	$5.66 \cdot 10^8$	0.9730

3.12. táblázat. Kárszám- és kárnagyság modellek az önrész bevezetése után

A kárszám esetében hasonló eredményeket kapunk, mint a generált mintára. Ugyanazt mondhatjuk el itt is, a Poisson-eloszlás jól viseli az önrész megjelenését, a modell továbbra is jól működik. A kárnagyság esetében viszont - hasonlóan a generált adatokhoz - már lényegesen rosszabb eredményeket látunk. Ezen felül a gamma és lognormális modell közti különbség is tovább nő, az önrész megjelenésével a lognormális modell az összes általunk vizsgált mutatóban jobbnak bizonyul.

### 3.4. Eloszlások illeszkedése

A fentiekén túl megvizsgáltam még azt, hogy mi történik "félremodellezés" esetén. Azaz ha ismerem a háttéreloszlást, akkor biztosan ki lehet-e jelteni, hogy a modell paramétereinek választott ugyanaz az eloszlás minden esetben a legjobb eredményt fogja produkálni. A kárnagyságot fogom vizsgálni, és az általam generált mintában rendelkezésre áll gamma- és lognormális-eloszlásból generált kárnagyság, aminek a paramétereit minden egyes szerződés esetén a fejezet elején ismertetett módon négy változóval befolyásoltam.

Először nézzük meg az önrész nélküli esetet, azon belül is a gamma eloszlásból generált kárnagyságra ezúttal lognormális eloszláson alapuló modellt illesztettem. Mivel az előző alfejezetekben látott gamma modell nem adott eredményt, így itt is csak azokat az eseteket vizsgálom, ami összevethető a megfelelő gamma modellel.

Ha összehasonlítjuk a táblázat értékeit a gamma eloszlású modell értékeivel, akkor olyan érzésünk lehet, mintha ugyanazt a táblázatot néznénk. Gyakorlatilag az összes mutató mindegyik mintaelemszám esetében közel

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 25000$	110817.0	65577.8	$3.67 \cdot 10^8$	0.0285
$n = 50000$	112292.5	68475.1	$7.78 \cdot 10^8$	0.0153
$n = 100000$	104534.4	64648.1	$1.47 \cdot 10^9$	0.0070
$n = 250000$	101805.9	62222.2	$3.58 \cdot 10^9$	0.0029

3.13. táblázat. A gamma eloszlású kárnagyság modellezése különböző mintanagyságok mellett lognormális eloszlású modellel

ugyanazt az eredményt adja. Itt különösen érdekes lesz megfigyelni, hogy mi történik az önrész megjelenésével, melyik irányba billen el az egyensúly, vagy elbillen-e egyáltalán.

A lognormális eloszlásból származó adatokra illesztett gamma alapú modell már csak a három legnagyobb elemszámú mintára adott eredményt. Ez még a konkrét számok összevetése előtt azt mutatja, hogy a gamma eloszlású modell még kevésbé tud eredményt adni a lognormális eloszlású kárnagyság esetében.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 50000$	111161.4	59465.0	$7.91 \cdot 10^8$	0.0136
$n = 100000$	117419.2	60149.0	$1.59 \cdot 10^9$	0.0062
$n = 250000$	119944.4	59175.7	$4.01 \cdot 10^9$	0.0027

3.14. táblázat. A lognormális eloszlású kárnagyság modellezése különböző mintanagyságok mellett gamma eloszlású modellel

Itt is vessük össze az eredményeket a lognormális eloszláson alapuló modellel. Tulajdonképpen ugyanazt mondhatjuk el, mint az imént. Gyakorlatilag nem tudjuk eldönteni, hogy melyik modell a jobb, hiszen az egyes táblázatok alig térnek el egymástól. Ez így elsőre meglepő eredmény, de emlékezzünk, hogy minden szerződés esetén befolyásoltuk különböző változókkal az egyes eloszlások paramétereit, ezért nem beszélhetünk úgymond "tisza" eloszlásokról. Habár a generálás az egyes esetekben azonos eloszlásból történt, de minden rekordra más és más paraméterekkel, és ennek következtében már nehezebb dolga van a modelleknek.

Változik-e vajon valami, ha megjelenik az önrész a rendszerben? Hasonlóan végezzük el a "félremodellezést" és vessük össze az eredményeket az eddig

tapasztaltakkal. Kezdjük itt is a gamma eloszlásból származó kárnagysággal és nézzük meg, hogy mi változik.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 25000$	147688.9	98742.2	$8.77 \cdot 10^8$	0.1543
$n = 50000$	138960.1	93651.3	$1.76 \cdot 10^9$	0.0721
$n = 100000$	141607.6	95556.2	$3.50 \cdot 10^9$	0.0360
$n = 250000$	130237.1	89837.3	$8.63 \cdot 10^9$	0.0150

3.15. táblázat. A gamma eloszlású kárnagyság modellezése különböző mintanagyságok mellett lognormális eloszlású modellel az önrész bevezetése után

Gyakorlatilag pontosan azt látjuk, mint az önrész nélküli esetben, vagyis nem tudunk dönteni, az eredmények gyakorlatilag itt is megegyeznek. Az *RMSE* és *MAE* értékek alapvetően alacsonyabbak, ha a saját eloszlásával modellezzük az adatokat, de itt is akad kivétellel ( $n = 250000$  és *RMSE*). Tulajdonképpen az önrész nélküli esetben látottak után ez egy teljesen érthető eredmény, hiszen ha már ott sem sikerült egyértelműen választani, akkor az önrész bevezetésével ez csak még nehezebbé válik.

És végül nézzük meg a lognormális eloszlásból generált károkat az önrész bevezetése után, úgy hogy gamma eloszlású modellt illesztünk rá.

	<i>RMSE</i>	<i>MAE</i>	$\overline{AE}$	$\bar{b}$
$n = 50000$	131019.6	71900.6	$1.25 \cdot 10^9$	0.0336
$n = 100000$	125923.0	72137.9	$2.53 \cdot 10^9$	0.0149
$n = 250000$	146753.4	75864.3	$6.43 \cdot 10^9$	0.0065

3.16. táblázat. A lognormális eloszlású kárnagyság modellezése különböző mintanagyságok mellett gamma eloszlású modellel az önrész bevezetése után

Az eddig elhangzottakhoz semmi újdonságot nem lehet hozzáfűzni. Pontosán azt látjuk, mint az összes eddigi esetben. Az értékek itt is megfeleltethetők egymásnak, nem lehet egyértelműen állást foglalni. Az egyetlen dimenzió, ahol a különbség mindig egyértelmű volt, és még nem beszéltünk róla, az a mintaelemszám. Mint láttuk, gamma eloszlású modell esetén a kisebb elemszámokra nem kapunk eredményt, míg a lognormális modell minden esetben működött. És bár ez nem igazán kapcsolódik szorosan a "félremodellezéshez", de tekinthetjük úgy, hogy a lognormális és gamma eloszlású

modellek ilyen feltételezések mellett lényegében nem különböztethetők meg. Vagy az általuk adott eredmények gyakorlatilag azonosak, azonban a lognormális modell működése stabilabb, kisebb elemszámra is működik, ezért ebből a szempontból egy jobb választásnak tűnik. Ez a vizsgálat csak felszínesen érintette magát a problémát, éppen ezért még sok lehetőség rejlik a további kutatásokban.

## 4. fejezet

# Összefoglalás

A szakdolgozat célja az önrészmodellek vizsgálata volt. Először az elméleti bevezetőben láthattuk egyszerűbb példákon, hogy az önrész hogyan befolyásolja a különböző eloszlásokból származó károkat. Különböző önrésztípusok esetén levezettük nevezetes eloszlásokra az eloszlás- és sűrűségfüggvényeket. Három kárnagyság eloszlás esetében (gamma, lognormális és Pareto) pedig alaposabban megismerhettük az önrész okozta változásokat. A tapasztalatok alapján azt láttuk, hogy a gamma eloszlást érintette a "legérzékenyebben" az önrész megjelenése, arányaiban az ő esetében történtek a legnagyobb változások a különböző mutatók tekintetében.

Komplexebb adatokon az általánosított lineáris modell segítségével készítettünk különböző modelleket. Az önrész bevezetése előtt készítettük az első modelleket generált és valós adatokra egyaránt. A generált minta esetében a gamma és lognormális modellek hasonló eredményekre vezettek, a legszembetűnőbb különbség a mintaelemszám esetében volt megfigyelhető. A lognormális eloszlás alapú modell kis mintaelemszámra is adott eredményt, mint a gamma modellnek nagyobb mintára volt "szüksége". Valós adatokon már a lognormális modell jobbnak bizonyult.

Az önrész bevezetése után minden esetben csökkent a megfigyeléseink száma, csonkított mintával dolgoztunk. Hasonló tapasztalatokat szereztünk, mint az elméleti bevezetőben egyszerű esetekre. A generált adatokra is a gamma eloszlást érintette a legérzékenyebben az önrész megjelenése. Hason-



ló modellekkel vizsgáltuk az önrészes mintákat, mint az eredeti adatokat. A kárszámokat Poisson-eloszlásból generáltuk, és mint ahogy az elméleti részben láttuk, az önrész megjelenésével csak a paraméter változott meg. Emiatt az önrész megjelenése után a kárszám modellezésével nem adódott probléma, nagyon hasonló eredményeket kaptunk, mint az önrész bevezetése előtt. A kárnagyság esetében ugyanez már nem mondható el, az eredményeink lényegesen rosszabbak lettek, de igazából ezt is vártuk, hiszen gamma és lognormális eloszlás esetében az önrész megjelenésével megváltozik az eloszlás (nem úgy, mint a Poisson-eloszlásnál).

Végül megnéztük, hogy mit lehet mondani "félremodellezés" esetében. Ehhez a gamma- és lognormális eloszlásból generált mintákat tulajdonképpen egymással modelleztük. Itt is külön néztük az önrész nélküli és önrészes adatokat. Az általunk használt mutatók segítségével nem lehetett egyértelműen győztest választani, ami további kutatási kérdésekre ad lehetőséget. Érdekes lehetne több, különféle mutató alapján az összehasonlítást elvégezni (pl. logisztikus score).

Munkám során azt láttam, hogy az önrész megjelenésével felmerülő különböző problémáknak tulajdonképpen csak egy kis szegletébe nyertem betekintést, és rengeteg kutatási lehetőséget rejt a téma. Mind a statisztika, mind a valószínűségi számítás terén rengeteg irányban lehet vizsgálni az önrész okozta kihívásokat. Mint említettem, akár új mutatók/mérőszámok bevezetésével, akár több, különböző modellel megvizsgálni az eredményeket további kutatásra ad lehetőséget.

# Irodalomjegyzék

- [1] Douglas G Chapman. Estimating the parameters of a truncated gamma distribution. *The Annals of Mathematical Statistics*, pages 498–506, 1956.
- [2] Gail Gong and Francisco J Samaniego. Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, pages 861–869, 1981.
- [3] Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.
- [4] Gee Y Lee. General insurance deductible ratemaking. *North American Actuarial Journal*, 21(4):620–638, 2017.
- [5] Arató Miklós. Nem-élet biztosítási matematika. *ELTE Eötvös Kiadó*, 2001.
- [6] Jostein Paulsen and Knut Stubø. On maximum likelihood and pseudo-maximum likelihood estimation in compound insurance models with deductibles. *ASTIN Bulletin: The Journal of the IAA*, 41(1):1–28, 2011.

# Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Arató Miklósnak, aki szakértelmével, tanácsaival, türelmével és precizitásával segítette munkámat. Ezen kívül szeretném megköszönni szaktársaimnak a segítséget, amit az egyetemi évek alatt nyújtottak, és szeretteimnek, hogy végig kitartóan támogattak tanulmányaimban.