

Eötvös Loránd Tudományegyetem
Természettudományi Kar



Budapesti Corvinus Egyetem
Közgazdaságtudományi Kar



Egyéni modellek a tartalékolásban

Készítette: Knódel Máté János
Biztosítási és Pénzügyi Matematika MSc
Aktuárius szakirány

Témavezető: Arató Miklós

2016. május

Köszönetnyilvánítás

Szeretnék köszönetet mondani konzulensemnek, Arató Miklósnak a témaválasztásban nyújtott segítségéért, valamint a szakdolgozat elkészítésében nyújtott támogatásáért, és szakmai tanácsaiért. Emellett szeretném megköszönni családomnak, hogy tanulmányaim alatt mindvégig, mindenben támogattak, és kitartottak mellettem a nehéz időszakokban is. Végül, de nem utolsó sorban szeretném megköszönni az összes barátomnak és csoporttársamnak az egyetemi éveket, és a sok vidámságot, amit együtt megéltünk.

Tartalomjegyzék

1. Bevezetés	4
2. A vizsgált modell	6
2.1. A modell változói, jelölések	6
2.1.1. A függőkárok típusai	6
2.1.2. Időváltozók	6
2.1.3. Kárfejlődési folyamat	7
2.1.4. Kárbejelentési és kárbekövetkezési intenzitás	8
2.2. A likelihood függvény	9
2.3. A függőkárok tartalékának becslése	10
3. Többdimenziós Pareto-eloszlás	14
3.1. Kétdimenziós Pareto-eloszlás	14
3.2. Többdimenziós Pareto-eloszlás	16
3.2.1. Peremeloszlás	16
3.2.2. Feltételes eloszlás	17
3.2.3. Paraméter becslés	20
3.2.4. A tartalékok meghatározásához szükséges állítások	24
4. Az adatok jellemzése	39
5. Paraméterbecslés	41
5.1. A kárfejlődési vektor paramétereinek becslése	41
5.2. Az időváltozók paramétereinek becslése	41
5.2.1. Bejelentési késés	42
5.2.2. Első fizetési késés	43
5.2.3. Kifizetések száma az első kifizetés után	44
5.2.4. Kifizetési késlekedések	45
5.3. A kárbekövetkezési intenzitás becslése	45
6. A függőkárok tartalékának becslése, összehasonlítás	47
6.1. A függőkárok tartalékának becslése	47
6.2. Tényadatok és becslés lánc-létra módszerrel	47
6.3. Összehasonlítás, az eltérések magyarázata	48
7. Összegzés, további vizsgálati lehetőségek	50
Hivatkozások	52

1. Bevezetés

A nem-életbiztosítók számára kardinális kérdés, hogy a jövőben várható kárkifizetéseiket minél pontosabban előre tudják jelezni. Napjainkban ez a téma még nagyobb aktualitást élvez, hiszen a Szolvencia II. és az IFRS 4 keretrendszerek bevezetése miatt minden biztosító köteles egy olyan modellt kidolgozni, amely a jövőbeli kötelezettségeit a lehető legpontosabban megragadja.

A függőkárok tartalékolásának témájában rengeteg cikk és modell született az elmúlt közel 20-25 évben, amelyek között vannak széles körben elterjedt, és kevésbé ismert módszerek is. Ezen dolgozat célja az, hogy a hagyományos ösvényről, a klasszikus keretrendszerektől eltérve egy egyéni, sztochasztikus kártartalékolási módszeren alapuló modellt mutasson be.

Az irodalomban a legtöbb modell aggregált kárkifizetési adatokkal dolgozik, azaz egy szerződés csoport adatait vizsgálja. A két legismertebb és legelterjedtebb módszer, amelyben az aggregált adatok alapján, kárkifizetési háromszögek segítségével számítható ki a tartalékszükséglet, a lánc-létra (chain-ladder, CL) módszer, valamint a Bornhuetter-Ferguson módszer (Bornhuetter és Ferguson (1972) cikke). A lánc-létra módszernél a jövőbeni kifizetések sztochasztikus viselkedését is vizsgálták, egy bővebb összefoglaló található Mack (1993) és Mack (1999) úttörő cikkeiben. Az eljárások előnye, hogy viszonylag kevés feltételezés alapján, a gyakorlatban is nagyon jó illeszkedést mutató becsléseket adnak.

A fent bemutatott modellek mind aggregált kárkifizetési adatokkal dolgoznak, de jogosan merül(t) fel a kérdés: kaphatunk-e pontosabb becslést az összkárkifizetésekre, ha egyéni szinten végezzük az előrejelzést? A (többdimenziós) statisztika 20. és 21. századi rohamos fejlődése lehetővé tette, hogy az újonnan kidolgozott módszereket felhasználjuk és bevonjuk az elemzésbe, és ezáltal jobb előrejelzést kapjunk a várható kárkifizetésekre, tartalékszükségletre. Emellett az informatika fejlődése is egy lökést adhat az új tartalékszámítások elterjedésének. Amikor a fenti bekezdésekben bemutatott módszerek alapjait lefektették, akkor jóval kisebb (akár semmilyen) számítógépkapacitással rendelkeztek a biztosítók, így fontos volt, hogy olyan módszereket dolgozzanak ki, amelyek zárt képletekkel, gyorsan számolhatók. Manapság azonban ezen problémák a modern informatika révén könnyebben áthidalhatók, így az egyéni módszereken alapuló kártartalékolási modellek is teret nyerhetnek.

Az első, folytonos idejű, egyéni kártartalékolási modellek Arjas (1989) és Norberg (1993) nevéhez fűződnek. Haastrup és Arjas (1996) cikkükben egy folytonos idejű, egyéni modellt mutatnak be, ahol a kárbekövetkezési intenzitást egy sztochasztikus folyamattal

jellemzik, majd ezek segítségével megadják a függőkárok eloszlását is.

A folytonos időkeretet elhagyva diszkrét időben építkezik a Pigeon et al. (2013) cikk, amelyhez hasonló keretrendszerben vizsgálom majd az általam felépített modellt is. Pigeon et al. (2014) cikkében az előző modell egy módosított változata található, ahol a kárkifizetések helyett a bekövetkezett károk adatai alapján is becsülhető az összkár eloszlása.

Dolgozatom felépítése a következő: Pigeon et al. (2013) cikke alapján bemutatom az általam vizsgált modellt, majd önálló, analitikus eredményeket adok a függőkárok várható értékére, az idézett cikktől eltérő többdimenziós eloszlást feltételezve. Ez után bevezetem a kárfejlődési vektor eloszlására használt többdimenziós, európai típusú Pareto-eloszlást, megvizsgálom ennek alapvető tulajdonságait, majd kimondom és bizonyítom a modellemhez szükséges állításokat. A dolgozat második felében bemutatom az elemzéshez használt adatokat, majd a maximum likelihood módszer segítségével becslém az eloszlások ismeretlen paramétereit. Végül az önálló eredmények felhasználásával meghatározom a függőkárok tartalékszükségletét, valamint megvizsgálom a modellem illeszkedését az adathalmazon.

2. A vizsgált modell

Az alábbi fejezetben az általam vizsgált, diszkrét időkeretű, egyéni káradatokon alapuló sztochasztikus kártartalékolási modellt mutatom be. Szakdolgozatom, illetve ezen fejezet alapjául M. Pigeon et al. (2013) cikke szolgált, így a jelölésekben is igyekeztem az eredeti struktúrát követni.

A fejezet első részében a hivatkozott cikk nyomán bemutatom a modellt, és annak változóit, majd a második alfejezetben felírom a minta likelihood függvényét. Ebből meghatározhatóak a változók eloszlásának ismeretlen paraméterei. A harmadik alfejezetben pedig önálló eredményként analitikus formulákat adok a függőkárok tartalékszükségletének becslésére.

2.1. A modell változói, jelölések

2.1.1. A függőkárok típusai

A modell diszkrét időben dolgozik, azaz a károk bekövetkezését, és további fejlődését diszkrét periódusokban (pl. egy éves időtartamokban) vizsgálja. Az analízis során a függőkárokat három kategóriába soroljuk: IBNR (bekövetkezett, de még be nem jelentett) károk, RBNP (bejelentett károk, amelyekre még nem történt kifizetés) károk, illetve RBNS (bejelentett károk, amelyekre már történt kifizetés, de még nem lezártak) károk. Ha egy kárra egy adott periódusban több kifizetés is történt, akkor ezeket aggregálom, mivel csak a kárkifizetési periódusok közt akarom vizsgálni a kumulált kifizetés növekedését.

2.1.2. Időváltozók

A modell felépítését az egyéni károk szintjén kezdjük. Egy adott kár a következő adatokkal írható le: bekövetkezési idő, bejelentési késlekedés, a kárkifizetések dátuma és nagysága, illetve a lezárási dátum. Ezek számszerűsítéséhez vezessük be a következő (diszkrét) jelöléseket: legyen az i . periódus k . kára (ik), továbbá $i = 1, \dots, I$, ahol I a vizsgált periódusok száma, valamint $k = 1, \dots, K_i$, ahol K_i az i . periódusban bekövetkezett károk száma. A károkat jellemző diszkrét időváltozók pedig legyenek a következők:

- jelölje T_{ik} az (ik) kár *bejelentési késését*, azaz a bekövetkezés és a bejelentés periódusai közt eltelt periódusok számát
- jelölje Q_{ik} az (ik) kár *első fizetési késését*, azaz a bejelentés és az első kifizetés periódusai közt eltelt periódusok számát

- jelölje U_{ik} azon periódusok számát, amelyben történt pozitív kárkifizetés az első után
- jelölje N_{ikj} az (ik) kár j . és $j + 1$. (aggregált) kárkifizetése közt eltelt periódusok számát, azaz a két kifizetés közti késlekedését $j = 0, \dots, U_{ik}$ -ra. $j = U_{ik} + 1$ -re pedig N_{ikj} jelölje a kár lezárási periódusát. Továbbá legyen $N_{ik} := \sum_{j=1}^{U_{ik}+1} N_{ikj}$ az első kifizetés és a lezárás közt eltelt periódusok száma.

A fent definiált változók diszkrét eloszlásúak, eloszlásuk és eloszlásfüggvényük rendre $f_1(t; \nu), f_2(q; \psi), f_3(u; \beta), f_4(n; \Phi)$ és $F_1(\cdot; \nu), F_2(\cdot; \psi), F_3(\cdot; \beta), F_4(\cdot; \Phi)$.

A modell és a diszkrét változók jobb megértését illusztrálja a következő példa. Tekintsünk egy kárt, amely 2010. május 13-án következett be, és 2010. június 8-án jelentették be a biztosító társaságnak. Így a bejelentés éve (első periódus) 2010, a bejelentési késlekedés $t_{ik} = 0$. Az első kárkifizetés 2010. december 16-án történt meg, így az első fizetési késés $q_{ik} = 0$. Ezt követően még 7 kárkifizetés volt, melynek dátumai: 2011. január 19., 2011. február 19., 2011. október 27., 2012. április 19., 2012. július 6., 2013. március 4. és 2014. január 22. A kárt 2014. január 30-án zárták le. Összegezve az adott években a károkat, az első kifizetés után még 4 évben (periódusban) fizettünk ki kárt, azaz $u_{ik} = 4$, valamint minden aggregált kifizetés között 1 periódus telt el, azaz $n_{ik1} = n_{ik2} = n_{ik3} = n_{ik4} = 1$, valamint az utolsó kifizetés a lezárással megegyező periódusban történt, így $n_{ik5} = 0$.

2.1.3. Kárfejlődési folyamat

Az előző alfejezetben bemutattam az adott kárt jellemző változókat, míg az alábbi részben annak időbeni fejlődését írom le. Jelölje $Y_{ikj} (> 0)$ a j . részki fizetést a k . kárra ($k = 1, \dots, K_i$) az i . periódusból ($i = 1, \dots, I$). Egy adott kárra a kumulált kárkifizetést ezen részki fizetések összegeként kapjuk meg. Legyen az (ik) -dik kárt a j . és a $j + 1$. kárkifizetési periódusok közt jellemző $\lambda_j^{(ik)}$ növekedési faktor:

$$\lambda_j^{(ik)} = \frac{\sum_{r=1}^{j+1} Y_{ikr}}{\sum_{r=1}^j Y_{ikr}}, \quad (1)$$

valamint a szigorúan pozitív $U_{ik} = u_{ik}$ mellett az $u_{ik} + 1$ hosszúságú $\Lambda_{u_{ik}+1}^{(ik)}$ vektor a kár fejlődési mintája:

$$\Lambda_{u_{ik}+1}^{(ik)} = \left[Y_{ik1} \quad \lambda_1^{(ik)} \quad \dots \quad \lambda_{u_{ik}}^{(ik)} \right]^T \quad (2)$$

Ha az első kifizetést nem követte további, azaz $u_{ik} = 0$, akkor a kár fejlődési mintája az egy elemű Y_{ik1} vektor.

A fent bemutatott fejlődési minta hasonlít a jól ismert lánc-létra modellnél megismert változatra. Fontos azonban megemlíteni egy lényeges különbséget: az utóbbi, klasszikus modell azt vizsgálja a növekedési faktorokban, hogy az egyes periódusok között hányszorosára változott a kumulált kárkifizetés. Ezzel szemben a fent bemutatott kárfejlődési vektor azt ragadja meg, hogy egy *újabb részki fizetés során* hányszorosára nőtt a kumulált kárkifizetés. Más szavakkal, azon periódusok közt, ahol történt kifizetés, hányszorosára nőtt az összkifizetés. Ezt nevezik *payment-to-payment* alapú szemléletnek.

A másik lényeges különbség, hogy a sztochasztikus lánc-létra modellnél a növekedési faktorok függetlenek a múlttól, sőt a kezdeti kárkifizetés is független a növekedési faktortól is. Ez a feltevés azonban a valóságban sérülhet, így a vizsgálat során érdemes egy többdimenziós megközelítést alkalmazni a $\Lambda_{u_{ik}+1}^{(ik)}$ vektor eloszlására. Ehhez tekintsük a következő definíciót Mardia (1962) cikke alapján:

2.1. definíció. Az (X_1, X_2, \dots, X_k) vektor k -dimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ elhelyezkedés és $p > 0$ lecsengés paraméterekkel, ha az együttes sűrűségfüggvény:

$$MP(x_1, x_2, \dots, x_k) = \begin{cases} \frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k}} & x_i > a_i > 0 \\ 0 & \text{egyébként} \end{cases} \quad (3)$$

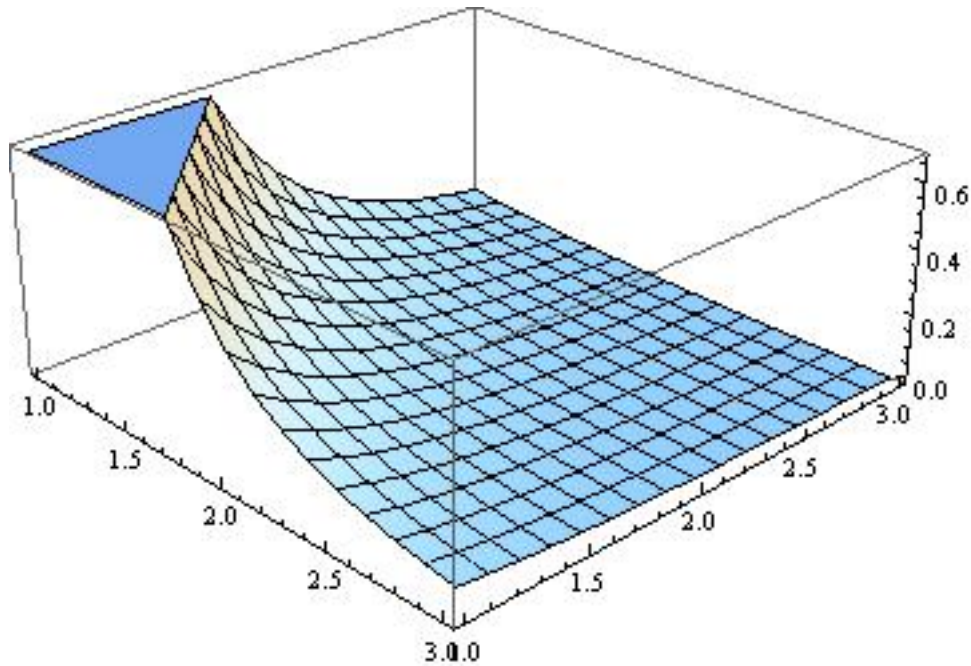
Az előző definícióban bevezetett eloszlást mutatja be az 1. ábra $\mathbf{a} = (1, 1)$ elhelyezkedés és $p = 2$ lecsengés paraméterekkel:

A fent bemutatott $\Lambda_{u_{ik}+1}^{(ik)}$ kárfejlődési vektor együttes eloszlását a 2.1. definícióban bevezetett többdimenziós Pareto-eloszlásúnak feltétezem a modellben. Ezen eloszlás részletes tulajdonságait a 3. fejezetben tárgyalom. Ott megvizsgálom a peremeloszlásokat, a feltételes eloszlást, illetve kitérek a paraméterek maximum likelihood becslésére is.

2.2. Megjegyzés. *M. Pigeon et al. (2013) cikkében a többdimenziós Pareto-eloszlás helyett egy rugalmas, többváltozós, ferde normális eloszlást alkalmaz a kárfejlődési vektor jellemzésére.*

2.1.4. Kárbejelentési és kárbekövetkezési intenzitás

Az előbbi alfejezetekben bevezetett változók segítségével egy kár jellemzése már lehetségessé vált, viszont szükséges az is, hogy becslést tudjunk adni az adott periódusban bekövetkezett károk számára. Ezt az adott periódusban kockázatban álló szerződések számával becsljük: tegyük fel, hogy az i . periódusban bekövetkezett károk száma, K_i Poisson eloszlású $\Theta w(i)$ paraméterekkel. Itt Θ jelöli a Poisson-folyamat intenzitását, $w(i)$



1. ábra. 2-dimenziós Pareto eloszlás sűrűségfüggvénye $\mathbf{a} = (1, 1)$ elhelyezkedés és $p = 2$ lecsengés paraméterekkel. Forrás: saját ábra

pedig az i . periódusban kockázatban lévő szerződések számát. Jelöljük t_{ik}^* -gal az értékelés periódusát, valamint a vizsgálatok kezdőperiódusának tekintsük az első periódust. Mivel csak a tényadatokat (azaz a bejelentett károkat) ismerjük, így a bekövetkezett, de még be nem jelentett (azaz IBNR) károk Poisson eloszlásúak $\Theta w(i) (1 - F_1(t_i^* - 1; \nu))$ paraméterrel. Így erre az időszakra az IBNR károk várható darabszáma $\Theta w(i) (1 - F_1(t_i^* - 1; \nu))$.

2.2. A likelihood függvény

A károkat jellemző, a fentiekben bemutatott változók (pl. bejelentési késlekedés) ismeretlen paramétereinek meghatározását maximum likelihood módszerrel végzem. A likelihood függvény a vizsgált kártípustól függően három részre osztható: a lezárt károk, az RBNP és az RBNS károk likelihood tagja. A továbbiakban t_{ik}^* jelöli az értékelés időpontját, valamint u_{ik}^* az értékelés pillanatában, az első kifizetés után ismert kifizetések darabszámát. A károk értékelését mindig az értékelő periódus (pl. az adott év) végén végezzük el.

Lezárt károk A lezárt károk (Cl) likelihood tagja a következőképpen adható meg:

$$\begin{aligned}
\mathcal{L}^{Cl} \propto & \prod_{(ik)_{Cl}} MP(\Lambda_{u_{ik}+1}; \mathbf{a}_{u_{ik}+1}; p|u_{ik}) \\
& \cdot \prod_{(ik)_{Cl}} f_1(t_{ik}; \boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1) \cdot f_2(q_{ik}; \boldsymbol{\psi}|Q_{ik} \leq t_{ik}^* - t_{ik} - 1) \\
& \cdot \prod_{(ik)_{Cl}} f_3(u_{ik}, \boldsymbol{\beta}|U_{ik} \leq t_{ik}^* - q_{ik} - t_{ik} - 1) \\
& \cdot \prod_{(ik)_{Cl}} I(u_{ik} = 0)(1) + I(u_{ik} = 1)f_4(n_{ik1}, \boldsymbol{\Phi}|0 < N_{ik1} \leq t_{ik}^* - q_{ik} - t_{ik} - u_{ik}) + \\
& \quad + I(u_{ik} > 1)f_4(n_{ik1}, \boldsymbol{\Phi}|0 < N_{ik1} \leq t_{ik}^* - q_{ik} - t_{ik} - u_{ik}) \cdot \\
& \quad \cdot \prod_{j=2}^{u_{ik}} f_4 \left(n_{ikj}, \boldsymbol{\Phi}|0 < N_{ikj} \leq t_{ik}^* - q_{ik} - t_{ik} - (u_{ik} - j + 1) - \sum_{p=1}^{j-1} n_{ikp} \right).
\end{aligned}$$

A likelihood első tagja adja a fejlődési minta többdimenziós eloszlását, míg a többi tag a bejelentési, az első kárkifizetési késlekedésre, a kárkifizetési periódusok számára és a két ilyen periódus közti késlekedésre vonatkozik.

RBNS károk Az RBNS károk likelihood tagja a következőképpen adható meg:

$$\begin{aligned}
\mathcal{L}^{RBNS} \propto & \prod_{(ik)_{RBNS}} MP(\Lambda_{u_{ik}^*+1}; \mathbf{a}_{u_{ik}^*+1}; p|u_{ik}^*) \\
& \cdot \prod_{(ik)_{RBNS}} f_1(t_{ik}; \boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1) \cdot f_2(q_{ik}; \boldsymbol{\psi}|Q_{ik} \leq t_{ik}^* - t_{ik} - 1) \\
& \cdot \prod_{(ik)_{RBNS}} (1 - F_3(u_{ik}^* - 1, \boldsymbol{\beta})) \\
& \cdot \prod_{(ik)_{RBNS}} I(u_{ik}^* = 0)(1) + I(u_{ik}^* = 1)f_4(n_{ik1}, \boldsymbol{\Phi}|0 < N_{ik1} \leq t_{ik}^* - q_{ik} - t_{ik} - u_{ik}^*) + \\
& \quad + I(u_{ik}^* > 1)f_4(n_{ik1}, \boldsymbol{\Phi}|0 < N_{ik1} \leq t_{ik}^* - q_{ik} - t_{ik} - u_{ik}^*) \\
& \quad \cdot \prod_{j=2}^{u_{ik}^*} f_4 \left(n_{ikj}, \boldsymbol{\Phi}|0 < N_{ikj} \leq t_{ik}^* - q_{ik} - t_{ik} - (u_{ik}^* - j + 1) - \sum_{p=1}^{j-1} n_{ikp} \right).
\end{aligned}$$

RBNP károk Az RBNP károk likelihood tagja a következőképpen adható meg:

$$\mathcal{L}^{RBNP} \propto f_1(t_{ik}; \boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1)(1 - F_2(t_{ik}^* - t_{ik} - 1; \boldsymbol{\psi}))$$

2.3. A függőkárok tartalékának becslése

Miután a likelihood függvény maximalizálásából megkaptuk a paraméterek becslését, ezek segítségével már tudjuk becsülni a várható kifizetést az IBNR, RBNP, illetve RBNS károkra. Az adott alfejezetben kártípusonként analitikus formulákat adok az összkifizetés

várható értékére, valamint az IBNR és RBNP károkra megadom az összkifizetés második momentumát is. Az egyszerűsítés végett az (ik) kárindexet most elhagyom, azaz a bemutatott tételek egy adott kárt jellemeznek.

2.3. Tétel. *[egy IBNR vagy RBNP kár első és második momentumai] Jelölje C egy IBNR (vagy RBNP) kár összértékét, azaz*

$$C = Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_U. \quad (4)$$

Tegyük fel, hogy rögzített U esetén a Λ_{U+1} kárfejlődési vektor a 2.1. definícióban megismert többdimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_{U+1})$ elhelyezkedés és p lecsengés paraméterekkel. Ha teljesül a $p > U + 1$ feltétel, akkor C várható értékét az U (azaz az első kifizetés utáni, kifizetéses periódusok számára) függvényében a következő kifejezés adja:

$$\mathbb{E}_U(C) = \mathbb{E}_U \left\{ \sum_{i=0}^U \binom{U}{i} \frac{a_1 \dots a_{U+1} \cdot p(p+1) \dots (p+U)}{(p+U)(p+U-1) \dots (p+1-i)(p-1-i)} \right\}, \quad (5)$$

$p \leq U + 1$ esetén a szorzat várható értéke végtelen.

Ha teljesül a $p > 2(U + 1)$ feltétel is, akkor C második momentumát a következő kifejezés adja meg:

$$\mathbb{E}_U(C^2) = \mathbb{E}_U \left\{ \sum_{i=0}^{2U} F_i \frac{a_1^2 \dots a_{U+1}^2 \cdot p(p+1) \dots (p+U)}{(p+U)(p+U-1) \dots (p+1-i)(p-2-i)} \right\}, \quad (6)$$

ahol

$$F_i = \sum_{s=\max(0, i-U)}^{\lfloor \frac{i}{2} \rfloor} \binom{U}{s} \binom{U-s}{i-2s} 2^{i-s} \quad (7)$$

$p \leq 2(U + 1)$ esetén pedig a szorzat második momentuma végtelen.

2.4. Megjegyzés. *A valóságban tetszőlegesen nagy U esetén a fenti várható érték végtelen lenne, így a gyakorlati alkalmazás során érdemes korlátozni az U értékét annak érdekében, hogy véges várható értéket kapjunk.*

Az RBNS károknál azt is figyelembe kell vennünk, hogy a fejlődési minta egy részét már megfigyeltük. Tegyük fel, hogy már $0 < m (< U + 1)$ periódusban történt kifizetés az adott kárra (beleértve az első kifizetést is). Jelölje A a már megfigyelt kifizetések halmazát, B a hátralévő kifizetéseket, azaz $\mathbf{\Lambda}_A = [Y_1 \ \lambda_1 \ \dots \ \lambda_{m-1}]^T$ és $\mathbf{\Lambda}_B = [\lambda_m \ \dots \ \lambda_U]^T$, így $\mathbf{\Lambda} = [\mathbf{\Lambda}_A \ \mathbf{\Lambda}_B]^T$.

2.5. Tétel. [egy RBNS kár első és második momentuma] Jelölje C egy RBNS kár összértékét, azaz

$$C = Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_U. \quad (8)$$

C értéke az ismert kifizetések mellett:

$$[C | \mathbf{\Lambda}_A = \boldsymbol{\ell}_A] = y_1 \cdot \ell_1 \cdot \dots \cdot \ell_{m-1} \cdot \lambda_m \cdot \dots \cdot \lambda_U, \quad (9)$$

Ekkor, ha a Λ_{U+1} kárfejlődési vektor a 2.1. definícióban megismert többdimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_{U+1})$ és p paraméterekkel, akkor C feltételes várható értékét és feltételes második momentumát (a $\mathbf{\Lambda}_A$ -ra nézve) a következő kifejezések adják:

$$\begin{aligned} \mathbb{E}(C | \mathbf{\Lambda}_A = \boldsymbol{\ell}_A) &= y_1 \cdot \ell_1 \cdot \dots \cdot \ell_{m-1} \cdot \mathbb{E} \{ \lambda_m \cdot \dots \cdot \lambda_U \} = \\ &= y_1 \cdot \ell_1 \cdot \dots \cdot \ell_{m-1} \cdot \\ &\cdot \mathbb{E} \sum_{i=0}^{U-m} G_i \cdot \frac{a'_{m+1;m} \cdot \dots \cdot a'_{U+1;m} \cdot (p+m) \cdot \dots \cdot (p+U)}{(p+U) \cdot \dots \cdot (p+m-i)} \cdot \left(\frac{a_{m+1}}{a'_{m+1;m}} + \frac{1}{p+m-1-i} \right), \end{aligned} \quad (10)$$

ahol az $a'_{:,l}$ paraméterek a következő rekurzív képletből kaphatóak meg minden $1 \leq l \leq m$ -dik lépésre:

$$\begin{aligned} a'_{i,l} &= \frac{\lambda_l + D_{l-1}}{a'_{i;l-1}} a'_{i;l-1} \quad , \quad \text{ahol } i = l+1, \dots, m \quad \text{és} \\ C_l &= \frac{\lambda_l}{\lambda_l + D_{l-1}} + \frac{a'_{i;l-1}}{\lambda_l + D_{l-1}} C_{l-1} \quad \text{és} \\ D_{l-1} &= a'_{i;l-1} \left(\sum_{i=l+1}^{U+1} \frac{a_i}{a'_{i;l-1}} + C_{l-1} \right) \end{aligned}$$

továbbá az $a'_{i,0} = a_i$, $C_0 = -U$, $D_0 = 0$ és $\lambda_0 = Y_1$ értékek a kezdő paraméterek, valamint a G_i együtthatók a következő módszerrel kaphatóak meg minden $i = 0, \dots, U-m$ indexre:

1. Vegyük az $\mathcal{A} = \left(\frac{a_{m+2}}{a'_{m+2}}, \dots, \frac{a_{U+1}}{a'_{U+1}} \right)$ halmaz összes $U-m-i$ elemű részhalmazát. Ha az \mathcal{A} halmaz üres, akkor legyen $G_i = 1$.
2. Számítsuk ki minden egyes részhalmazon belül az elemek szorzatát.
3. A kapott szorzatokat összegezzük, és jelöljük az így kapott együtthatót G_i -vel.

2.6. Megjegyzés. Az RBNS károk várható összkifizetését leíró formula elég bonyolult, mivel ki kell számítani hozzá a rekurzív a'_i paramétereket, illetve az analitikusan nem számolható G_i együtthatókat. Ez jelentős mértékben megnehezíti a gyakorlatban ezen kártípus összkifizetésének várható érték számítását. Az RBNS károk második momentumát leíró képletek még összetettebbek lennének, így azokra nem térek ki.

A függőkárok összkifizetését leíró állítások bizonyításai a 3. fejezetben találhatóak. Ott egy általános, k -dimenziós Pareto-eloszlásra határozom meg a koordináták szorzatának feltételes, illetve feltétel nélküli várható értékét és második momentumát. Az ott kapott eredményeket alkalmazva $k = U + 1$ esetre megkapjuk az állításokban szereplő képleteket. A 2.3. tétel bizonyításához lásd a 3.10. és a 3.13. állításokat, a 2.5. tétel bizonyításához pedig a 3.15. állítást.

A következő tétel egyfajta összegzés: az előző tételekben bemutattam kártípusonként az egy kárra várható összkifizetést. Ennek segítségével pedig az összes kárra vonatkozó összkárfizetés az alábbiak szerint számítható:

2.7. Tétel (becslés az IBNR, RBNP és RBNS károk összkifizetésére). *Jelölje \mathcal{I} a károkra az adathalmazból megfigyelt információt. Így a 2. fejezetben ismertett modellben az IBNR, RBNP és RBNS károk várható értékét a következőképpen kapjuk meg: Az IBNR károk várható értéke:*

$$\mathbb{E}[IBNR | \mathcal{I}] = \mathbb{E}(K_{IBNR}) \cdot \mathbb{E}_U(C_{IBNR}), \quad (11)$$

ahol az IBNR károk várható darabszáma ($\mathbb{E}(K_{IBNR})$) a Poisson-eloszlásból kapható meg, valamint a teljes kárkifizetés egy IBNR kárra (C_{IBNR}) a 2.3. tétel alapján számolható.

Az RBNP károk várható értéke:

$$\mathbb{E}[RBNP | \mathcal{I}] = k_{RBNP} \cdot \mathbb{E}_U(C_{RBNP}), \quad (12)$$

ahol az RBNP károk várható darabszáma (k_{RBNP}) adott, valamint a teljes kárkifizetés egy RBNP kárra (C_{RBNP}) a 2.3. tétel alapján számolható.

Az RBNS károk várható értéke:

$$\mathbb{E}[RBNS | \mathcal{I}] = \sum_{(ik)_{RBNS}} \mathbb{E}_U(C_{ik} | \Lambda_A^{ik} = \ell_A^{ik}) \quad (13)$$

ahol a szumma végigfut az RBNS károkon, és egy adott RBNS kár teljes várható kifizetése (a már ismert $\Lambda_A^{ik} = \ell_A^{ik}$ kifizetések alapján) a 2.5. tétel alapján számolható.

3. Többdimenziós Pareto-eloszlás

A kárnagyságok, különösen az extrém nagy károk eloszlásának vizsgálatára különösen jól használhatóak a Pareto-típusú eloszlások, mivel a tapasztalatok szerint gyakran nagyon jó illeszkedést mutatnak. Ennek oka, hogy a farokvalószínűség csak polinomiális lecsengésű, nem pedig exponenciális (mint például a gamma eloszlásnál), így Pareto-eloszlás esetén a nagy károk bekövetkezésének valószínűsége nagyobb, mint exponenciális farokvalószínűségű eloszlások esetén.

Ezek alapján jogosan merült fel az igény, hogy ezen eloszlástípusnak is szülessenek többdimenziós általánosításai. Az egyik legismertebb általánosítás Mardia (1962) cikkében található: az 1-es típusú, többdimenziós Pareto-eloszlás az európai típusú Pareto-eloszlás általánosítása, melynek peremeloszlásai 1-dimenziós Pareto-eloszlások. A 2-es típusú, többdimenziós Pareto-eloszlást pedig mint a többdimenziós Gamma-eloszlás egy speciális esetét mutatja be.

Emellett találkozhatunk más általánosításokkal is: Rootzén és Tajvidi (2006) cikkében egy, az extrém érték elméletéhez kapcsolódó változat szerepel, míg Asimit (2009) cikkében egy többváltozós, pozitív ferdeségű Pareto-eloszlás található.

Ebben a fejezetben Mardia (1962) cikkében bevezetett 1-es típusú európai (2, illetve k -dimenziós) Pareto-eloszlást mutatom be, majd ennek tulajdonságait vizsgálom. Célom az, hogy ezen eloszlás tulajdonságainak feltérképezése után a vizsgált modellben a kárfejlődési vektort modellezem, meghatározom a várható összkárkifizetést, majd ennek segítségével meghatározom a szükséges tartalékképzés mértékét.

3.1. Kétdimenziós Pareto-eloszlás

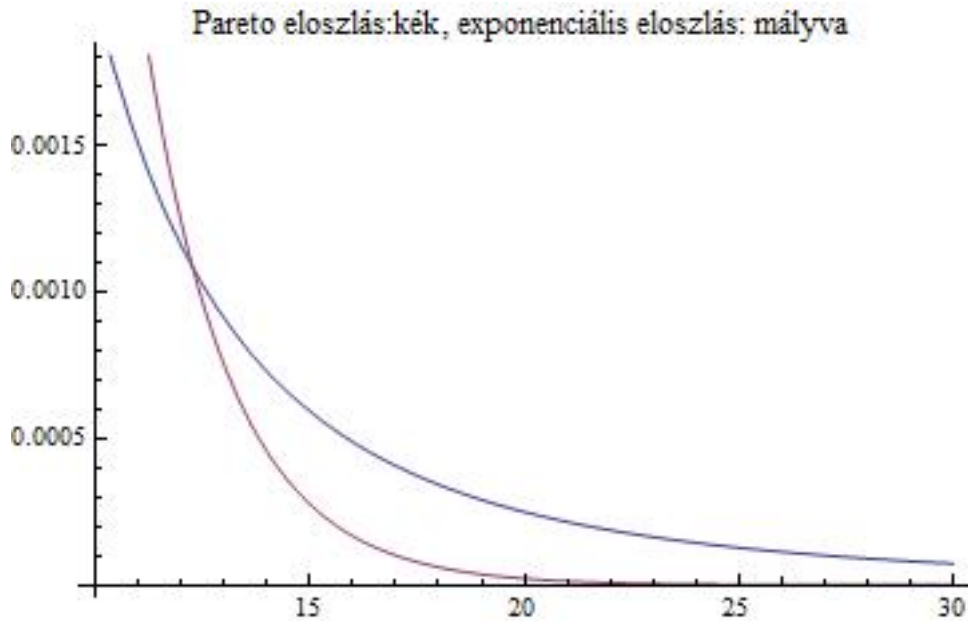
3.1. definíció. Az X valószínűségi változó európai típusú Pareto-eloszlású (a, p) paraméterekkel ($p > 0$), ha a sűrűségfüggvény:

$$f(x) = \begin{cases} \frac{pa^p}{x^{p+1}} & x > a > 0 \\ 0 & \text{egyébként} \end{cases} \quad (14)$$

Ezen eloszlás várható értéke és varianciája az alábbi képletekkel adható meg:

$$\begin{aligned} \mathbb{E}(X) &= \frac{ap}{p-1}, & \text{ha } p > 1 \\ \mathbb{D}^2(X) &= \frac{a^2p}{(p-1)^2(p-2)}, & \text{ha } p > 2 \end{aligned}$$

A következőkben tekintsük ezen eloszlásnak egy kétdimenziós általánosítását:



2. ábra. Azonos várható értékű Pareto-eloszlás és exponenciális eloszlás lecsengése.

Forrás: saját ábra

3.2. definíció. Az (X, Y) vektor 2-dimenziós Pareto-eloszlású $((a, b), p)$ paraméterekkel $(p > 0)$, ha az együttes sűrűségfüggvény:

$$f(x, y) = \begin{cases} p(p+1)(ab)^{p+1}/[(bx+ay-ab)^{p+2}] & x > a > 0, y > b > 0 \\ 0 & \text{egyébként} \end{cases} \quad (15)$$

Érdeemes megvizsgálni a fenti definícióból következő néhány egyszerű tulajdonságot.

Peremeloszlások: A 3.2 definícióban definiált eloszlás marginális eloszlásai 1 dimenziós Pareto-eloszlások, azaz X és Y egydimenziós, európai Pareto eloszlásúak (a, p) és (b, p) paraméterekkel.

Feltételes eloszlások: A 3.2 definícióban definiált eloszlásnál X feltételes eloszlása Y -ra nézve szintén 1-dimenziós Pareto-eloszlás, a feltételes sűrűségfüggvény:

$$f(x|y) = \begin{cases} [b(p+1)(ay)^{p+1}]/[(bx+ay-ab)^{p+2}] & x > a > 0, y > b > 0 \\ 0 & \text{egyébként} \end{cases} \quad (16)$$

Így a feltételes várható érték és variancia:

$$\begin{aligned} \mathbb{E}(X|Y) &= a + \frac{aY}{bp} \\ \mathbb{D}^2(X|Y) &= \frac{a^2Y^2(p+1)}{b^2(p-1)p^2} \end{aligned} \quad (17)$$

3.2. Többdimenziós Pareto-eloszlás

A 3.2. definícióban szereplő 2-dimenziós sűrűségfüggvényt írhatjuk a következő alakban is:

$$f(x, y) = \begin{cases} \frac{p(p+1)}{ab(a^{-1}x+b^{-1}y-1)^{p+2}} & x > a > 0, y > b > 0 \\ 0 & \text{egyébként} \end{cases} \quad (18)$$

Ennek mintájára Mardia (1962) nyomán a k -dimenziós általánosítás:

3.3. definíció. Az (X_1, X_2, \dots, X_k) vektor k -dimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ elhelyezkedés és $p > 0$ lecsengés paraméterekkel, ha az együttes sűrűségfüggvény:

$$f(x_1, x_2, \dots, x_k) = \begin{cases} \frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k}} & x_i > a_i > 0 \\ 0 & \text{egyébként} \end{cases} \quad (19)$$

3.2.1. Peremeloszlás

A fent definiált k -dimenziós vektor változó peremeloszlásai szintén 1-dimenziós Pareto-eloszlások (a_i, p) paraméterekkel.

Bizonyítás. A bizonyítást rekurzív módon végzem el: első lépésben belátom, hogy a $k-1$ dimenziós peremeloszlás $k-1$ dimenziós Pareto-eloszlás (a_1, \dots, a_{k-1}, p) paraméterekkel. Ezt az eljárást folytatva megkapható, hogy az 1-dimenziós peremeloszlások Pareto-eloszlást követnek (a_i, p) paraméterekkel.

Nézzük a $k-1$ dimenziós peremeloszlást:

$$\begin{aligned} f(x_1, \dots, x_{k-1}) &= \int_{x_k=a_k}^{\infty} \frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k}} dx_k = \\ &= \left[\frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k-1} (-p+k-1) a_k^{-1}} \right]_{a_k}^{\infty} = \\ &= \frac{p(p+1) \cdot \dots \cdot (p+k-2)}{\left(\prod_{i=1}^{k-1} a_i\right) \left\{ \left(\sum_{i=1}^{k-1} a_i^{-1} x_i\right) - (k-2) \right\}^{p+k-1}}, \end{aligned}$$

azaz valóban egy $k-1$ dimenziós, (a_1, \dots, a_{k-1}, p) paraméterű Pareto-eloszlást kaptunk. \square

3.2.2. Feltételes eloszlás

A k dimenziós Pareto-eloszlású \mathbf{X} vektor feltételes eloszlása az első l koordinátára nézve egy $k - l$ dimenziós, módosított Pareto-eloszlás lesz. Mielőtt rátérnék a levezetésre, bevezetem az említett módosított Pareto-eloszlást:

3.4. definíció. Az (X_1, \dots, X_k) vektor k -dimenziós, módosított Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ elhelyezkedés, $p > 0$ lecsengés és $\mathbf{a}' = (a'_1, \dots, a'_k) \geq \mathbf{a}$ módosító paraméterekkel, valamint C_k eltolással, ha a sűrűségfüggvény a következő alakú:

$$f(x_1, x_2, \dots, x_k) = \begin{cases} \frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=1}^k a'_i\right) \left\{ \left(\sum_{i=1}^k \frac{x_i}{a'_i}\right) + C_k \right\}^{p+k}} & x_i > a_i > 0 \\ 0 & \text{egyébként,} \end{cases} \quad (20)$$

valamint a teljesül az $\int_{x_1=a_1}^{\infty} \dots \int_{x_k=a_k}^{\infty} f(x_1, \dots, x_k) = 1$ feltétel.

3.5. Megjegyzés. A k -dimenziós, módosított Pareto-eloszlás speciális eseteként megkapható a 3.3-ban definiált k -dimenziós Pareto eloszlás \mathbf{a} elhelyezkedés, p lecsengés és $\mathbf{a}' = \mathbf{a}$ módosító paraméterekkel, valamint $C_k = -(k-1)$ eltolással.

Ezek után pedig következzen a feltételes eloszlást leíró állítás, mely önálló eredmény:

3.6. Állítás. Legyen $\mathbf{X} = (X_1, \dots, X_k)$ k -dimenziós Pareto-eloszlású vektor $\mathbf{a} = (a_1, \dots, a_k)$ elhelyezkedés és $p > 0$ lecsengés paraméterekkel, valamint $0 < m < k$ tetszőleges. Ekkor (X_{m+1}, \dots, X_k) feltételes eloszlása (X_1, \dots, X_m) -re nézve egy $k-m$ dimenziós, módosított Pareto-eloszlás $\mathbf{a} = (a_{m+1}, \dots, a_k)$ elhelyezkedés, $p+m$ lecsengés és $\mathbf{a}' = (a'_{m+1;m}, \dots, a'_{k;m})$ módosító paraméterekkel, valamint C_m eltolással, ahol a módosító paraméterek és az eltolás a következő rekurzív képletből kaphatóak meg minden $1 \leq l \leq m$ -dik lépésre:

$$\begin{aligned} a'_{i,l} &= \frac{x_l + D_{l-1}}{a'_{i,l-1}} a'_{i,l-1} \quad , \text{ ahol } i = l+1, \dots, m \quad \text{és} \\ C_l &= \frac{x_l}{x_l + D_{l-1}} + \frac{a'_{l;l-1}}{x_l + D_{l-1}} C_{l-1} \quad \text{és} \\ D_{l-1} &= a'_{l;l-1} \left(\sum_{i=l+1}^k \frac{a_i}{a'_{i;l-1}} + C_{l-1} \right) \end{aligned}$$

továbbá $a'_{i,0} = a_i$, $C_0 = -(k-1)$ és $D_0 = 0$.

Bizonyítás. A bizonyítást teljes indukcióval végzem. Első lépésként vizsgáljuk meg az állítást $l = 1$ -re, azaz írjuk fel (X_2, \dots, X_k) feltételes eloszlását X_1 -re nézve:

$$\begin{aligned}
f(x_2, \dots, x_k | x_1) &= \frac{f(x_1, \dots, x_k)}{f(x_1)} = \frac{\frac{p(p+1) \cdots (p+(k-1))}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i \right) - (k-1) \right\}^{p+k}}}{\frac{p a_1^p}{x_1^{p+1}}} = \\
&= \frac{p(p+1) \cdots (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i \right) - (k-1) \right\}^{p+k}} \frac{a_1}{p(a_1 x_1^{-1})^{p+1}} = \\
&= \frac{p(p+1) \cdots (p+k-1)}{\left((a_1 x_1^{-1})^{p+1} \prod_{i=2}^k a_i \right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i \right) - (k-1) \right\}^{p+k}} = \\
&= \frac{p(p+1) \cdots (p+k-1)}{\left((a_1 x_1^{-1})^{-(k-1)} \prod_{i=2}^k a_i \right) (a_1 x_1^{-1})^{p+k} \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i \right) - (k-1) \right\}^{p+k}} = \\
&= \frac{p(p+1) \cdots (p+k-1)}{\left(\prod_{i=2}^k \left(\frac{x_1}{a_1} a_i \right) \right) \left\{ \left(\sum_{i=2}^k \left(\frac{x_1}{a_1} a_i \right)^{-1} x_i \right) + 1 - (k-1) \frac{a_1}{x_1} \right\}^{p+k}}
\end{aligned}$$

A fentiekből látszik, hogy (X_2, \dots, X_k) feltételes eloszlása X_1 -re nézve egy $k - 1$ dimenziós módosított Pareto-eloszlás $\mathbf{a}' = \left(\frac{x_1}{a_1} a_2, \dots, \frac{x_1}{a_1} a_k \right)$ módosító és $p + 1$ lecsengés paraméterekkel és $C_1 = \left(1 - (k-1) \frac{a_1}{x_1} \right)$ eltolással. Így $l = 1$ -re igaz az állítás.

Tegyük fel a továbbiakban, hogy az állítás igaz $0 < l - 1 < k$ -ra. A képletek egyszerűsítése végett az $a'_{i;l-1}$ helyett ebben a lépésben az a'_i jelölést alkalmazom. Azaz az indukciós feltétel szerint (X_l, \dots, X_k) feltételes eloszlása (X_1, \dots, X_{l-1}) -re nézve egy $k - (l - 1)$ dimenziós, módosított Pareto-eloszlás a'_i módosító paraméterekkel és C_{l-1} eltolással, azaz a feltételes sűrűségfüggvény:

$$\frac{(p+l-1) \cdots (p+k-1)}{\left(\prod_{i=l}^k a'_i \right) \left\{ \left(\sum_{i=l}^k \frac{x_i}{a'_i} \right) + C_{l-1} \right\}^{p+k}} \quad (21)$$

A következő lépésben belátom, hogy az állítás igaz l -re is, azaz kiszámolom (X_{l+1}, \dots, X_k) feltételes eloszlását (X_1, \dots, X_l) -re nézve. (21)-ből könnyen kiszámolható, hogy X_l marginális sűrűségfüggvénye a következő alakú:

$$f_{X_l}(x_l | (x_1, \dots, x_{l-1})) = \frac{p+l-1}{a'_l \left(\frac{x_l}{a'_l} + \sum_{i=l+1}^k \frac{a_i}{a'_i} + C_{l-1} \right)^{p+l}} = \frac{(p+l-1) (a'_l)^{p+l-1}}{(x_l + D_{l-1})^{p+l}}$$

Ezt felhasználva az indukciós lépésben a következő feltételes eloszlás:

$$\begin{aligned}
f(x_{l+1}, \dots, x_k | x_1, \dots, x_l) &= \frac{f(x_l, \dots, x_k | (x_1, \dots, x_{l-1}))}{f(x_l | (x_1, \dots, x_{l-1}))} = \frac{\frac{(p+l-1)\dots(p+k-1)}{(\prod_{i=l}^k a'_i) \{(\sum_{i=l}^k a_i'^{-1} x_i) + C_{l-1}\}^{p+k}}}{\frac{(p+l-1)a_l'^{p+l-1}}{(x_l + D_{l-1})^{p+l}}} = \\
&= \frac{(p+l-1)(p+l) \dots (p+k-1)}{(\prod_{i=l}^k a'_i) \left\{ \left(\sum_{i=l}^k a_i'^{-1} x_i \right) + C_{l-1} \right\}^{p+k}} \frac{\left(\frac{x_l + D_{l-1}}{a'_l} \right)^{p+l}}{(p+l-1)a_l'^{-1}} = \\
&= \frac{(p+l) \dots (p+k-1)}{\left(\left(\frac{a'_l}{x_l + D_{l-1}} \right)^{p+l} \prod_{i=l+1}^k a'_i \right) \left\{ \left(\sum_{i=l}^k a_i'^{-1} x_i \right) + C_{l-1} \right\}^{p+k}} = \\
&= \frac{\left(\left(\frac{a'_l}{x_l + D_{l-1}} \right)^{-(k-l)} \prod_{i=l+1}^k a'_i \right) \left(\frac{a'_l}{x_l + D_{l-1}} \right)^{p+k} \left\{ \left(\sum_{i=l}^k \frac{x_i}{a'_i} \right) + C_{l-1} \right\}^{p+k}}{(p+l) \dots (p+k-1)} = \\
&= \frac{\left(\prod_{i=l+1}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right) \right) \left\{ \left(\sum_{i=l}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right)^{-1} x_i \right) + \frac{a'_l}{x_l + D_{l-1}} C_{l-1} \right\}^{p+k}}{(p+l) \dots (p+k-1)} = \\
&= \frac{\left(\prod_{i=l+1}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right) \right) \left\{ \left(\sum_{i=l+1}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right)^{-1} x_i \right) + \frac{x_l}{x_l + D_{l-1}} + \frac{a'_l}{x_l + D_{l-1}} C_{l-1} \right\}^{p+k}}{(p+l) \dots (p+k-1)} = \\
&= \frac{\left(\prod_{i=l+1}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right) \right) \left\{ \left(\sum_{i=l+1}^k \left(\frac{x_l + D_{l-1}}{a'_i} a'_i \right)^{-1} x_i \right) + C_l \right\}^{p+k}}{(p+l) \dots (p+k-1)},
\end{aligned}$$

ahol

$$C_l = \frac{x_l}{x_l + D_{l-1}} + \frac{a'_l}{x_l + D_{l-1}} C_{l-1} \quad (22)$$

A kapott formula alapján (X_{l+1}, \dots, X_k) feltételes eloszlása (X_1, \dots, X_l) -re nézve egy $k-l$ dimenziós, módosított Pareto-eloszlás (a_{l+1}, \dots, a_k) elhelyezkedés, $(a'_{l+1;l}, \dots, a'_{k;l})$ módosító és $p+l$ lecsengés paraméterekkel, valamint a (22)-ben definiált C_l eltolással. Azaz az indukciós lépés igaz l -re is, így az állítás helyes. \square

3.7. Megjegyzés. *A fenti állításnak van egy fontos következménye, melyre a 3.15. állítás bizonyításánál hivatkozni fogok:*

$$\sum_{s=m+1}^k \frac{a_s}{a'_{s;m}} + C_m = 1 \quad (23)$$

Ez az egyenlet nagyon könnyen belátható abból, hogy a feltételes sűrűségfüggvény is sűrűségfüggvény, azaz integrálja a $k-m$ dimenzós térben 1. Az $a'_{s;m} = a'_s$ egyszerűsítő jelölést

használva:

$$1 = \int_{x_{m+1}=a_{m+1}}^{\infty} \cdots \int_{x_k=a_k}^{\infty} \frac{(p+m) \cdot \dots \cdot (p+k-1)}{\left(\prod_{i=m+1}^k a_i' \right) \left\{ \sum_{i=m+1}^k \frac{x_i}{a_i'} + C_m \right\}^{p+k}} = \frac{1}{\left(\sum_{s=m+1}^k \frac{a_s}{a_s'} + C_m \right)^{p+m}}$$

Innen már adódik a keresett egyenlet.

3.2.3. Paraméter becslés

A paraméterek maximum likelihood becslésénél ügyelni kell arra, hogy nem feltétlen azonos hosszúságú vektorok a megfigyelések. Azaz van olyan eset, amikor csak egy kárkifizetés történt, míg más alkalommal kettő, stb., így a többdimenziós Pareto-eloszlású kárfejlődési vektoraink különböző hosszúságúak. A következő állítás önálló eredmény.

3.8. Állítás. Tegyük fel, hogy van N_j darab megfigyelésünk, ahol j változós a többdimenziós Pareto-eloszlású $(X_1^{l_j}, X_2^{l_j}, \dots, X_j^{l_j})$ vektor, ahol $l_j = 1, \dots, N_j$. Emellett $j = 1, \dots, J$, azaz maximálisan J dimenziójú megfigyeléseink vannak, valamint legyen $N = \sum_{j=1}^J N_j$ az összes adat száma. Ekkor $\hat{a}_i = \min_{l_j} X_i^{l_j}$ a maximum likelihood becslés a paraméterekre, ha kielégíti a következő egyenlőtlenség rendszert:

$$\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} - (j-1) \leq (\hat{p} + j - 1) \frac{x_v^{l_j}}{\hat{a}_v} \quad \text{és} \quad (24)$$

$$\sum_{j=1}^J N_j \left(\frac{1}{\hat{p}} + \frac{1}{\hat{p}+1} + \dots + \frac{1}{\hat{p}+j-1} \right) = \sum_{j=1}^J \sum_{l_j=1}^{N_j} \log \left(\sum_{i=1}^j \frac{x_i^{l_j}}{a_i} - (j-1) \right) \quad (25)$$

minden $l_j = 1, \dots, N_j$ vektorra és $v = 1, \dots, j$ koordinátára, ahol $j = 1, \dots, J$.

Bizonyítás. Felírva a likelihood függvényt:

$$\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_J) = \prod_{j=1}^J \prod_{l_j=1}^{N_j} \frac{p(p+1) \cdot \dots \cdot (p+j-1)}{\left(\prod_{i=1}^j a_i \right) \left\{ \left(\sum_{i=1}^j a_i^{-1} x_i^{l_j} \right) - (j-1) \right\}^{p+j}} \quad (26)$$

Nézzük először a p szerinti optimumot! (26)-ból a loglikelihood függvény:

$$\begin{aligned} \ell(\mathbf{x}_1, \dots, \mathbf{x}_n) = & \sum_{j=1}^J \sum_{l_j=1}^{N_j} (\log p + \log(p+1) + \dots + \log(p+j-1)) - \\ & \sum_{j=1}^J N_j \log \left(\prod_{i=1}^j a_i \right) - \sum_{j=1}^J (p+j) \sum_{l_j=1}^{N_j} \log \left(\sum_{i=1}^j \frac{x_i^{l_j}}{a_i} - (j-1) \right) \end{aligned}$$

Ennek p szerinti optimumához tekintsük a p szerinti elsőrendű parciális deriváltat:

$$0 = \frac{\partial \mathcal{L}(x_1, \dots, x_n)}{\partial p} = \sum_{j=1}^J N_j \left(\frac{1}{p} + \frac{1}{p+1} + \dots + \frac{1}{p+j-1} \right) - \sum_{j=1}^J \sum_{l_j=1}^{N_j} \log \left(\sum_{i=1}^j \frac{x_i^{l_j}}{a_i} - (j-1) \right)$$

Innen már adódik a (25) egyenlet:

$$\sum_{j=1}^J N_j \left(\frac{1}{\hat{p}} + \frac{1}{\hat{p}+1} + \dots + \frac{1}{\hat{p}+k-1} \right) = \sum_{j=1}^J \sum_{l_j=1}^{N_j} \log \left(\sum_{i=1}^j \frac{x_i^{l_j}}{a_i} - (j-1) \right)$$

Az egyenlet bal oldala szigorúan pozitív, és a jobb oldala is szigorúan pozitív, mivel $x_i^{l_j} > a_i$, így a logaritmus belsejében 1-nél nagyobb számok állnak, azaz nemnegatív számokat adunk össze. Az egyenlet jobb oldala konstans, míg a bal oldal p szigorúan monoton csökkenő függvénye, így az egyenletnek csak 1 megoldása van. Továbbá könnyen látszik, hogy a parciális derivált p -nek szigorúan monoton csökkenő függvénye, így valóban maximumot találtunk.

A következőkben tekintsük, hogy milyen \hat{a}_i becslések esetén tudjuk maximalizálni a likelihood függvényt. Mivel $a_i \leq X_i^{l_j}$ minden $i = 1, \dots, j$ koordinátára és az összes lehetséges l_j vektorra, így $\hat{a}_i \leq \min X_i^{l_j}$ -nek teljesülnie kell.

Tegyük fel, hogy kiszámítottuk az $\hat{\mathbf{a}} = (\min X_1^{l_1}, \dots, \min X_j^{l_j})$ becslés vektort a koordinátákra. Tekintsünk a v . koordinátára egy másik becslést: $0 < b_v < \hat{a}_v \leq X_v^{l_v}$. Az előző feltétel miatt $\exists r \in (0, 1)$, melyre $b_v = r \cdot \hat{a}_v$. A következőkben vizsgáljuk meg azt, hogy milyen r (azaz milyen b_v) esetén tudjuk növelni a likelihood függvény értékét. Ehhez tekintsünk egy általános tagot, ahol \hat{a}_v helyébe b_v -t helyettesítünk:

$$\mathcal{L}_b = \frac{p(p+1) \cdot \dots \cdot (p+j-1)}{\left(\prod_{i \neq v} \hat{a}_i \right) b_v \left\{ \left(\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} + \frac{x_v^{l_j}}{b_v} \right) - (j-1) \right\}^{p+j}}$$

A javítás szükséges feltétele, hogy

$$\mathcal{L}_b > \mathcal{L}_a = \frac{p(p+1) \cdot \dots \cdot (p+j-1)}{\left(\prod_{i \neq v} \hat{a}_i \right) \hat{a}_v \left\{ \left(\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} + \frac{x_v^{l_j}}{\hat{a}_v} \right) - (j-1) \right\}^{p+j}}$$

A fenti egyenlőtlenséget átrendezve:

$$\hat{a}_v \left\{ \left(\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} + \frac{x_v^{l_j}}{\hat{a}_v} \right) - (j-1) \right\}^{p+j} > b_v \left\{ \left(\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} + \frac{x_v^{l_j}}{b_v} \right) - (j-1) \right\}^{p+j}$$

Legyen $C_v^{l_j} = \sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} - (j-1)$, ezt behelyettesítve:

$$\hat{a}_v \left\{ C_v^{l_j} + \frac{x_v^{l_j}}{\hat{a}_v} \right\}^{p+j} > b_v \left\{ C_v^{l_j} + \frac{x_v^{l_j}}{b_v} \right\}^{p+j}$$

Az egyenlet mindkét oldalát b_v^{p+j} -nel szorozva:

$$\begin{aligned} \hat{a}_v \left\{ C_v^{l_j} b_v + x_v^{l_j} \frac{b_v}{\hat{a}_v} \right\}^{p+j} &> b_v \left\{ C_v^{l_j} b_v + x_v^{l_j} \right\}^{p+j} \\ \left\{ \frac{C_v^{l_j} b_v + x_v^{l_j} r}{C_v^{l_j} b_v + x_v^{l_j}} \right\}^{p+j} &> \frac{b_v}{\hat{a}_v} = r \end{aligned}$$

Mivel $p+j > 0$, így gond nélkül $\frac{1}{p+j}$ -dik hatványra emelhetjük mindkét oldalt:

$$\frac{C_v^{l_j} b_v + x_v^{l_j} r}{C_v^{l_j} b_v + x_v^{l_j}} > r^{\frac{1}{p+j}}$$

Ismét keresztbe szorozva:

$$C_v^{l_j} b_v r^{-\frac{1}{p+j}} + x_v^{l_j} r^{1-\frac{1}{p+j}} > C_v^{l_j} b_v + x_v^{l_j}$$

Az egyenlőtlenségbe $b_v = \hat{a}_v \cdot r$ -t helyettesítve:

$$C_v^{l_j} \hat{a}_v r^{1-\frac{1}{p+j}} + x_v^{l_j} r^{1-\frac{1}{p+j}} > C_v^{l_j} \hat{a}_v r + x_v^{l_j}$$

Legyen $k(j) = 1 - \frac{1}{p+j}$ az r kitevője. Az egyszerűség kedvéért a továbbiakban ezt k -val jelölöm. Így az utolsó egyenlőtlenség:

$$\begin{aligned} C_v^{l_j} \hat{a}_v r^k + x_v^{l_j} r^k &> C_v^{l_j} \hat{a}_v r + x_v^{l_j} \\ C_v^{l_j} \hat{a}_v (r^k - r) &> x_v^{l_j} (1 - r^k) \end{aligned}$$

Azaz a javításhoz szükséges feltétel:

$$C_v^{l_j} \frac{\hat{a}_v}{x_v^{l_j}} > \frac{1 - r^k}{r^k - r} \quad (27)$$

Az átosztásnál a relációs jel nem fordul meg, hiszen $r^k - r > 0$, mivel $r < 1$ és $k < 1$, így $r^k > r$.

Összefoglalva a kapott eredményt: ha $\exists r \in (0, 1)$, amelyre (27) teljesül, akkor a $b_v < \hat{a}_v$ becslés javít a likelihood függvényen. Más szavakkal, ha $\forall r \in (0, 1)$ -re (27) nem teljesül, azaz $C_v^{l_j} \frac{\hat{a}_v}{x_v^{l_j}} \leq \frac{1-r^k}{r^k-r}$, akkor nem tudunk javítani az \hat{a}_v becslésen.

Ehhez vizsgáljuk meg (27) egyenlőtlenség jobb oldalát, azaz elemezzük a

$$g(r) = \frac{1 - r^k}{r^k - r}$$

függvényt monotonitás, illetve korlátosság szempontjából. Felírva a deriváltfüggvényt:

$$\begin{aligned} g'(r) &= \frac{-kr^{k-1}(r^k - r) - (1 - r^k)(kr^k - 1)}{(r^k - r)^2} = \frac{-kr^{2k-1} + kr^k - kr^{k-1} + 1 + kr^{2k-1} - r^k}{(r^k - r)^2} = \\ &= \frac{(k-1)r^k - kr^{k-1} + 1}{(r^k - r)^2} \end{aligned}$$

$g'(r)$ számlálójára adható egy r -től független becslés, ehhez tekintsük a számláló (mint r függvénye) deriváltját:

$$\begin{aligned} \frac{d}{dr}(k-1)r^k - kr^{k-1} + 1 &= k(k-1)r^{k-1} - k(k-1)r^{k-2} = \\ &= k(k-1)r^k \left(\frac{1}{r} - \frac{1}{r^2} \right) = k(k-1)r^k \cdot \frac{r-1}{r^2} \end{aligned}$$

Mivel $0 < k < 1$ és $0 \leq r \leq 1$, így ez a derivált nagyobb egyenlő mint 0, azaz a számláló (mint r függvénye) monoton növekvő. Így maximumát a jobb oldali végpontban, azaz $r = 1$ -ben éri el, ahol értéke 0.

Azaz a becslés alapján $g'(r) \leq 0$, így $g(r)$ monoton csökkenő. Sőt, ha $r < 1$, akkor teljesül a szigorú egyenlőtlenség is, így $g(r)$ szigorúan monoton csökkenő r -ben ($r \in (0, 1)$). Ezt az észrevételt felhasználva számítsuk ki a határértéket a végpontokban:

$$\lim_{r \rightarrow 0} g(r) = \lim_{r \rightarrow 0} \frac{1 - r^k}{r^k - r} = +\infty$$

$$\lim_{r \rightarrow 1} g(r) = \lim_{r \rightarrow 1} \frac{1 - r^k}{r^k - r} = \lim_{r \rightarrow 1} \frac{-kr^{k-1}}{kr^{k-1} - 1} = \frac{-k}{k-1} = \frac{k}{1-k} = \frac{1 - \frac{1}{p+j}}{\frac{1}{p+j}} = p + j - 1 > 0$$

Az eredmények alapján $g(r)$ alulról korlátos, a legjobb korlát pedig $p + j - 1$. Ennek segítségével a (27) egyenlőtlenségben megfogalmazott észrevételt újrafogalmazhatjuk: ha $C_v^{l_j} \frac{\hat{a}_v}{x_v} \leq p + j - 1$ igaz, azaz ha teljesül a

$$\sum_{i \neq v} \frac{x_i^{l_j}}{\hat{a}_i} - (j-1) \leq (p+j-1) \frac{x_v^{l_j}}{\hat{a}_v}$$

egyenlőtlenség minden l_j vektorra, és ezeknek minden $i = 1, \dots, j$ koordinátájára, akkor nincs az $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_j)$ vektornál jobb becslés az $\mathbf{a} = (a_1, \dots, a_j)$ paraméterekre. \square

3.9. Megjegyzés. A (3.8) állítását vizsgáljuk meg a $j = 1$ és $j = 2$ dimenziók esetében. $j = 1$ -re az egyenlőtlenség rendszer a következő alakot ölti:

$$0 - (1 - 1) \leq (p + 1 - 1) \frac{x^l}{a} = p \frac{x^l}{a},$$

ami nyilván teljesül, mivel p , a , illetve x^l is pozitív minden l vektorra. Így nincs az $\hat{a} = \min X^l$ -nél jobb becslés. Ez összhangban van az egydimenziós Pareto-eloszlásra ismert maximum likelihood becsléssel.

$j = 2$ -re az egyenlőtlenség rendszer a következő alakot ölti:

$$\begin{aligned} \frac{x_1^l}{a_1} &\leq (p + 1) \frac{x_2^l}{a_2} \\ \frac{x_2^l}{a_2} &\leq (p + 1) \frac{x_1^l}{a_1} \end{aligned}$$

Ezt már nem feltétlen elégíti ki az $\hat{\mathbf{a}} = (\min x_1^l, \min x_2^l)$ becslés, elképzelhető, hogy adott p értékre találunk ennél jobbat.

3.2.4. A tartalékok meghatározásához szükséges állítások

Ahhoz, hogy a 2. fejezetben bemutatott modellt azon feltétel mellett vizsgáljuk, hogy a 2.1.3. fejezetben bemutatott kárfejlődési minta többdimenziós Pareto-eloszlású, szükség van néhány állítás kimondására. Az alábbi alfejezetben található állítások mind önálló eredmények. Közülük az első egy IBNR (és RBNP) kár várható értékének kiszámolásához nyújt segítséget:

3.10. Állítás. Tegyük fel, hogy az $\mathbf{X} = (X_1, X_2, \dots, X_k)$ vektor többdimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ és p paraméterekkel. Ekkor $p > k$ feltétel mellett a koordináták szorzatának várható értéke a következőképpen kapható meg:

$$\mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_k) = \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{a_1 \dots a_k \cdot p(p+1) \dots (p+k-1)}{(p+k-1)(p+k-2) \dots (p+1-i)} \cdot \frac{1}{p-1-i} \quad (28)$$

Bizonyítás. A várható érték definíciójából:

$$\begin{aligned} \mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_k) &= \int_{x_1=a_1}^{\infty} \int_{x_2=a_2}^{\infty} \dots \int_{x_k=a_k}^{\infty} x_1 \cdot x_2 \cdot \dots \cdot x_k \cdot f(x_1, x_2, \dots, x_k) dx_k \dots dx_1 = \\ &= \int_{a_1}^{\infty} \int_{a_2}^{\infty} \dots \int_{a_k}^{\infty} x_1 \cdot x_2 \cdot \dots \cdot x_k \cdot \frac{p(p+1) \dots (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k}} dx_k \dots dx_1 = \\ &= \int_{a_1}^{\infty} \int_{a_2}^{\infty} \dots \int_{a_k}^{\infty} \frac{x_1}{a_1} \cdot \frac{x_2}{a_2} \cdot \dots \cdot \frac{x_k}{a_k} \cdot \frac{p(p+1) \dots (p+k-1)}{\left\{ \left(\sum_{i=1}^k \frac{x_i}{a_i}\right) - (k-1) \right\}^{p+k}} dx_k \dots dx_1 \quad (29) \end{aligned}$$

Az integrál kiszámításához alkalmazzuk a $t_i = \frac{x_i}{a_i}$ helyettesítést. Így a Jacobi-determináns $|\mathbf{J}| = a_1 \cdot a_2 \cdot \dots \cdot a_k$ és $t_i \in (1, \infty)$. A helyettesítés után (29) a következő alakban írható fel:

$$a_1 \cdot \dots \cdot a_k \int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \dots \int_{t_k=1}^{\infty} t_1 \cdot t_2 \cdot \dots \cdot t_k \cdot \frac{p(p+1) \dots (p+k-1)}{\left\{ \left(\sum_{i=1}^k t_i \right) - (k-1) \right\}^{p+k}} dt_k \dots dt_1$$

A későbbiekben tekintsük a konstansok nélküli k -változós integrált:

$$\int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \dots \int_{t_k=1}^{\infty} t_1 \cdot t_2 \cdot \dots \cdot t_k \cdot \frac{1}{\{t_1 + t_2 + \dots + t_k - (k-1)\}^{p+k}} dt_k \dots dt_1$$

A k -változós integrál kiszámításához először végezzük el a t_k szerinti integrálást:

$$\begin{aligned} & \int_{t_k=1}^{\infty} t_k \frac{1}{(t_1 + t_2 + \dots + t_k - (k-1))^{p+k}} dt_k = \\ & = \left[\frac{t_k}{(-p+k-1)(t_1 + t_2 + \dots + t_k - (k-1))^{p+k-1}} \right]_{t_k=1}^{\infty} - \\ & \quad - \int_{t_k=1}^{\infty} \frac{1}{(-p+k-1)(t_1 + t_2 + \dots + t_k - (k-1))^{p+k-1}} dt_k = \\ & = \frac{1}{(p+k-1)(t_1 + t_2 + \dots + t_{k-1} + 1 - (k-1))^{p+k-1}} - \\ & \quad - \left[\frac{1}{(p+k-1)(p+k-2)(t_1 + t_2 + \dots + t_k - (k-1))^{p+k-2}} \right]_{t_k=1}^{\infty} = \\ & = \frac{1}{(p+k-1)(t_1 + t_2 + \dots + t_{k-1} - (k-2))^{p+k-1}} + \\ & \quad + \frac{1}{(p+k-1)(p+k-2)(t_1 + t_2 + \dots + t_{k-1} - (k-2))^{p+k-2}} \end{aligned}$$

Látható, hogy a következő lépésben az első lépésben elvégzett integrálhoz hasonló alakú integrálokkal lesz dolgunk, azaz megfigyelhető egy rekurzív formula a többi változó szerinti egyszeres integrálok elvégzésére. Ehhez vezessük be a következő jelölést:

$$I_n^j = \frac{1}{(p+k-1)(p+k-2) \dots (p+k-n)(t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n}} \quad (30)$$

Az n -es alsó index azt mutatja, hogy hány darab szorzótag jött be a hatvány elé az első lépéstől kezdve (illetve a hatványkitevő hogyan változott ezzel párhuzamosan). A j felső index pedig azt mutatja, hogy az integrálási lépések során a hatványalapban hány darab t_i változó esett ki a lépések során, azaz hány darab egyszeres integrált végeztünk már el.

A rekurzív formula megalkotásához számítsuk ki $\int_{t_{k-j}=1}^{\infty} t_{k-j} I_n^j dt_{k-j}$ -t. Mivel az integrálás során a konstans szorzók kivihetők az integrál elé, így jelöljük C -vel I_n^j konstans tagjait, azaz $C = (p+k-1)(p+k-2)\dots(p+k-n)$. Ennek felhasználásával

$$\begin{aligned}
& \int_{t_{k-j}=1}^{\infty} t_{k-j} I_n^j dt_{k-j} = \int_{t_{k-j}=1}^{\infty} t_{k-j} \frac{1}{C(t_1+t_2+\dots+t_{k-j}-(k-j-1))^{p+k-n}} dt_{k-j} = \\
& = \left[\frac{t_{k-j}}{C(-(p+k-n-1))(t_1+t_2+\dots+t_{k-j}-(k-j-1))^{p+k-n-1}} \right]_{t_{k-j}=1}^{\infty} - \\
& \quad - \int_{t_{k-j}=1}^{\infty} \frac{1}{C(-(p+k-n-1))(t_1+t_2+\dots+t_{k-j}-(k-j-1))^{p+k-n-1}} dt_{k-j} = \\
& = \frac{1}{C(p+k-n-1)(t_1+t_2+\dots+t_{k-j-1})+1-(k-j-1))^{p+k-n-1}} - \\
& \quad - \left[\frac{1}{C(p+k-n-1)(p+k-n-2)(t_1+t_2+\dots+t_{k-j}-(k-j-1))^{p+k-n-2}} \right]_{t_{k-j}=1}^{\infty} = \\
& = \frac{1}{C(p+k-(n+1))(t_1+t_2+\dots+t_{k-(j+1)}-(k-(j+1)-1))^{p+k-(n+1)}} + \\
& \quad + \frac{1}{C(p+k-(n+1))(p+k-(n+2))(t_1+t_2+\dots+t_{k-(j+1)}-(k-j-2))^{p+k-(n+2)}} \\
& = I_{n+1}^{j+1} + I_{n+2}^{j+1} \tag{31}
\end{aligned}$$

Az előző jelöléssel az első integráltként:

$$\int_{t_k=1}^{\infty} \frac{t_k}{(t_1+\dots+t_k-(k-1))^{p+k}} dt_k = I_1^1 + I_2^1 \tag{32}$$

Ennek felhasználásával kapjuk a 2 változós integrált:

$$\int_{t_{k-1}=1}^{\infty} \int_{t_k=1}^{\infty} \frac{t_{k-1} \cdot t_k}{(t_1+\dots+t_k-(k-1))^{p+k}} dt_k dt_{k-1} = \int_{t_{k-1}=1}^{\infty} t_{k-1} \cdot (I_1^1 + I_2^1) dt_{k-1}$$

A (31) rekurziós egyenlet felhasználásával az integrál tagjai:

$$\begin{aligned}
& \int_{t_{k-1}=1}^{\infty} t_{k-1} \cdot I_1^1 dt_{k-1} = I_2^2 + I_3^2, \quad \text{és} \\
& \int_{t_{k-1}=1}^{\infty} t_{k-1} \cdot I_2^1 dt_{k-1} = I_3^2 + I_4^2, \quad \text{így}
\end{aligned}$$

$$\begin{aligned} \int_{t_{k-1}=1}^{\infty} \int_{t_k=1}^{\infty} \frac{t_{k-1} \cdot t_k}{(t_1 + \dots + t_k - (k-1))^{p+k}} dt_k dt_{k-1} &= I_2^2 + I_3^2 + I_3^2 + I_4^2 \\ &= I_2^2 + 2I_3^2 + I_4^2 \end{aligned}$$

Tovább folytatva a gondolatmenetet:

$$\begin{aligned} &\int_{t_{k-2}=1}^{\infty} \int_{t_{k-1}=1}^{\infty} \int_{t_k=1}^{\infty} \frac{t_{k-2} \cdot t_{k-1} \cdot t_k}{(t_1 + \dots + t_k - (k-1))^{p+k}} dt_k dt_{k-1} dt_{k-2} = \\ &= \int_{t_{k-2}=1}^{\infty} t_{k-2} \cdot (I_2^2 + 2I_3^2 + I_4^2) dt_{k-2} = I_3^3 + I_4^3 + 2(I_4^3 + I_5^3) + I_5^3 + I_6^3 = \\ &= I_3^3 + 3I_4^3 + 3I_5^3 + I_6^3 \end{aligned} \quad (33)$$

Az eljárás folytatható a 4,5,... változós integrál kiszámítására is, de már itt is kirajzolódik a binomiális együtthatós struktúra, amelynek pontos megfogalmazását a következő lemma adja:

3.11. Lemma. *Az előző jelölésekkel, $N = 0 \dots (k-2)$ -re:*

$$\int_{t_{k-N}=1}^{\infty} \dots \int_{t_k=1}^{\infty} \frac{t_{k-N} \cdot \dots \cdot t_k}{(t_1 + \dots + t_k - (k-1))^{p+k}} dt_k \dots dt_{k-N} = \sum_{i=0}^{N+1} \binom{N+1}{i} I_{N+1+i}^{N+1} \quad (34)$$

Bizonyítás. A bizonyítást teljes indukcióval végzem. A lemma állítása $N = 0$ -ra (32) miatt igaz. Tegyük fel, hogy N -re igaz az állítás, azaz

$$\int_{t_{k-N}=1}^{\infty} \dots \int_{t_k=1}^{\infty} \frac{t_{k-N} \cdot \dots \cdot t_k}{(t_1 + \dots + t_k - (k-1))^{p+k}} dt_k \dots dt_{k-N} = \sum_{i=0}^{N+1} \binom{N+1}{i} I_{N+1+i}^{N+1}$$

Vizsgáljuk meg, hogy $N + 1$ -re is igaz-e az állítás (feltéve, hogy $N + 1 < k - 2$).

$$\begin{aligned}
& \int_{t_{k-(N+1)}=1}^{\infty} \int_{t_{k-N}=1}^{\infty} \cdots \int_{t_k=1}^{\infty} \frac{t_{k-N} \cdot \dots \cdot t_k}{(t_1 + \dots + t_k - (k-1))^{p+k}} dt_k \dots dt_{k-N} dt_{k-(N+1)} = \\
&= \int_{t_{k-(N+1)}=1}^{\infty} t_{k-(N+1)} \sum_{i=0}^{N+1} \binom{N+1}{i} I_{N+1+i}^{N+1} dt_{k-(N+1)} = \\
&= \sum_{i=0}^{N+1} \binom{N+1}{i} \int_{t_{k-(N+1)}=1}^{\infty} t_{k-(N+1)} \cdot I_{N+1+i}^{N+1} dt_{k-(N+1)} = \\
&= \sum_{i=0}^{N+1} \binom{N+1}{i} (I_{N+2+i}^{N+2} + I_{N+3+i}^{N+2}) = \sum_{i=0}^{N+1} \binom{N+1}{i} I_{N+2+i}^{N+2} + \sum_{i=1}^{N+2} \binom{N+1}{i-1} I_{N+2+i}^{N+2} = \\
&= I_{N+2}^{N+2} + \left\{ \sum_{i=1}^{N+1} \left[\binom{N+1}{i-1} + \binom{N+1}{i} \right] I_{N+2+i}^{N+2} \right\} + I_{N+2+(N+2)}^{N+2} = \sum_{i=0}^{N+2} \binom{N+2}{i} I_{N+2+i}^{N+2}
\end{aligned}$$

Az utolsó lépésben a binomiális együtthatókra vonatkozó ismert összefüggést használtam. □

Az integrálási lépések elvégzése során a nevezőben eltűnik egy változó, azaz $k - 1$ lépés után a nevezőben csak egy változó, a t_k marad. Azaz $N = k - 1$ esetén már csak egy hatványfüggvény integrálját kell kiszámítani, nincs szükség parciális integrálásra. Így a keresett integrál:

$$\begin{aligned}
& \int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \cdots \int_{t_k=1}^{\infty} \frac{t_1 \cdot t_2 \cdot \dots \cdot t_k}{\{t_1 + t_2 + \dots + t_k - (k-1)\}^{p+k}} dt_k \dots dt_2 dt_1 = \\
&= \int_{t_1=1}^{\infty} t_1 \cdot \sum_{i=0}^{k-1} \binom{k-1}{i} I_{k-1+i}^{k-1} dt_1 = \sum_{i=0}^{k-1} \binom{k-1}{i} \int_{t_1=1}^{\infty} t_1 I_{k-1+i}^{k-1} dt_1 = \\
&= \sum_{i=0}^{k-1} \binom{k-1}{i} \int_{t_1=1}^{\infty} \frac{t_1}{(p+k-1)(p+k-2) \dots (p+k-(k-1+i))(t_1)^{p+k-(k-1+i)}} dt_1 = \\
&= \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{1}{(p+k-1)(p+k-2) \dots (p+1-i)} \left[\frac{1}{-(p-1-i)t_1^{p-1-i}} \right]_{t_1=1}^{\infty} = \\
&= \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{1}{(p+k-1)(p+k-2) \dots (p+1-i)(p-1-i)}
\end{aligned}$$

Az elvégzett lépések során végig kihasználtam a $p > k$ feltételt, hiszen így minden impromprius integrál konvergens volt.

Összerakva a kapott eredményeket, a keresett várható érték a következő zárt formulával adható meg:

$$\mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_k) = \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{a_1 \dots a_k \cdot p(p+1) \dots (p+k-1)}{(p+k-1)(p+k-2) \dots (p+1-i)(p-1-i)}$$

□

3.12. Megjegyzés. Az előző állításban $k = 1$ esetén a $(p+k-1)(p+k-2) \dots (p+1-i)$ szorzat értéke alatt 1-et értek. Ezzel a jelöléssel visszakapjuk az 1-dimenziós Pareto-eloszlás várható értékét.

Továbbá fontosnak tartottam, hogy az előző állításban kapott formula helyességét számítógép segítségével ellenőrizzem $k = 2, 3$ esetén. A Wolfram Mathematica szimbolikus programcsomag segítségével kiszámítottam a várható értéket, és a kapott eredmények megegyeztek a formulából kiszámolható értékekkel.

A várható kárnagság mellett fontos információt rejt magában az összkárkifizetés szórása is, amelynek meghatározása nehezebb feladat. A következő állítás a szorzat második momentumát adja meg, ennek segítségével már könnyen számolható a szórásnégyzet is.

3.13. Állítás. Tegyük fel, hogy az $\mathbf{X} = (X_1, X_2, \dots, X_k)$ vektor többdimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ és p paraméterekkel. Ekkor $p > 2k$ esetben a koordináták szorzatának második momentuma a következőképpen kapható meg:

$$\mathbb{E}((X_1 \cdot X_2 \cdot \dots \cdot X_k)^2) = \sum_{i=0}^{2k-2} F_i \frac{a_1^2 \dots a_k^2 \cdot p(p+1) \dots (p+k-1)}{(p+k-1)(p+k-2) \dots (p+1-i)} \frac{1}{(p-2-i)}, \quad (35)$$

ahol

$$F_i = \sum_{s=\max(0, i-(k-1))}^{\lfloor \frac{i}{2} \rfloor} \binom{k-1}{s} \binom{k-1-s}{i-2s} 2^{i-s} \quad (36)$$

Bizonyítás. A bizonyítás a 3.10. állítás bizonyításának analógiája, induljunk ki a keresett várható érték definíciójából:

$$\begin{aligned} \mathbb{E}((X_1 \cdot X_2 \cdot \dots \cdot X_k)^2) &= \int_{x_1=a_1}^{\infty} \dots \int_{x_k=a_k}^{\infty} x_1^2 \cdot x_2^2 \cdot \dots \cdot x_k^2 \cdot f(x_1, x_2, \dots, x_k) dx_k \dots dx_1 = \\ &= \int_{a_1}^{\infty} \int_{a_2}^{\infty} \dots \int_{a_k}^{\infty} x_1^2 \cdot x_2^2 \cdot \dots \cdot x_k^2 \cdot \frac{p(p+1) \dots (p+k-1)}{\left(\prod_{i=1}^k a_i\right) \left\{ \left(\sum_{i=1}^k a_i^{-1} x_i\right) - (k-1) \right\}^{p+k}} dx_k \dots dx_1 = \\ &= a_1 \cdot \dots \cdot a_k \int_{a_1}^{\infty} \int_{a_2}^{\infty} \dots \int_{a_k}^{\infty} \frac{x_1^2}{a_1^2} \cdot \frac{x_2^2}{a_2^2} \cdot \dots \cdot \frac{x_k^2}{a_k^2} \cdot \frac{p(p+1) \dots (p+k-1)}{\left\{ \left(\sum_{i=1}^k \frac{x_i}{a_i}\right) - (k-1) \right\}^{p+k}} dx_k \dots dx_1 \quad (37) \end{aligned}$$

Az integrál kiszámításához alkalmazzuk a $t_i = \frac{x_i}{a_i}$ helyettesítést. Így a Jacobi-determináns $|\mathbf{J}| = a_1 \cdot a_2 \cdot \dots \cdot a_k$ és $t_i \in (1, \infty)$. A helyettesítés után (37) a következő alakban írható fel:

$$a_1^2 \cdot \dots \cdot a_k^2 \int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \dots \int_{t_k=1}^{\infty} t_1^2 \cdot t_2^2 \cdot \dots \cdot t_k^2 \cdot \frac{p(p+1) \dots (p+k-1)}{\left\{ \left(\sum_{i=1}^k t_i \right) - (k-1) \right\}^{p+k}} dt_k \dots dt_1 \quad (38)$$

A későbbiekben tekintsük a konstansok nélküli k -változós integrált:

$$\int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \dots \int_{t_k=1}^{\infty} t_1^2 \cdot t_2^2 \cdot \dots \cdot t_k^2 \cdot \frac{1}{\{t_1 + t_2 + \dots + t_k - (k-1)\}^{p+k}} dt_k \dots dt_1 \quad (39)$$

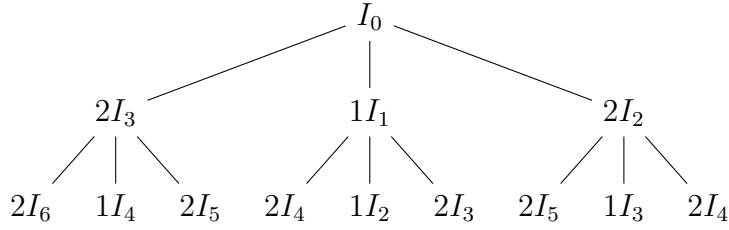
A cél itt is egy rekurzív formula megalkotása, ezért legyen ismételt:

$$I_n^j = \frac{1}{(p+k-1)(p+k-2) \dots (p+k-n) (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n}},$$

valamint számítsuk ki $\int_{t_{k-j}=1}^{\infty} t_{k-j}^2 I_n^j dt_{k-j}$, ahol $j = 1, \dots, k-1$. Mivel az integrálás során a konstans szorzók kivihetőek az integrál elé, így jelöljük C -vel I_n^j konstans tagjait, azaz $C = (p+k-1)(p+k-2) \dots (p+k-n)$. Ennek felhasználásával:

$$\begin{aligned} \int_{t_{k-j}=1}^{\infty} t_{k-j}^2 I_n^j dt_{k-j} &= \int_{t_{k-j}=1}^{\infty} t_{k-j}^2 \frac{1}{C (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n}} dt_{k-j} = \\ &= \left[\frac{t_{k-j}^2}{C (-(p+k-n-1)) (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n-1}} \right]_{t_{k-j}=1}^{\infty} - \\ &\quad - \int_{t_{k-j}=1}^{\infty} \frac{2t_{k-j}}{C (-(p+k-n-1)) (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n-1}} dt_{k-j} = \\ &= \frac{1}{C(p+k-n-1) (t_1 + t_2 + \dots + t_{k-j-1}) + 1 - (k-j-1)} - \\ &\quad - \left[\frac{2t_{k-j}}{C(p+k-n-1)(p+k-n-2) (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n-2}} \right]_{t_{k-j}=1}^{\infty} + \\ &\quad + \int_{t_{k-j}=1}^{\infty} \frac{2}{C(p+k-n-1)(p+k-n-2) (t_1 + t_2 + \dots + t_{k-j} - (k-j-1))^{p+k-n-2}} dt_{k-j} \\ &= I_{n+1}^{j+1} + 2I_{n+2}^{j+1} + 2I_{n+3}^{j+1} \end{aligned} \quad (40)$$

Az integrálási lépések rekurzív elvégzése során a nevezőben eltűnik egy változó, azaz $k-1$ lépés után már csak egy változó, t_k marad, és egy hatványfüggvényt kell kiintegrálni. Így a cél egy olyan formula megalkotása, amely $j = 1, \dots, k-1$ -re megadja az integrátagokat azok együtthatóival együtt.



A rekurzió alapján látható, hogy minden $\int_{t_{k-j}=1}^{\infty} t_{k-j}^2 I_n^j dt_{k-j}$ alakú integrál elvégzése után 3 tag, azaz 3 utód keletkezik. Tekintsük a folyamat "ábráját" egy gráfként, ahogy az ábra is mutatja: a fa gyökere I_0^0 , azaz (39)-ben keresett integrál, a j . szinten pedig a j . lépés után a rekurzió alapján kapott integráltak találhatók, a hozzájuk tartozó együtthatóval együtt. A fa egy $k-1$ mélységű, teljes 3-adfokú (másnéven trinomiális) fa, mivel a $k-1$. lépés után csak hatványfüggvények integrálját kell kiszámolnunk, valamint minden szülőnek 3 gyereke van.

A j . szinten a következő alakú integráltak fordulnak elő: $I_j, I_{j+1}, \dots, I_{3j}$ (a jelölések egyszerűsítése végett elhagyom a felső indexet). Fontos észrevétel, hogy ezen a szinten összesen 3^j csúcs van, de csak $2j+1$ darab különböző indexű tag, így némely indexű tag többször, különböző együtthatóval szerepel. A végső cél tehát a $k-1$. szinten minden $n = k-1, \dots, 3(k-1)$ indexre meghatározni az együtthatók összegét.

A kérdés megválaszolásához vezessük be a következő jelölésrendszert: legyen minden csúcs középső gyereke 1-es, jobb oldali gyereke 2-es, valamint bal oldali gyereke 3-as számú. Ezek a kódok adják meg, hogy a szülő indexéhez képest mennyivel nagyobb a gyerek indexe, valamint a szülő együtthatóját mivel kell megszorozni, hogy megkapjuk az adott gyerek együtthatóját: 1-es esetében 1-gyel, 2-es esetében 2-vel, míg 3-as esetében szintén 2-vel.

Mivel teljes trinomiális fával dolgozunk, így a j . szinten minden csúcs indexe és együtthatója megadható egy j hosszúságú, csak az $(1, 2, 3)$ számokat tartalmazó listával, aminek m . eleme a csúcs m . szinten lévő ősenek kódja ($m = j$ esetén az adott csúcs kódja). Jelölje s a 3-asok, t a 2-esek, míg $j-s-t$ az 1-esek számát a kódban, feltéve, hogy $s \in [0, j]$ és $t \in [0, j-s]$. Így minden j . szinten lévő csúcs indexe a lista elemeinek összege (azaz a gyökérhez képest hányal növeltük az indexet), formálisan: $3s+2t+(j-s-t) = 2s+t+j$. Az adott csúcs együtthatója pedig 2^{s+t} lesz, mivel a gyökér együtthatóját (azaz 1-et) $s+t$ alkalommal szoroztuk 2-vel és $j-t-s$ alkalommal 1-gyel.

A kódrendszer szemléltetéséhez tekintsük példaként a 2. szint 2. elemét, I_4 -et. Itt $j = 2$, azaz 2 hosszúságú listát keresünk. Ennek első eleme 3 (az első szinten lévő ő, I_3 kódja), második eleme pedig 1 (I_4 , a vizsgált csúcs kódja). A kódban $s = 1$ db 3-as, $t = 0$ db 2-es szerepel, azaz keresett index $3+1 = 4$, a keresett szorzótag pedig $2^{1+0} = 2$.

Mivel nekünk a fa legalsó szintjén kell összegeznünk, ezért tekintsük egy L mélységű, teljes trinomiális fa leveleit, és az ezeket leíró L hosszúságú kódokat. Az s darab 3-ast, t darab 2-est tartalmazó listák száma $\binom{L}{s} \binom{L-s}{t}$, ezen kódolású levelek indexe $2s + t + L$, együtthatója 2^{s+t} . Mivel az indexek az $[L, 3L]$ tartományban helyezkednek el, ezért minden $z = 2s + t \in [0, 2L]$ -re meg kell határoznunk az együtthatók összegét. Ehhez tekintsük az L . szinten lévő összes együttható összegét, majd rendezzük át úgy a szummákat, hogy $z = 2s + t$ szerint menjen az összegzés:

$$\sum_{s=0}^L \sum_{t=0}^{L-s} \binom{L}{s} \binom{L-s}{t} 2^{s+t} \quad (41)$$

Alkalmazva a $z = 2s + t$ helyettesítést, $t = z - 2s$ és $z \in [2s, L + s]$:

$$\sum_{s=0}^L \sum_{z=2s}^{L+s} \binom{L}{s} \binom{L-s}{z-2s} 2^{z-s} \quad (42)$$

Itt még szükség van a szummák sorrendjének felcserélésére, hogy a megfelelő alakot kapjuk. Ehhez tekintsünk két esetet: ha $z = 0 \dots L$, akkor $s = 0 \dots \lfloor \frac{z}{2} \rfloor$, továbbá ha $z = L + 1 \dots 2L$, akkor $s = z - L \dots \lfloor \frac{z}{2} \rfloor$. Összefoglalva $s = \max(0, z - L) \dots \lfloor \frac{z}{2} \rfloor$, és ennek segítségével a szummacsere után (42):

$$\sum_{z=0}^{2L} \sum_{s=\max(0, z-L)}^{\lfloor \frac{z}{2} \rfloor} \binom{L}{s} \binom{L-s}{z-2s} 2^{z-s}, \quad (43)$$

azaz tetszőleges $z \in [0, 2L]$ -re a $z + L$ indexű integráltag együtthatója:

$$\sum_{s=\max(0, z-L)}^{\lfloor \frac{z}{2} \rfloor} \binom{L}{s} \binom{L-s}{z-2s} 2^{z-s} \quad (44)$$

Mivel az általunk vizsgált fa $k - 1$ mélységű, így a fenti eredményt alkalmazva kapjuk a keresett F_i együtthatókat:

$$F_i = \sum_{s=\max(0, i-(k-1))}^{\lfloor \frac{i}{2} \rfloor} \binom{k-1}{s} \binom{k-1-s}{i-2s} 2^{i-s} \quad (45)$$

Utolsó lépésként pedig ki kell számítani a megmaradt, egyváltozós hatványfüggvények integrálját:

$$\begin{aligned}
& \int_{t_1=1}^{\infty} \int_{t_2=1}^{\infty} \cdots \int_{t_k=1}^{\infty} \frac{t_1^2 \cdot t_2^2 \cdot \dots \cdot t_k^2}{\{t_1 + t_2 + \dots + t_k - (k-1)\}^{p+k}} dt_k \dots dt_2 dt_1 = \\
&= \int_{t_1=1}^{\infty} t_1^2 \cdot \sum_{i=0}^{2k-2} F_i I_{k-1+i}^{k-1} dt_1 = \sum_{i=0}^{2k-2} F_i \int_{t_1=1}^{\infty} t_1^2 I_{k-1+i}^{k-1} dt_1 = \\
&= \sum_{i=0}^{2k-2} F_i \int_{t_1=1}^{\infty} \frac{t_1^2}{(p+k-1)(p+k-2)\dots(p+k-(k-1+i))(t_1)^{p+k-(k-1+i)}} dt_1 = \\
&= \sum_{i=0}^{2k-2} F_i \frac{1}{(p+k-1)(p+k-2)\dots(p+1-i)} \left[\frac{1}{-(p-2-i)t_1^{p-2-i}} \right]_{t_1=1}^{\infty} = \\
&= \sum_{i=0}^{2k-2} F_i \frac{1}{(p+k-1)(p+k-2)\dots(p+1-i)(p-2-i)}
\end{aligned}$$

Az elvégzett lépések során végig kihasználtam a $k > 2p$ feltételt, hiszen így minden improprius integrál konvergens volt.

Összerakva a kapott eredményeket, a keresett várható érték a következő zárt formulával adható meg:

$$\mathbb{E}((X_1 \cdot X_2 \cdot \dots \cdot X_k)^2) = \sum_{i=0}^{2k-2} F_i \frac{a_1^2 \dots a_k^2 \cdot p(p+1) \dots (p+k-1)}{(p+k-1)(p+k-2)\dots(p+1-i)(p-2-i)}$$

□

3.14. Megjegyzés. Az előző állításban $k = 1$ esetén a $(p+k-1)(p+k-2)\dots(p+1-i)$ szorzat értéke alatt 1-et értek. Ezzel a jelöléssel visszkapjuk az 1-dimenziós Pareto-eloszlás második momentumát.

Továbbá fontosnak tartottam, hogy az előző állításban kapott formula helyességét ismét ellenőrizsem számítógép segítségével $k = 2, 3$ esetén. Szintén a Wolfram Mathematica szimbolikus programcsomag segítségével kiszámítottam a második momentumokat, és a kapott eredmények itt is megegyeztek a formulából kiszámolható értékekkel.

Az előző két állításban kiszámítottam a koordináták szorzatának várható értékét, illetve második momentumát, amely az IBNR (illetve RBNP) károk tartalékszükségletét adja meg. Az RBNS károk esetében viszont a kárfejlődési vektor egy része már ismert, így az összkárkifizetést az ismert koordináták függvényében kell megadni. A következő állítás ebben nyújt segítséget:

3.15. Állítás. Tegyük fel, hogy az $\mathbf{X} = (X_1, X_2, \dots, X_k)$ vektor többdimenziós Pareto-eloszlású $\mathbf{a} = (a_1, \dots, a_k)$ és p paraméterekkel. Ekkor a $p > k - m$ feltétel teljesülése esetén az utolsó $k - m$ koordináta szorzatának feltételes várható értéke az első m koordinátára nézve a következőképpen kapható meg:

$$\begin{aligned} \mathbb{E}(X_{m+1} \cdots X_k | (X_1, \dots, X_m)) &= \\ &= \sum_{i=0}^{k-m-1} G_i \cdot \frac{a'_{m+1;m} \cdots a'_{k;m} \cdot (p+m) \cdots (p+k-1)}{(p+k-1) \cdots (p+m-i)} \left(\frac{a_{m+1}}{a'_{m+1;m}} + \frac{1}{(p+m-i-1)} \right), \end{aligned} \quad (46)$$

ahol az $a'_{i;m}$ módosító paraméterek a 3.6. állításban leírtak alapján számíthatóak, valamint a G_i együtthatók a következő eljárás alapján kaphatóak meg:

1. Vegyük az $\mathcal{A} = \left(\frac{a_k}{a'_{k;m}}, \dots, \frac{a_{m+2}}{a'_{m+2;m}} \right)$ halmaz összes $k - m - 1 - i$ elemű részhalmazát. Ha az \mathcal{A} halmaz üres, akkor $G_i = 1$.
2. Ha \mathcal{A} nem üres, akkor számítsuk ki minden egyes részhalmazon belül az elemek szorzatát.
3. A kapott szorzatokat összegezzük, és jelöljük az így kapott együtthatót G_i -vel.

Bizonyítás. A bizonyítás a 3.10. állítás bizonyításának analógiája. A 3.6. állításban belátottak miatt a feltételes eloszlás egy $k - m$ dimenziós módosított Pareto-eloszlás $\mathbf{a} = (a_{m+1}, \dots, a_k)$ elhelyezkedés, $p + m$ lecsengés és $\mathbf{a}' = (a'_{m+1;m}, \dots, a'_{k;m})$ módosító paraméterekkel, valamint C_m eltolással. Az $a'_{i;m} = a'_i$ egyszerűsítő jelölést alkalmazva a feltételes sűrűségfüggvény a következő alakban írható fel:

$$f(x_{m+1}, \dots, x_k | x_1, \dots, x_m) = \frac{(p+m) \cdots (p+k-1)}{\left(\prod_{i=m+1}^k a'_i \right) \left(\sum_{i=m+1}^k \frac{x_i}{a'_i} + C_m \right)^{p+k}}$$

Ennek felhasználásával a feltételes várható érték:

$$\begin{aligned} \mathbb{E}(X_{m+1} \cdots X_k | (X_1, \dots, X_m)) &= \\ &= \int_{x_{m+1}=a_{m+1}}^{\infty} \cdots \int_{x_k=a_k}^{\infty} x_{m+1} \cdots x_k \cdot f(x_{m+1}, \dots, x_k | x_1, \dots, x_m) dx_k \cdots dx_{m+1} = \\ &= \int_{a_{m+1}}^{\infty} \cdots \int_{a_k}^{\infty} \frac{x_{m+1} \cdots x_k (p+m) \cdots (p+k-1)}{\left(\prod_{i=m+1}^k a'_i \right) \left(\sum_{i=m+1}^k a_i'^{-1} x_i + C_m \right)^{p+k}} dx_k \cdots dx_{m+1} \end{aligned} \quad (47)$$

Alkalmazzuk a $t_i = \frac{x_i}{a_i}$ helyettesítést minden $i > m$ -re, ekkor $t_i \in (\frac{a_i}{a_i}, \infty)$ és a Jacobi-determináns $|\mathbf{J}| = a'_{m+1} \cdot a'_{m+2} \cdot \dots \cdot a'_k$ értéke miatt a (47) integrál a következő alakban írható fel:

$$a'_{m+1} \dots a'_k \int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} \dots \int_{t_k=\frac{a_k}{a'_k}}^{\infty} \frac{t_{m+1} \dots t_k (p+m) \dots (p+k-1)}{\left(\sum_{i=m+1}^k t_i + C_m \right)^{p+k}} dt_k \dots dt_{m+1}$$

Tekintsük most is a konstansok nélküli, $k - m$ dimenziós integrált:

$$\int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} \dots \int_{t_k=\frac{a_k}{a'_k}}^{\infty} \frac{t_{m+1} \dots t_k}{\left(\sum_{i=m+1}^k t_i + C_m \right)^{p+k}} dt_k \dots dt_{m+1}$$

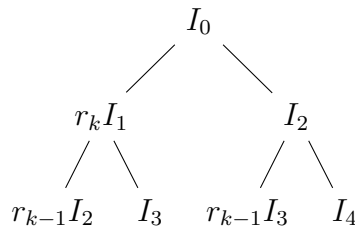
A fenti integrál a 3.10. állítás bizonyításának mintájára, rekurzív módon számítható ki. Ehhez vezessük be a következő jelölést:

$$I_n^j = \frac{1}{(p+k-1) \dots (p+k-n) \left(t_{m+1} + \dots + t_{k-j} + \sum_{s=k-j+1}^k \frac{a_s}{a'_s} + C_m \right)^{p+k-n}}$$

Egy általános tag integrálása után belátható a következő rekurziós formula:

$$\int_{t_{k-j}=\frac{a_{k-j}}{a'_{k-j}}}^{\infty} t_{k-j} I_n^j dt_{k-j} = \frac{a_{k-j}}{a'_{k-j}} I_{n+1}^{j+1} + I_{n+2}^{j+1},$$

A 3.13. állítás bizonyításának mintájára itt is egy faszerkezet segítségével lehet meghatározni az egyes szinteken az együtthatók összegét. Mivel itt minden tagnak két utódja lesz, így egy binomiális fával kell dolgoznunk. Jelölje $r_{k-j} = \frac{a_{k-j}}{a'_{k-j}}$ -t, így a keresett fa:



A fa gyökerében I_0 , azaz a keresett $k - m$ dimenziós integrál helyezkedik el. Minden $\int_{t_{k-j}=\frac{a_{k-j}}{a'_{k-j}}}^{\infty} t_{k-j} I_n^j dt_{k-j}$ alakú integrálás után két utód keletkezik, így egy teljes binomiális fát kapunk.

A továbbiakban vizsgáljunk egy L mélységű fát. A j . szinten a következő integrálok fordulnak elő: $I_j, I_{j+1}, \dots, I_{2j}$, így a cél az, hogy az alsó szinten meghatározzam az ezekhez tartozó együtthatók összegét.

Vezessük be a következő jelöléstrendszert: legyen minden csúcs bal oldali gyermeke 1-es, míg a jobb oldali gyermeke 2-es számú. Ezek a kódok adják meg, hogy a szülő indexéhez képest mennyivel nagyobb a gyerek indexe, valamint a szülő együtthatóját mivel kell megszorozni, hogy megkapjuk az adott gyerek együtthatóját: 1-es esetben r_{k-j+1} -gyel, míg 2-es esetben 1-gyel.

Mivel teljes binomiális fával dolgozunk, így a j . szinten minden csúcs indexe és együtthatója megadható egy j hosszúságú, csak az $(1,2)$ számokat tartalmazó listával, aminek m . eleme a csúcs m . szinten lévő együtthatójának kódja. Jelölje s a 2-esek számát a kódban, $L - s$ pedig az 1-esek számát a kódban, feltéve hogy $s \in [0, L]$. Így minden j . szinten lévő csúcs indexe a lista elemeinek összege. A csúcs együtthatója nehezebben határozható meg, mivel a különböző szinteken különböző r_{k-j} -vel szorzódott az addigi együttható. Például a bemutatott fa 3. szintjén az egyik I_3 -as tag együtthatója $r_{k-1} \cdot 1$, míg a másiké $1 \cdot r_{k-2}$. Azonban a kódból kiolvasható, hogy az $L - s$ indexű tagok s alkalommal szorzódtak meg valamilyen r_{k-j} -vel az öröklődés során, így az együtthatójuk megkapható, ha összeszorozzuk a kódban a 2-es indexű tagokhoz tartozó r_{k-j} -ket. Azaz tekintsük a csúcs kódjában azokat az elemeket, amik értéke 2, legyen ennek indexe m_i . Ekkor az ehhez tartozó együttható értéke r_{k-m_i} . A csúcs együtthatója pedig ezen r_{k-m_i} -k szorzataként számolható.

Összefoglalva, a fa L . szintjén az adott $s = 0, \dots, L$ indexhez tartozó együtthatók a következő módszerrel határozhatók meg:

1. Vegyük az (r_k, \dots, r_{k-L+1}) halmaz összes $L - s$ elemű részhalmazát. Ha ez a halmaz üres, akkor legyen $G_s = 1$.
2. Ha a fenti halmaz nem üres, akkor számítsuk ki minden egyes részhalmazon belül az elemek szorzatát.
3. A kapott szorzatokat összegezzük, és jelöljük az így kapott együtthatót G_s -sel.

A fenti eljárás segítségével megkapható minden keresett együttható egy L mélységű fában. Mivel az általam vizsgált fa $k - m - 1$ mélységű, így az eredményeket $L = k - m - 1$ -re alkalmazva kimondható a következő lemma:

3.16. Lemma. *A fenti jelölésekkel:*

$$\int_{\frac{a_{m+2}}{a_{m+2}}}^{\infty} \dots \int_{\frac{a_k}{a_k}}^{\infty} t_{m+2} \cdot \dots \cdot t_k \frac{1}{(t_{m+1} + \dots + t_k + C_m)^{p+k}} dt_k \dots dt_{m+2} = \sum_{i=0}^{k-m-1} G_i \cdot I_{k-m-1+i}^{k-m-1}$$

A rekurzió befejező lépéseként a fenti lemma eredményét alkalmazva kapjuk, hogy:

$$\begin{aligned}
& \int_{\frac{a_{m+1}}{a'_{m+1}}}^{\infty} \dots \int_{\frac{a_k}{a'_k}}^{\infty} \frac{t_{m+1} \cdot \dots \cdot t_k}{(t_{m+1} + \dots + t_k + C_m)^{p+k}} dt_k \dots dt_{m+1} = \\
& = \int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} t_{m+1} \sum_{i=0}^{k-m-1} G_i I_{k-m-1+i}^{k-m-1} dt_{m+1} = \\
& = \sum_{i=0}^{k-m-1} G_i \int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} t_{m+1} I_{k-m-1+i}^{k-m-1} dt_{m+1} =
\end{aligned}$$

Tekintsük a szumma egy tagját a G_i együttható nélkül!

$$\begin{aligned}
& \int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} \frac{t_{m+1}}{(p+k-1) \dots (p+m+1-i) \left(t_{m+1} + \sum_{s=m+2}^k \frac{a_s}{a'_s} + C_m \right)^{p+m+1-i}} dt_{m+1} = \\
& = \left[\frac{t_{m+1}}{(p+k-1) \dots (p+m-i) \left(t_{m+1} + \sum_{s=m+2}^k \frac{a_s}{a'_s} + C_m \right)^{p+m-i}} \right]_{\frac{a_{m+1}}{a'_{m+1}}}^{\infty} - \\
& - \int_{t_{m+1}=\frac{a_{m+1}}{a'_{m+1}}}^{\infty} \frac{1}{(p+k-1) \dots (p+m-i) \left(t_{m+1} + \sum_{s=m+2}^k \frac{a_s}{a'_s} + C_m \right)^{p+m-i}} dt_{m+1} = \\
& = \frac{\frac{a_{m+1}}{a'_{m+1}}}{(p+k-1) \dots (p+m-i) \left(\sum_{s=m+1}^k \frac{a_s}{a'_s} + C_m \right)^{p+m-i}} + \\
& + \frac{1}{(p+k-1) \dots (p+m-i)(p+m-i-1) \left(\sum_{s=m+1}^k \frac{a_s}{a'_s} + C_m \right)^{p+m-i-1}}
\end{aligned}$$

Végül pedig kihasználva a 3.7. megjegyzésben megállapítottakat, azaz $\sum_{s=m+1}^k \frac{a_s}{a'_s} + C_m = 1$, a rekurzió befejező lépéséből a szumma egy tagja:

$$\begin{aligned}
& \frac{\frac{a_{m+1}}{a'_{m+1}}}{(p+k-1) \dots (p+m-i)} + \frac{1}{(p+k-1) \dots (p+m-i)(p+m-i-1)} = \\
& = \frac{1}{(p+k-1) \dots (p+m-i)} \left(\frac{a_{m+1}}{a'_{m+1}} + \frac{1}{p+m-i-1} \right)
\end{aligned}$$

A bizonyítás során végig kihasználtam a $p > k - m$ feltételt, hiszen így minden improprius integrál konvergens volt.

Végül összerakva a kapott eredményeket, a keresett várható érték a következő alakot ölti:

$$\begin{aligned} \mathbb{E}(X_{m+1} \cdot \dots \cdot X_k | (X_1, \dots, X_m)) &= \\ &= \sum_{i=0}^{k-m-1} G_i \cdot \frac{a'_{m+1} \dots a'_k \cdot (p+m) \dots (p+k-1)}{(p+k-1) \dots (p+m-i)} \left(\frac{a_{m+1}}{a'_{m+1}} + \frac{1}{p+m-i-1} \right) \end{aligned}$$

□

3.17. Megjegyzés. A 3.15. állítást $m = 0$ -ra alkalmazva (azaz a k -dimenziós kiinduló vektort nézve, feltétel nélkül) visszakapjuk a 3.10. állítás eredményeit. Ez egyrészt $a'_{i;0} = a_i$ miatt teljesül, másrészt a G_i együtthatók meghatározásánál a $k - 1$ elemű, csupa 1-est tartalmazó halmaz részhalmazait kell vennünk, az i elemű részhalmazok száma $\binom{k-1}{i}$, az ebben lévő elemek szorzata pedig 1. Ezeket az értékeket szorzat várható értékének formulájába ((46) egyenlet) helyettesítve könnyen látható, hogy a feltétel nélküli vektorra kapott eredményhez ((28) egyenlet) jutunk.

A feltételes várható érték második momentumának meghatározása analitikusan nehezebb feladat, illetve még bonyolultabb képleteket és eljárásokat eredményezne, így arra nem térek ki.

4. Az adatok jellemzése

Az előző fejezetekben bemutatam a modellt, valamint a vizsgálatok során használt Pareto-eloszlást, és annak tulajdonságait. A következő fejezetekben pedig megvizsgálom a bemutatott elméleti keretrendszer gyakorlatban való alkalmazását, illetve a modellem illeszkedését valós káradatokon.

Az elemzéshez egy biztosító nem-életbiztosítási adatait használom fel. A számolás egyszerűsítése miatt minden kifizetést elosztottam 100.000-rel (ez csak az a_1 paraméter, azaz az első kifizetés nagyságrendjén változtat, a többi paraméter becslését nem befolyásolja), illetve a tartalékok is 100.000 Ft-szorosai az eredményeknek. Az elérhető adathalmaz 11 év kárkifizéseit mutatja be, ahol a károk az 1. és a 6. év közt következtek be, és közülük mindegyiket lezárták a 11. év végén.

Az előrejelzéshez, illetve a függőkárok tartalékának becsléséhez az eredeti adathalmazból csak azokat a károkat tekintetem, amelyeknek a bejelentési periódusa nem későbbi a 6. évnél. Emellett egy kárra az adott évbe eső kifizetéseket aggregáltam, így a kár fejlődése során ténylegesen a kifizetési periódusok közti növekedést vizsgálom. További egyszerűsítésként az elemzésből kihagytam azokat a károkat, amelyekre volt negatív kifizetés.

Ilyen módon a vizsgált adathalmaz 36599 kárt tartalmaz, amelyek közül 29688 darab lezárt, 6639 darab RBNP (még nem történt kifizetés) és 272 darab RBNS (volt már részkifizetés, de még nem zártuk le) állapotú. A bejelentett károkra legfeljebb 3 kifizetést rögzítettek: 29749 olyan kár van, ahol csak egy kifizetés történt, 210 olyan, ahol pontosan kettő, míg 1 olyan, ahol van harmadik kifizetés is. Így a modellben a kárkifizetések számát 3-ban maximalizálom, azaz az első kifizetés mellett maximum 2 növekedési faktor lehet.

Az első kifizetésekről (Y_1), illetve a növekedési faktorokról (λ_1, λ_2) egy rövid leíró statisztikát ad az alábbi táblázat:

Változó	Átlag	Szórás	Minimum	Maximum	Megf. száma
Y_1	1.218722	1.062091	1.000592	109.688789	29960
λ_1	2.239029	1.477195	1.120636	19.844485	211
λ_2	1.497919	—	1.497919	1.497919	1

1. táblázat. *Az első kárkifizetés és a növekedési faktorok jellemzése.* Forrás: saját számítás

Szükségesnek tartom kiemelni, hogy a harmadik kárkifizetésre csak egy adatom van, így a paraméter becslések során ez nagy bizonytalanságot okozhat.

A mintában az átlagos kárnagyságot és ennek szórását csak a lezárt károkra érdemes meghatározni, mivel ezeknél áll rendelkezésre a teljes fejlődési minta. Az egy kárra történt összkifizetés átlagértéke 1.235138, míg szórása 2.042792. Ezen utóbbi adat elég nagy (az átlaghoz viszonyítva), és ez azt sejteti, hogy a kifizetések eloszlása vastag farkú.

5. Paraméterbecslés

A vizsgálat során a következő cél, hogy a káradatokat jellemző, a 2.1. fejezetben bemutatott változók eloszlásainak ismeretlen paramétereit meghatározzam. Ezt maximum likelihood módszerrel végzem el, a 2.2. fejezetben leírt likelihood függvényt maximalizálom, tagonként. Az optimalizálás során az R programcsomagot használom. A vizsgálatához használt program forráskódja, illetve a nyers adatok a <http://bit.ly/23PJQqZ> linken érhetőek el.

5.1. A kárfejlődési vektor paramétereinek becslése

Első lépésként az elemzés során alkalmazott, a 2.1. definícióban bemutatott többdimenziós Pareto eloszlás ismeretlen \mathbf{a} elhelyezkedésvektor, illetve p lecsengés paramétereinek becslését határozom meg. Ezt a 3. fejezetben bemutatott maximum likelihood becslés eljárás (a 3.8. állítás) segítségével hajtom végre.

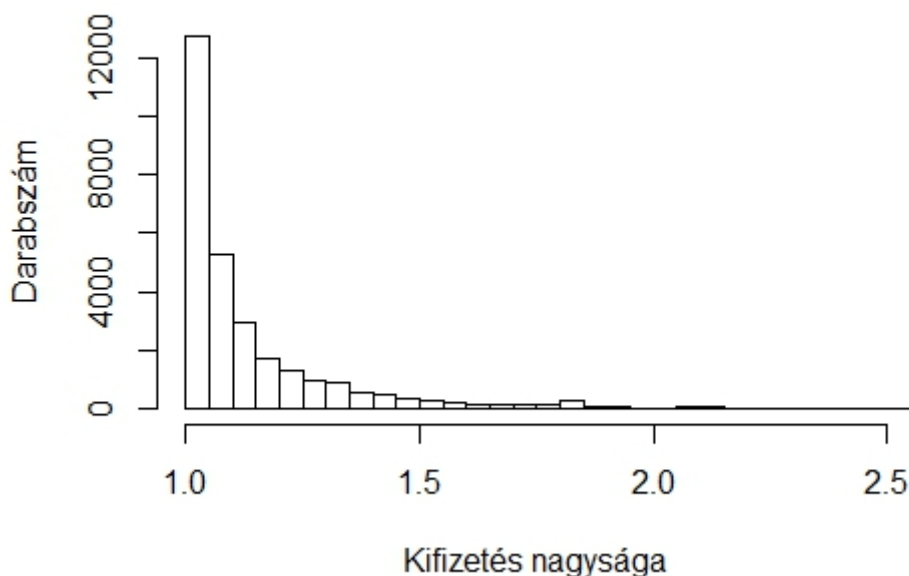
Először meghatározom az elhelyezkedés vektor paramétereit, amely a mintában az adott koordinátára az adatok minimuma. Azaz \hat{a}_1 adja meg a legkisebb ismert első kifizetést, \hat{a}_2 a legkisebb ismert első növekedési faktort, míg \hat{a}_3 a legkisebb második növekedési faktort. Ezek rendre: $\hat{a}_1 = 1.000592$, $\hat{a}_2 = 1.120636$ és $\hat{a}_3 = 1.497919$. A kapott becsléseket a p paraméter meghatározására kapott, numerikusan megoldható egyenletbe ((25) egyenlet) helyettesítve a lecsengés paraméter becslésére $\hat{p} = 6.738437$ adódik.

Következő lépésként meg kell vizsgálni, hogy a kapott $\hat{\mathbf{a}}$ és \hat{p} paraméter becslések optimálisak-e, azaz kielégítik-e az (24)-ben leírt egyenlőtlenség-rendszert. A behelyettesítések után azt kaptam, hogy minden feltétel teljesült, azaz nem tudunk javítani a likelihood függvény ezen tagján. Más szavakkal, a kezdő lépésben kapott paraméter becslések optimálisak.

A p -re kapott eredmény a modellbeli analitikus eredmények szempontjából jónak mondható, hiszen maximálisan 3 kifizetést feltételezve, teljesülnek a $p > U + 1$ és $p > 2(U + 1)$ feltételek ($U + 1$ a kárfejlődési vektor dimenziója), így az adatokra illesztett többváltozós Pareto-eloszlás véges várható értékű, illetve véges szórású. Érdeemes megemlíteni azt is, hogy a \hat{p} viszonylag magas értéke gyors lecsengésű eloszlásra utal. Ezt a megfigyelést alátámasztja a 3. ábra is, amely az első kifizetések hisztogramját mutatja.

5.2. Az időváltozók paramétereinek becslése

Második lépésként a 2.1.2. fejezetben bemutatott változók ismeretlen paramétereit becslöm meg a mintából. A likelihood függvényt szintén tagonként maximalizálom: minden



3. ábra. Az első kifizetések hisztogramja. Forrás: saját ábra

változóra diszkrét eloszlásokat használok, majd ezek paramétereinek változtatásával optimalizálom az adott likelihood-tagot. Az adatokra legjobban illeszkedő eloszlást az Akaike, illetve a Bayesi információs kritériumok alapján határozom meg.

5.1. Megjegyzés. A vizsgálatok során az R programcsomag beépített függvényeivel dolgozom. A használt diszkrét eloszlások eloszlásfüggvénye csak a negatív binomiális eloszlásnál tér el a klasszikus alaktól. A programban szereplő valószínűség eloszlás az (r, p) paraméterű negatív binomiális eloszlásra:

$$f(x) = \frac{\Gamma(x+r)}{\Gamma(x)} p^r (1-p)^x,$$

ahol $x = 0, 1, \dots$ és $0 < p \leq 1$. Az eloszlás várható értéke pedig $\frac{r(1-p)}{p}$.

5.2.1. Bejelentési késés

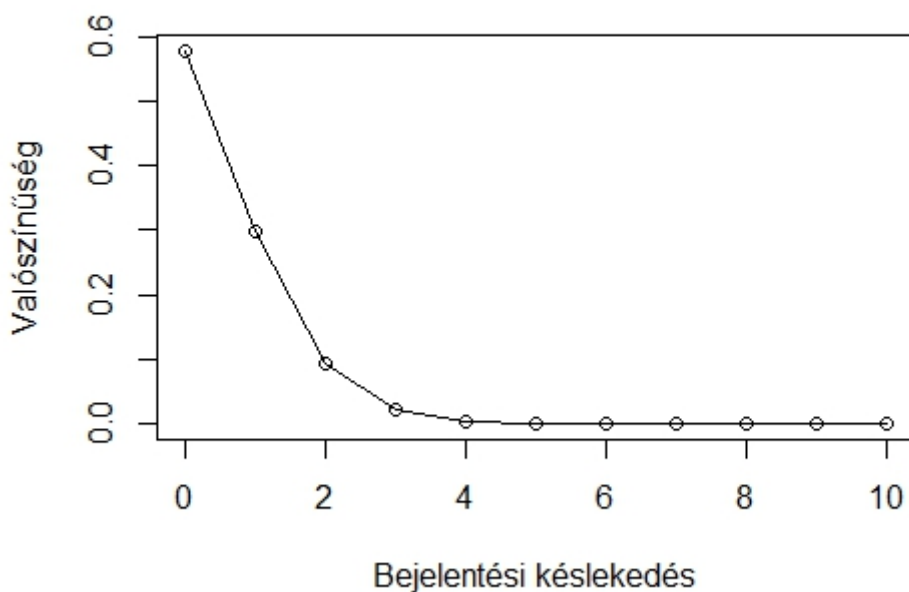
A bejelentési késés mind a 36599 bejelentett kárnál rendelkezésre áll, és értéke 0 és 5 között mozog (mivel a 6. periódus végén végzem az értékelést). A késést megadó T változó $F_1(t)$ eloszlására geometriai, Poisson, binomiális, illetve negatív binomiális eloszlást illetve, az információs kritériumok értékei az alábbi táblázatban láthatóak:

	Geometriai	Poisson	Binomiális	Neg. Binom.
AIC	75856.9	74828	-	74549.92
BIC	75865.41	74836.51	-	74566.93

2. táblázat. Az információs kritériumok értékei a bejelentési késére a különböző diszkrét eloszlások esetén. Forrás: saját számítás

A binomiális eloszlásra az optimalizálás nem konvergált, így a maradék 3 eloszlástípus közül mindkét információs kritérium alapján a negatív binomiális eloszlás illeszkedik a legjobban, melynek paraméterei $r_T = 4.4993076$ és $p_T = 0.8855756$.

Az eloszlás várható értéke 0.5965587, ami azt jelzi, hogy várhatóan a károk nagy részét 0 vagy 1 periódus késlekedéssel jelentik be, tehát a kár bekövetkezése után rövid a bejelentési késlekedés. A választott negatív binomiális eloszlást szemlélteti a 4. ábra.



4. ábra. A bejelentési késés becsült eloszlása. Forrás: saját ábra

5.2.2. Első fizetési késés

A mintában 29960 olyan kár szerepel, amire már ismert az értékelés pillanatában az első kifizetés (az RBNP károkra ezek még nem állnak rendelkezésre), és ezen késés értékei 0 és 5 között mozognak. A késést megadó Q változó $F_2(t)$ eloszlására geometriai, Poisson,

binomiális, illetve negatív binomiális eloszlást illesztve, az információs kritériumok értékei az alábbi táblázatban láthatóak:

	Geometriai	Poisson	Binomiális	Neg. Binom.
AIC	-	138456.8	151596.7	-
BIC	-	138465.1	151613.3	-

3. táblázat. *Az információs kritériumok értékei az első fizetési késére a különböző diszkrét eloszlások esetén.* Forrás: saját számítás

Az optimalizálás során csak a binomiális, illetve a Poisson-eloszlásra konvergált a maximumkereső eljárás. A táblázat eredményei azt mutatják, hogy az első kifizetési késés Poisson-eloszlást követ $\lambda = 1.703278$ paraméterrel, így várható értéke 1.703278. Így elmondható, hogy várhatóan a bejelentést követő első vagy második periódusban megtörténik az első kifizetés.

5.2.3. Kifizetések száma az első kifizetés után

A mintában csak 211 darab olyan kifizetés szerepel, ahol az első kifizetést követte további is, így viszonylag kis minta áll rendelkezésre ezen valószínűségi változó eloszlásának becslésére. Ezen U változó $F_3(\cdot)$ eloszlására geometriai, Poisson, binomiális, illetve negatív binomiális eloszlást illesztve, az információs kritériumok értékei az alábbi táblázatban láthatóak:

	Geometriai	Poisson	Binomiális	Neg. Binom.
AIC	4955.352	4959.006	-	4956.725
BIC	4963.659	4967.314	-	4973.34

4. táblázat. *Az információs kritériumok értékei az első fizetési utáni kifizetések darabszámára a különböző diszkrét eloszlások esetén.* Forrás: saját számítás

Az optimalizálás során a binomiális eloszlásra nem konvergált a maximumkereső eljárás, így csak a maradék 3 eloszlásra határoztam meg az információs kritériumok értékeit. A táblázat eredményei azt mutatják, hogy az első kifizetés utáni kifizetések darabszáma geometriai eloszlást követ, melynek paramétere $p_U = 0.9838264$, és várható értéke 0.01643952. A különböző eloszlásokra kapott információs kritérium értékek azonban nem térnek el szignifikánsan. A geometriai eloszlás paraméterei az mutatják, hogy nagy valószínűséggel az első kifizetést már nem követik továbbiak, ami összhangban van azzal, hogy a mintában 29749 kárra volt csak egy kifizetés, míg 211 darabra legalább kettő.

5.2.4. Kifizetési késlekedések

A mintában maximum 3 kifizetéses károk voltak, így az első kifizetés utáni első, illetve második kifizetés késésére adható csak becslés. 211 megfigyelés található a második kifizetés (N_1) késésére, míg 1 megfigyelés a harmadik kifizetés (N_2) késésének becslésére.

Fontos kiemelni, hogy amíg a többi időváltozó (bejelentési késés, első fizetési késés, valamint az első kifizetés utáni többi kifizetés darabszáma) *nemnegatív*, egész értékű valószínűségi változó (azaz 0 is lehet az értéke), addig az első utáni kifizetések késése *pozitív*, egész értékű valószínűségi változó (azaz 0 nem lehet az értéke). Emiatt a likelihood függvény maximalizálása során az késések eltoltt függvényére, azaz $N_i - 1$ -re illeszttem az eloszlásokat (ez már nemnegatív változó), ahol $i = 1, 2$.

A második késés eltoltt eloszlására geometriai, Poisson, binomiális, illetve negatív binomiális eloszlást illesztve az információs kritérium értékek adódnak:

	Geometriai	Poisson	Binomiális	Neg. Binom.
AIC	448.5755	438.2267	-	438.1363
BIC	456.8741	446.5253	-	446.4349

5. táblázat. *Az információs kritériumok értékei a második fizetési késésre a különböző diszkrét eloszlások esetén.* Forrás: saját számítás

Az optimalizálás során a binomiális eloszlásra nem konvergált a maximumkereső eljárás, így csak a maradék 3 eloszlásra határoztam meg az információs kritériumok értékeit. A táblázat eredményei azt mutatják, hogy kis különbséggel ugyan, de a negatív binomiális eloszlás illeszkedik a legjobban. Ennek paraméterei $r = 17.870478$ és $p = 0.966587$, ezek alapján a várható értéke pedig 0.6177471. Így az első kifizetés utáni első kifizetés várható késése $0.6177471 + 1 = 1.6177471$, ami azt mutatja, hogy várhatóan az első kifizetést követő 1-2 periódusban megtörténik a következő is (ha van több kifizetés).

A harmadik késésre csak egy megfigyelésem van, így arra nem érdemes eloszlást illeszteni. Ennek konkrét értéke $N_2 = 2$. Nagyobb minta esetén ennek eloszlására is illeszthető lenne egy becslés.

5.3. A kárbekövetkezési intenzitás becslése

A károkat jellemző változók ismeretlen paramétereinek becslése után szükséges, hogy megbecsüljük az IBNR károk várható darabszámát is. A 2.1.4. fejezetben leírtak szerint az IBNR károk darabszáma Poisson-eloszlást követ $\Theta w(i) (1 - F_1(t_i^* - 1))$ paraméterrel.

A vizsgálatok során sajnos az állomány mérete, és a kockázatban álló szerződések száma nem állt a rendelkezésemre, csak a káradatok. Így az IBNR károk becslését nem tudom elvégezni a Poisson-eloszlás alapján. Ehelyett a lánc-létra módszert használom a még be nem jelentett károk darabszámának becslésére, amelyre 6315.201 adódik.

6. A függőkárok tartalékának becslése, összehasonlítás

Az alábbi fejezetben az 5. fejezetben kapott, becsült paraméterek segítségével és a 2.3. fejezetben bemutatott analitikus eredmények felhasználásával megbecsülöm a függőkárok tartalékszükségletét. A kapott eredményeket összevetem a lánc-létra módszer által becsült tartalékokkal, valamint a ténykárfizetéssel. Ezek alapján vizsgálom a modellem jóságát és illeszkedését az adatokra, és ahol lehet, az eltéréseket megmagyarázom.

6.1. A függőkárok tartalékának becslése

Az IBNR és RBNP károk tartalékszükségletének becsléséhez két adatra van szükségünk: egy kárkifizetés várható értéke, valamint a károk (várható) darabszáma. Egy kárkifizetés várható értékét a 2.3. tétel alapján kaphatjuk meg, felhasználva, hogy az U változó geometriai eloszlású. A becsült paraméterek behelyettesítése után a keresett várható érték 1.181. Ez némileg alatta marad lezárt károk átlagának (1.235), de az eltérés csak 5% körüli.

Az IBNR károk várható darabszáma az előző fejezet alapján 6315.201, míg az RBNP károk ismert darabszáma 6639. Ezek alapján az IBNR károk tartalékszükséglete 7458.253, míg az RBNP károké 7843.197.

Az RBNS károkra a fejlődési minta egy része már ismert, így csak a többi, ismeretlen kifizetés várható értékére kell tartalékolnunk. Ezen típusnál minden egyes kárra megbecsülöm a várható összkifizetést a 2.5. tétel alapján, majd ezekből a tartalékszükségletet az adott kárra. Végül a kapott eredményeket összeadom.

A mintában összesen 272 darab RBNS kár szerepelt, ezek közül 268-ra volt csak egy ismert kifizetés, míg 4 darabra két ismert kifizetés. Minden kárra az egyéni paraméterek és együtthatók kiszámítása után az RBNS károk várható összertalék szükségletére 333.764 adódott.

Így a modellben a függőkárok teljes tartalékszükséglete 15635.21.

6.2. Tényadatok és becslés lánc-létra módszerrel

A rendelkezésemre álló adatokban olyan károk szerepeltek, amelyek az 1. és a 6. év között következtek be, de az 1. és a 11. év között jelentettek be a biztosító társaságnak. Így a kifizetések egy teljes kárkifizési négyzetet alkottak, amelyet a 6. táblázat foglal össze. Az ebben látható adatok kumulált kárkifizetéseket tartalmaznak, a félkövér értékek mutatják a még nem ismert kifizetéseket.

	0	1	2	3	4	5
1	1143.747	5372.256	6881.659	7722.883	8288.413	8933.527
2	615.515	3609.088	5477.649	6169.22	6785.449	7155.399
3	396.59	4528.703	6484.631	7528.212	8225.146	8845.121
4	916.06	5128.805	7806.32	8942.934	9942.753	10332.818
5	546.209	5345.121	7710.437	9286.894	11461.247	11954.056
6	832.193	5534.142	8290.022	9328.821	10064.728	10485.066

6. táblázat. *A ténykárfizetések a bejelentési év és a bejelentési késés szerint*

Ezekből a vizsgálat során csak azokat vettem figyelembe, amelyeknek a kárbejelentési periódusa nem nagyobb 6-nál. Ezáltal egy felső kárfizetési háromszöget kaptam, amely a modellem tényadataiként szolgált.

A rendelkezésre álló háromszögből a lánc-létra módszer segítségével is kiszámítottam a függőkárok tartalékszükségletét, ezt foglalja össze a 7. táblázat. A félkövér értékek adják meg a becsült kumulált kárfizetéseket a bejelentési késés függvényében.

	0	1	2	3	4	5
1	1143.747	5372.256	6881.659	7722.883	8288.413	8933.527
2	615.515	3609.088	5477.649	6169.22	6785.449	7313.582
3	396.59	4528.703	6484.631	7528.212	8168.614	8804.404
4	916.06	5128.805	7806.32	8873.614	9628.465	10377.88
5	546.209	5345.121	7642.577	8687.483	9426.501	10160.196
6	832.193	5516.48	7887.591	8965.996	9728.706	10485.922

7. táblázat. *A lánc-létra módszer által becsült kumulált kárfizetési háromszög.* Forrás: saját számítás

A táblázatokból kiszámolható, hogy a tény tartalékszükséglet 20475.165 lenne, míg a lánc-létra módszer által becsült érték, 18844.689 némileg alatta marad ennek (az eltérés kb. 8% a tényadatokhoz viszonyítva).

6.3. Összehasonlítás, az eltérések magyarázata

Az általam vizsgált modellben az össztartalék szükséglet (az IBNR, RBNP és RBNS károk tartalékának összegzése után) 15635.21 adódott, míg a lánc-létra módszer által becsült érték 18844.689, a tényadatokból kiolvasott tartalék pedig 20475.165 lenne. Látható,

hogy a modell által becsült érték alatta marad mind a lánc-létra becslésnek, mind a tény eredményeknek. Az eltérés a tényadatoktól 24% körüli, míg a lánc-létra módszer eltérésre 8% körüli. Ezért szükségesnek látom, hogy megvizsgáljam az eltérések okát, illetve a modellem jóságát.

A legnagyobb eltérést az okozza, hogy az RBNP státuszú károk (azaz 6. év végéig bejelentett, de kifizetés nélküli károk) átlagkifizetése (1.86) szignifikánsan, kb. 1.5-szer nagyobb a lezárt károk átlagkifizetésénél (1.235). Mivel a modellemben a "kifizetési" paraméterek becslését döntően a lezárt károk határozták meg, így az egy kárra várt átlagkifizetés (1.181) is ezekhez igazodik. Mivel ez az eltérés minden RBNP kárra jelentkezik, így összességében a tényadatoktól való eltérés jelentős részét magyarázza ez a jelenség.

Az eltéréshez hozzátesz még az IBNR károk tartalékának becslése is. Ezen károk darabszáma a tényadatok alapján 6533 volt, míg a lánc-létra módszerrel becsült érték 6315.201. Az IBNR károk becsült tartalékszükséglete 7458.253, míg a tényadatokban ezen kifizetések összértéke 7596.04 volt. Látható, hogy az IBNR károk darabszámát alulbecsüli a módszer, viszont tartalékszükségletükre nem volt szignifikáns eltérés.

Az RBNS károknál is van eltérés a becsült és a tényeredmények közt. Az RBNS károk becsült tartalékszükségletére 333.764 adódott, míg a tényleges tartalékszükséglet 523.538 lett volna. Tehát a modellem alulbecsülte az összkárfizetést, a ténylegestől való eltérés arányaiban nagy, viszont nincs szignifikáns hatása az össztartalékra.

Az eredmények bizonytalanságához hozzájárul még a paraméterbecslések hibája is. Az egynél több kifizetésű károkból az összes ismert kárhoz viszonyítva kevés adat áll rendelkezésre, így ez rontja a becslés pontosságát.

7. Összegzés, további vizsgálati lehetőségek

A szakdolgozatomban bemutattam egy lehetséges, diszkrét idejű tartalékolási modellt. A 2. fejezetben Pigeon et al. (2013) által vizsgált keretrendszert kiindulásként használva, leírtam a vizsgált modellt, annak változóit, valamint analitikus eredményeket adtam a várható összkárkifizetés becslésére. A 3. fejezetben Mardia (1962) nyomán bemutattam a k -dimenziós Pareto-eloszlást, melyet a kárfejlődési vektor eloszlásánál használtam. Röviden megvizsgáltam ennek tulajdonságait, majd a függőkárok tartalékszükségletének becsléséhez új, önálló eredményeket elérve mondtam ki és bizonyítottam 3 állítást. Az utolsó 3 fejezetben pedig egy biztosító nem-életbiztosítási adataira illesztettem a modellem, valamint kiszámítottam a tartalékszükségletet. Az eredményeket összevettem mind a tényadatokkal, mind a lánc-létra módszer által becsült értékekkel, az észlelt eltérések okát pedig megvizsgáltam.

Összességében megállapítható, hogy a modellem alulbecsülte a tartalékszükségletet, az eltérés mértéke 24% körüli, ennél jobb eredményt ad a lánc-létra módszer, amelynek hibája csak 8% (ez is negatív irányban értendő). Így a modellem a vizsgált adathalmazon nem adott jó eredményt. Sajnos az elemzéshez nem tudtam más kárstatisztikát szerezni, azonban elképzelhető, hogy más típusú adatokon sokkal pontosabb becslés kapható a függőkárok tartalékszükségletére.

A modellem legfőbb előnyeként, és egyben legfőbb hátrányaként annak *részletességét* emelném ki. Azáltal, hogy egyéni szinten vizsgáljuk a bekövetkezett károkat, és azok fejlődését, egy sokkal részletesebb és pontosabb képet kaphatunk azok időbeni alakulásáról, illetve összkifizetésükről. Így lehetőség nyílik arra, hogy a tartalékszükségletre jobb becslést adjunk, mint egy aggregált adatokkal dolgozó modell. Más részről viszont meg kell említeni, hogy a részletesség, a viszonylag sok paraméter becslés rengeteg apró hibát von maga után, amely pontatlanabbá teheti az eredményeket.

A modellemben a paraméterek becsléséből is származik hiba, hiszen ezeket az eddig ismert adatok, főként a lezárt károk alapján végeztem el. Egy érdekes képet kaphatnánk akkor, ha a becsült paraméterekből (kárkifizetési vektor, illetve időváltozók) károkat szimulálnánk, majd összevetném, hogy a kapott véletlen minta eloszlása mennyire hasonlít a tényadatok eloszlására. Ennek előfeltétele az, hogy tudjunk többdimenziós Pareto-eloszlású véletlen vektorokat generálni, de ez az eszköztár jelenleg nem áll rendelkezésemre. Így ennek fejlesztése egy kiaknázatlan lehetőség a modellemben.

Az RBNP károkra tapasztalt eltérés okán egy másik hiányosság is felmerült: a modell érzékeny a mintán belüli részminták eloszlásának egyenetlenségére. Ennek oka lehet a kárkifizetések időbeni változása is (egy megfigyelhető trend alapján), de akár naptári

hatások is állhatnak a háttérben. Ezek vizsgálata, illetve figyelembe vétele is további lehetőségeket rejt a modellemben.

Végül pedig érdemes lehet azt is megvizsgálni, hogy a többdimenziós Pareto-eloszlás helyett más többváltozós eloszlást használva pontosabb illeszkedést, illetve pontosabb becslést kaphatunk-e az adatok alapján. Ehhez azonban szükség lenne más eloszlások részletes vizsgálatára, illetve az összkárkifizetést leíró tételek kimondására is.

Összességében úgy vélem, az egyéni módszereken alapuló tartalékolás nagy jövő előtt áll, hiszen az ilyen jellegű modellek alapján pontosabb becslést adhatunk a függőkárok tartalékszükségletére, mint a klasszikus módszerek által adott értékek. Habár a vizsgált adathalmazra nem illeszkedett jól a modell, más kárstatisztikák vizsgálata és a fent említett módszerek fejlesztése révén egy pontosabb, jobban illeszkedő előrejelzést alkothatok meg a jövőben.

Hivatkozások

- [1] M. Pigeon, K. Antonio, M. Denuit (2013): Individual Loss Reserving with the Multivariate Skew Normal Framework, *ASTIN Bulletin*, Vol. 43, No. 03, 399-428.
- [2] T. Mack (1993): Distribution-free calculation of the standard error of chain-ladder reserve estimates, *ASTIN Bulletin* 23, 213-225.
- [3] T. Mack (1999): The standard error of chain-ladder reserve estimates: recursive calculation and inclusion of a tail factor, *ASTIN Bulletin* 29, 361-366.
- [4] R. L. Bornhuetter, R. E. Ferguson (1972): The Actuary and IBNR , *Proceedings of the Casualty Actuarial Society* 59, 181-195.
- [5] E. Arjas (1989): The claim reserving problem in non-life insurance: some structural ideas, *ASTIN Bulletin*, Vol. 19, No. 2, 139-152.
- [6] R. Norberg (1993): Prediction of Outstanding Liabilities in Non-Life Insurance, *ASTIN Bulletin*, Vol. 23, No. 1, 95-115.
- [7] S. Haastrup, E. Arjas (1996): Claims reserving in continuous time; a nonparametric Bayesian approach, *ASTIN Bulletin*, Vol. 26, No. 2, 139-164.
- [8] M. Pigeon, K. Antonio, M. Denuit (2014): Individual loss reserving using paid-incurred data, *Insurance: Mathematics and Economics*, No. 58, 121-131.
- [9] K. V. Mardia (1962): Multivariate Pareto Distributions, *The Annals of Mathematical Statistics*, Vol. 33, No. 3, 1008-1015.
- [10] H. Rootzén, N. Tajvidi (2006): Multivariate Generalized Pareto Distributions, *Bernoulli*, Vol. 12, No. 5, 917-930.
- [11] A. V. Asimit, E. Furman, R. Vernic (2009), *Insurance: Mathematics and Economics* 46, 308-316.