BIZTOSÍTÁS ÉS KOCKÁZAT

EGYÉNI MODELLEK A TARTALÉKOLÁSBAN

KNÓDEL MÁTÉ SZAKDOLGOZATA 2016-BAN ELNYERTE A MAGYAR AKTUÁRIUS TÁRSASÁG DÍJÁT, A BIZTOSÍTÁSI MATEMATIKA IFJÚ MESTERE CÍMET.

Knódel Máté János, (NN Biztosító Zrt., Junior Aktuárius) knodelmj@gmail.com,

ÖSSZEFOGLALÓ

A nem-életbiztosítók számára kardinális kérdés, hogy a jövőben várható kárkifizetéseiket minél pontosabban előre tudják jelezni, és a fennálló (már bekövetkezett károk miatti) kötelezettségeikre megfelelő tartalékot képezzenek. Ezek számítása napjainkban kárkifutási háromszögeken alapuló módszerek segítségével történik, de felmerülhet a kérdés, hogy kaphatunk-e pontosabb és stabilabb becslést, ha a károkat nem aggregált, hanem egyéni szinten vizsgáljuk?

Munkám célja az volt, hogy a hagyományos ösvényről letérve egy egyéni, sztochasztikus kártartalékolási módszeren alapuló modellt mutassak be, a választott többdimenziós Pareto-eloszlás segítségével. Meghatároztam a modell szerinti kifizetések várható értékét és szórását. A valós adatokon történő alkalmazás céljából becslési eljárásokat is kidolgoztam.

Jelen publikáció a Biztosítási és pénzügyi matematika mesterszak aktuárius specializációján a 2015-2016-os tanévében megvédett szakdolgozatom alapján készült, mely letölthető a *https://www.cs.elte.hu/blobs/diplomamunkak/msc_actfinmat/2016/knodel_mate_janos.pdf* címről.

SUMMARY

For non-life insurers it is a cardinal point to predict the outstanding claim amounts, determine the reserve requirements and form solvency capital risk. Nowadays reserves are calculated with methods based on run-off triangles, but one could challenge this framework: could we have more appropriate and accurate predictions by using individual methods?

The aim of my work was, by leaving the classical frameworks, to introduce and develop a stochastic, individual claim reserving method, using the multivariate Pareto-distribution.

In my thesis I calculated the expected value and variance of outstanding claim amounts. In order to apply the methodology for real insurance data, I proposed novel parameter estimation techniques.

This article is based on my thesis written at Actuarial Specialization in MSc in Insurance and Financial Mathematics in the academic year 2015-2016), which can be found on the following link: *https://www.cs.elte.hu/blobs/diplomamunkak/msc_actfinmat/2016/knodel_mate_janos.pdf*

Kulcsszavak: sztochasztikus kártartalékolás, egyéni károk, többváltozós Pareto-eloszlás, lánc-létra módszer

Keywords: stochastic loss reserving, individual claims, multivariate Pareto-distribution, chain-ladder method

JEL: G22

DOI: 10.18530/BK.2016.3.40
http://dx.doi.org/1018530/BK.2016.3.40

Bevezetés, motiváció

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

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.

Az irodalomban a legtöbb modell aggregált kárkifizetési adatokkal dolgozik, azaz egy szerződéscsoport adatait vizsgálja. A két legismertebb és legelterjedtebb módszer, amelyben az aggregált adatok alapján, kárkifutási háromszögek segítségével számítható ki a tartalékszükséglet, a lánc-létra (chain-ladder, CL) módszer, valamint a Bornhuetter-Ferguson módszer. Az eljárások előnye, hogy viszonylag kevés feltételezés alapján a gyakorlatban is elfogadható becsléseket adnak.

> **A Pareto-típusú eloszlások a nagy károkra gyakran jó illeszkedést mutatnak.**

Munkám és kutatásom egyik célja az volt, hogy a klasszikus keretrendszerektől eltérve egy egyéni, sztochasztikus kártartalékolási módszeren alapuló modellt mutasson be. Az első, egyéni adatokon alapuló tartalékolási módszerek Arjas (1989) és Norberg (1993) nevéhez fűződnek, de napjainkban újabb vizsgálatok tárgyait képezik az egyéni módszerek. Az általam vizsgált keretrendszer alapjául egy viszonylag friss cikk szolgált (M. Pigeon et al (2013)), ahol a szerzők egy többdimenziós, ferde normális eloszlást használnak az elemzés során. Így felmerült bennem a kérdés, hogy más típusú eloszláscsalád esetén milyen eredményeket kaphatunk?

Választásom a Pareto-típusú eloszlásokra esett, mivel az egydimenziós esetben a nagy károkra gyakran jó illeszkedést mutatnak. Vizsgálataimhoz a Mardia (1962) által bevezetett, többváltozós, 1-es típusú eloszlást használtam, amely a jól ismert európai típusú Pareto-eloszlás többdimenziós általánosítása.

A cikk első részében bemutatom a vizsgált modellt, a használt jelölésekkel és változóival. Röviden leírom a likelihood függvényt, majd bemutatom az önálló számítási eredményként kapott, a függőkárok becslésére használt analitikus eredményeket. Ezt követően megismerhetjük a vizsgált adathalmazt, illetve a paraméterek becslésére kapott eredményeket, majd meghatározom a tartalékszükségletet. Végezetül a kapott eredményeket összevetem a lánc-létra módszerrel számított értékekkel, illetve a tényadatokkal.

A vizsgált modell

Először az általam vizsgált, diszkrét időkeretű, egyéni káradatokon alapuló sztochasztikus kártartalékolási modell alapjait fektetem le: bemutatom a vizsgált függőkárok típusait, majd leírom az ezek jellemzésére szolgáló mennyiségeket.

A függőkárok típusai

A modell diszkrét időben, azaz a folytonos időt diszkrét periódusokra osztva vizsgálja a károk bekövetkezését, illetve fejlődését. Egy periódusnak tekinthetünk egy negyedévet is, azonban a modellemben egyéves intervallumot tekintettem alap időegységként.

A függőkárokat három csoportba soroltam, amelyekre a későbbiekben 3 mozaikszóval hivatkozom. A kárbekövetkezés és a kár bejelentése közti időszakban a biztosító még nem ismeri sem a kár tényét, sem a kifizetés nagyságát, azonban már kötelezettsége keletkezik (amennyiben nem érvényesül valamilyen kizáró tényező). Ebben az időszakban a kárt **IBNR** (incurred, but not yet reported), azaz bekövetkezett, de be nem jelentett kárként kategorizáljuk. A bejelentéstől az első kifizetésig eltelt időben a kár ténye, valamint annak elfogadása már ismert és eldöntött, azonban a kárkifizetés(ek) nagysága még nem. Ebben az időszakban a kárt **RBNP** (reported, but not paid), azaz bejelentett, de még kifizetés nélküli kárnak nevezzük. Végezetül az első kifizetést a kár lezárásáig követhetik továbbiak is, ebben az időszakban **RBNS** (reported, but not settled), azaz bejelentett, de le nem zárt kárként kategorizáljuk.

A bemutatott struktúra hasonlít a magyar gyakorlatban használtakhoz, azonban a tételes függőkárokat részletesebben, RBNP és RBNS kategóriák megbontásában vizsgálja. Ezáltal lehetőség nyílhat mélyebb elemzésre, amely szükség esetén szűkíthető is.

Időváltozók

A modell felépítését az egyéni károk szintjén kezdem. Ezek címkézéséhez, megkülönböztetéséhez a következő jelölést alkalmazom: legyen az *i.* periódus *k.* kára *(ik)*, továbbá $i=1,\ldots,I$, ahol I a vizsgált periódusok száma, valamint $k=1,\ldots,K_i$, ahol K_i az *i.* periódusban bekövetkezett károk száma.

Egy adott kár leírásához szükség van annak legfontosabb adataira: bekövetkezési idő, bejelentési késlekedés, a kárkifizetések dátuma és nagysága, illetve a lezárási dátum. Ezek számokba öntéséhez az alábbi diszkrét változókat vezetem be:

- 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, amelyekben 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** $j=0,\ldots,U_{ik}$-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.

Fontos kiemelni, hogy a fent definiált diszkrét változók paraméteres eloszlásból származnak, azaz az ismeretlen paraméterek mind az adatokból becsülhetők.

A modell és a diszkrét változók szemléltetéséhez következzen az alábbi példa. Tekintsünk egy kárt, amely 2013. november 13-án következett be, és 2013. december 19-én jelentették be a biztosítótársaságnak. Így a bejelentés éve (első periódus) 2013, a bejelentési késlekedés $t_{ik}=0$ (mivel a bekövetkezés és a bejelentés azonos évben volt). Az első kárkifizetés 2014. március 24-én történt meg, így az első fizetési késés $q_{ik}=1$. Ezt követően még 2 kárkifizetés volt, melynek dátumai: 2014. november 17. és 2016. január 11. A kárt 2016. február 15-én zárták le. Összegezve az adott években a károkat, az első kifizetés után még 1 évben (2016-ban) fizettünk ki kárt, azaz $u_{ik}=1$. Az aggregált kifizetések között 1 periódus telt el, azaz $n_{ik1}=1$, valamint az utolsó kifizetés a lezárással megegyező periódusban történt, így $n_{ik2}=0$.

Kárfejlődési folyamat

Az előző részben bemutattam az adott kárt jellemző (diszkrét) változókat, de fontos, hogy annak időbeli fejlődését is le tudjuk írni. Más szavakkal, a kárkifizetések nagyságát és az ebből származó fejlődési dinamikát is meg kell ragadnunk a modellben. A végső cél pedig az, hogy minden egyes kárra meg tudjuk becsülni a dinamika alapján az összkárkifizetést.

Jelölje Y_{ikj} (>0) a *j.* részkifizetést az *(ik)* kárra. Egy adott kárra a kumulált kárkifizetést ezen részkifizetések összegeként kapjuk meg. Legyen az *(ik)*-dik kárt a *j.* és a *j+1.* kárkifizetési periódusok közt jellemző $\lambda_j^{(ik)}$ *növekedési faktor*:

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

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

$$\boldsymbol{\Lambda}_{u_{ik}\,1}^{(ik)} = \begin{pmatrix} Y_{ik1} & \lambda_1^{(ik)} & \dots & \lambda_{u_{ik}}^{(ik)} \end{pmatrix}^T$$

Ha az első kifizetést nem követte további, azaz $u_{ik}=0$, akkor a kár fejlődési mintája az egyelemű 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 részkifizetés 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 különbség jobb szemléltetéséhez kanyarodjunk vissza az előbbi példához: itt a kárfejlődési vektor kételemű: $\boldsymbol{\Lambda}_{u_{ik}+1}^{(ik)} = \begin{pmatrix} Y_{ik1} & \lambda_1^{(ik)} \end{pmatrix}^T$

Az Y_{ik1} jelöli az első kifizetést, míg $\lambda_1^{(ik)}$ azt ragadja meg, hogy a két aggregált kárkifizetéses periódus (2014 és 2016) között hányszorosára nőtt a kumulált kárkifizetés. Így tehát elkerüljük azt a problémát, hogy a 2014 és 2015-ös évek közti kumulált növekedésre egy „üres" 1-es érték kerüljön be, feleslegesen növelve a vektor dimenzióját.

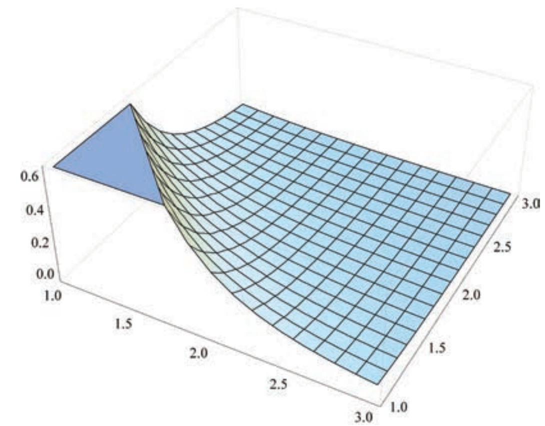
A másik lényeges különbség, hogy a sztochasztikus lánc-létra modellnél a növekedési faktorok függetlenek a múlttól, sőt, a kezdeti kárkifizetés is független a növekedési faktoroktól. Ez a feltevés azonban a valóságban sérülhet, így a vizsgálat során érdemes többdimenziós megközelítést alkalmazni a $\boldsymbol{\Lambda}_{u_{ik}+1}^{(ik)}$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

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

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

Az előző definícióban bevezetett eloszlást mutatja be az 1. ábra $\boldsymbol{a}$ = *(1,1)* elhelyezkedés és *p=2* lecsengés paraméterekkel. A vizsgált többdimenziós Pareto-eloszlás első látásra elég absztraktnak tűnik, azonban megemlíteném, hogy ennek egydimenziós eseteként megkapható a jól ismert európai-típusú Pareto-eloszlás.

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



Forrás: Saját ábra

A fent bemutatott $\boldsymbol{\Lambda}_{u_{ik}+1}^{(ik)}$ kárfejlődési vektor együttes eloszlását az előző definícióban bevezetett többdimenziós Pareto-eloszlásúnak feltétezem a modellben. Ezen eloszlás részletes tulajdonságait (peremeloszlások, feltételes eloszlás, paraméterek maximum likelihood becslése) a szakdolgozatomban tárgyalom.

Kárbejelentési és kárbekövetkezési intenzitás

Az előbbiekben bevezetett változók segítségével egy kár jellemzése már lehetségessé vált. Viszont szükséges az is, hogy becslést tudjunk adni az adott periódusban bekövetkezett károk számára, ezáltal majd a portfólió teljes kárkifizetésére.

A károk számát az adott periódusban kockázatban álló (in-force) szerződések számával becsülöm: tegyük fel, hogy az *i*. periódusban bekövetkezett károk száma, K_i Poisson-eloszlású $\Theta\, w(i)$ paraméterekkel. Itt Θ jelöli a Poisson-folyamat intenzitását, $w(i)$ 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 $\Theta\, w(i)\ (1-F_1\ (t_{ik}^{*}-1;\nu))$ paraméterrel, ahol $F_1\ (*;\nu)$ jelöli a bejelentési késés eloszlását. Így erre az időszakra az IBNR károk várható darabszáma $\Theta\, w(i)\ (1-F_1\ (t_i^{*}-1;\nu))$.

A másik két kártípusra (RBNP és RBNS) már nem szükséges hasonló becslés alkalmazása, hiszen ezek már bejelentettek, így darabszámuk is ismert.

A likelihood függvény, paraméterbecslés

A modell valós adatokon történő alkalmazásához szükséges, hogy a bemutatott diszkrét időváltozók (pl. bejelentési késlekedés), illetve a kárfejlődési vektor eloszlásának ismeretlen paramétereit meghatározzam. Ezt maximum likelihood módszerrel végzem, azaz felírom a mintából a likelihood függvényt, majd ennek maximumhelyeit keresve megkapom a paraméterek becslését is. A likelihood függvény a vizsgált kártípustól függően három részre osztható: a lezárt károk, az RBNP és az RBNS károk likelihood tagja, a maximalizálandó függvény pedig ezen tagok szorzataként adódik. Mivel a likelihood függvényt leíró képlet elég terjedelmes, így a cikkben nem térek ki a konkrét formulára.

A függvény maximumhelyének megtalálásához tényezőnként optimalizálom azt, azaz első lépésben meghatározom a kárfejlődési vektorra illesztett többdimenziós Pareto-eloszlás ismeretlen elhelyezkedés, illetve lecsengés paramétereit. Második lépésben pedig megkeresem a diszkrét eloszlások ismeretlen paramétereinek legjobb becsléseit is.

A két lépés közül a második végezhető el egyszerűbben, hiszen egy adott diszkrét eloszlás esetén (pl. geometriai eloszlás) ismertek a maximum likelihood becslést megadó formulák. Az első lépéshez azonban szükséges volt, hogy eljárást adjak a többdimenziós Pareto-eloszlás paramétereinek becslésére. Ennek pontos leírása a szakdolgozatomban megtalálható, röviden két pontból áll:

1. *Paraméterek becslése:* az a_i elhelyezkedés paraméterek a koordináták minimumaiként, a p lecsengés paraméter pedig egy egyenlet egyértelmű megoldásaként kapható meg.
2. *Optimalitás-vizsgálat:* megvizsgálom, hogy az előző lépésben kapott becslések optimálisak-e (ez egy egyenletrendszerbe helyettesítés segítségével kapható meg). Ha nem optimálisak, akkor valamelyik paraméter mentén javítás található.

A függőkárok tartalékának becslése

Miután a likelihood függvény maximalizálásából megkaptuk a paraméterek becslését, ezek segítségével már meg tudjuk becsülni a várható kifizetést az IBNR, RBNP, illetve RBNS károkra. Önálló számítási eredményként meghatároztam minden kártípusra (IBNR, RBNP és RBNS), egy adott kárra a teljes kárkifizetés várható értékét. Kiemelném, hogy ezek mind leírhatók zárt formulákkal, így a likelihood módszerből kapott paraméterek becsült értékeivel kiszámíthatóak a várható értékek.

Az egyszerűsítés végett az *(ik)* kárindexet most elhagyom, azaz a bemutatott tételek egy adott kárt jellemeznek. Az alábbiakban pedig következzen a függőkárokra vonatkozó két analitikus állítás:

I. tétel: [egy IBNR vagy RBNP kár várható értéke] Jelölje C egy IBNR (vagy RBNP) kár összértékét, azaz

$$C = Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_U$$

Tegyük fel, hogy rögzített U esetén a $\boldsymbol{\Lambda}_{u+1}$ kárfejlődési vektor a fenti definícióban megismert többdimenziós Pareto-eloszlású $\boldsymbol{a} = (a_1,\ldots,a_{U+1})$ 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:

$$\mathbb{E}_U(C) = \mathbb{E}_U\left\{\binom{U}{i}\frac{a_1 \ldots a_{U+1}\, p(p+1)\ldots(p+U)}{(p+U)(p+U-1)\ldots\ (p+1-i)(p-1-i)}\right\}$$

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

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 kifizetésekét, azaz $\boldsymbol{\Lambda}_{\boldsymbol{A}}=[Y_1\ \lambda_1 \ldots \lambda_{m-1}]^T$ *és* $\boldsymbol{\Lambda}_{\boldsymbol{B}}=[\lambda_m \ldots \lambda_U]^T$, *így* $\boldsymbol{\Lambda}=[\Lambda_A^{\ T}\ \Lambda_B^{\ T}]^T$.

II. tétel: [egy RBNS kár várható értéke] Jelölje C egy RBNS kár összértékét, azaz

$$C = Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_U$$

C értéke az ismert kifizetések mellett:

$$[C|\boldsymbol{\Lambda}_{\boldsymbol{A}}=\boldsymbol{l}_{\boldsymbol{A}}] = y_1 \cdot l_1 \cdot \ldots \cdot l_{m-1} \cdot \lambda_m \cdot \ldots \cdot \lambda_U$$

Ekkor, ha a $\boldsymbol{\Lambda}_{U+1}$ kárfejlődési vektor a fenti definícióban megismert többdimenziós Pareto-eloszlású $\boldsymbol{a} = (a_1,\ldots,a_{U+1})$ elhelyezkedés és p lecsengés paraméterekkel, akkor C feltételes várható értékét (a $\boldsymbol{\Lambda}_A$-ra nézve) a következő kifejezés adja:

$$\mathbb{E}_U(C|\boldsymbol{\Lambda}_A = \boldsymbol{l}_A) = y_1 \cdot l_1 \cdot \ldots \cdot l_{m-1} \cdot$$

$$\cdot\, \mathbb{E}_U \sum_{i=0}^{U-m} G_i \frac{a_{m+1;m} \ldots a_{U+1;m} (p+m) \ldots (p+U)}{(p+U) \ldots (p+m-i)} \cdot \left(\frac{a_{m+1}}{a_{m+1;m}} + \frac{1}{p+m-1-i} \right)$$

ahol a G_i együtthatók és az $a_{m+1;m}$ paraméterek egy rekurziós eljárással kaphatóak meg (ennek pontos leírása a szakdolgozatomban szerepel).

Megjegyzés: Az RBNS károk várható összkifizetését leíró formula bonyolult, mivel némely elemei csak rekurzióval kaphatóak meg. Ez jelentős mértékben megnehezít(het)i a gyakorlatban ezen kártípusra a teljes kárkifizetés várható értékének számítását.

A függőkárok összkifizetését leíró állítások bizonyításai technikailag elég bonyolultak, a zárt formulák megadása „körmönfont okoskodást kíván", idézve bírálóm szavait. Ezen okból kifolyólag a cikkben nem térek ki erre, de az érdeklődő Olvasó megtalálhatja ezeket a szakdolgozatomban.

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

III. tétel: [becslés az IBNR, RBNP és RBNS károk összkifizetésére] Jelölje $\mathcal{I}$ a károkra az adathalmazból megfigyelt információt. Így az IBNR, RBNP és RBNS károk várható értékét a következőképpen kapjuk meg:

1. Az *IBNR* károk várható értéke: $\mathbb{E}[IBNR\,|\,\mathcal{I}]=\mathbb{E}(K_{IBNR})\cdot\mathbb{E}_U(C_{IBNR})$, ahol az IBNR károk várható darabszáma ($\mathbb{E}(K_{IBNR})$) a Poisson-eloszlásból kapható meg, valamint a teljes kárkifizetés várható értéke egy IBNR kárra (C_{IBNR}) az I. tétel alapján számolható.
2. Az *RBNP* károk várható értéke: $\mathbb{E}[RBNP\,|\,\mathcal{I}]=k_{RBNP}\cdot\mathbb{E}_U(C_{RBNP})$, ahol az RBNP károk darabszáma (k_{RBNP}) adott, valamint a teljes várható kárkifizetés egy RBNP kárra (C_{RBNP}) az I. tétel alapján számolható.
3. Az *RBNS* károk várható értéke: $\mathbb{E}[RBNS\,|\,\mathcal{I}]=\sum_{(ik)RBNS}\mathbb{E}_U(C_{ik}\,|\,\Lambda_A^{ik}=l_A^{ik})$, ahol a szumma végigfut az RBNS károkon, és egy adott RBNS kár teljes várható kifizetése (a már ismert $\boldsymbol{\Lambda}_A^{ik}=\boldsymbol{l}_A^{ik}$ kifizetések alapján) a II. tétel alapján számolható.

Az összegző tételek tulajdonképpen matematikai formulákba öntik az intuitív módszereket: IBNR (RBNP) károkra számítsuk ki egy kár várható értékét, majd szorozzuk meg ezt a várható (ismert) kárdarabszámmal. RBNS károk esetében pedig tételesen, minden egyes kárra számítsuk ki a hozzá tartozó módosított paramétereket a rekurzió segítségével, majd a már ismert kárkifizetések függvényében megkapjuk a várható, még fennálló kötelezettségek nagyságát.

A modell alkalmazása valós adatokon

Az előző részben bemutattam a modellt és az ismeretlen paraméterek becslését a maximum likelihood módszer segítségével. Ezek ismeretében jogosan merül fel a kérdés: a leírt elméleti keretrendszer és a módszerek hogyan alkalmazhatóak az aktuáriusi gyakorlatban?

Az adatok jellemzése

Az elemzéshez egy biztosító nem-életbiztosítási adatait használtam fel. A számolás egyszerűsítése miatt minden kifizetést elosztottam 100 000-rel, ami csak a várható összkifizetés, illetve a tartalékok nagyságára hat (mindkét érték 100 000 Ft-szorosa az itt leírt eredményeknek). Az elérhető adathalmaz 11 év kárkifizetéseit 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 tekintettem, 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 36 599 kárt tartalmaz, amelyek közül 29 688 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: 29 749 olyan kár van, ahol csak egy kifizetés történt, 210 olyan, ahol pontosan kettő, míg egy 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 a Függelékben az 1. táblázat. Szükségesnek tartom kiemelni, hogy a harmadik kárkifizetésre csak egy adatom van, így a paraméterbecslések 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.

Paraméterbecslés

A vizsgálat során a következő cél, hogy a káradatokat jellemző változók eloszlásainak ismeretlen paramétereit meghatározzam. Ezt a már bemutatott maximum likelihood módszerrel végzem el, az optimalizálás során az R programcsomagot használom. A vizsgálathoz használt program forráskódja, illetve a nyers adatok a *http://bit.ly/23PJQqZ* linken érhetőek el.

A becslés első lépéseként az elemzés során alkalmazott, többdimenziós Pareto-eloszlás ismeretlen $\boldsymbol{a}$ elhelyezkedésvektor, illetve p lecsengés paramétereinek becslését határoztam meg. Ezt a 2. fejezetben bemutatott kétlépéses eljárás segítségével kaptam meg. A becslésekre adódott értékek a Függelék 2. táblázatában láthatóak.

Kiemelném, hogy a $\breve{p}$ -ra kapott eredmény a modellbeli analitikus eredmények szempontjából jónak mondható, hiszen maximálisan 3 kifizetést feltételezve, teljesül a $p>U+1$ feltétel ($U+1$ a kárfejlődési vektor dimenziója), így az adatokra illesztett többváltozós Pareto-eloszlás véges várható értékű (illetve véges szórású is, ezt a szakdolgozatomban szintén beláttam). Érdemes megemlíteni azt is, hogy a $\breve{p}$ viszonylag magas értéke gyors lecsengésű eloszlásra utal. Ezt a megfigyelést alátámasztja a Függelékben az 1. ábra is, amely az első kifizetések hisztogramját mutatja.

Az időváltozók paramétereinek becslése

Második lépésként a bemutatott változók (időváltozók, első kifizetések utáni további kifizetések darabszáma) ismeretlen paramétereit becsülöm meg a mintából. A likelihood függvényt szintén tagonként maximalizálom: minden változóra diszkrét eloszlásokat (geometriai, Poisson, binomiális, illetve negatív binomiális eloszlás) használok, majd ezek paramétereinek változtatásával optimalizálom az adott likelihood-tagot. Az adatokra legjobban illeszkedő eloszlást az Akaike, illetve a Bayesi információs kritériumok alapján határoztam meg. Az eredményül kapott maximum likelihood becslések szintén a 2. táblázatban találhatóak, a legjobban illeszkedő eloszlással együtt.

A kapott eredmények tekintetében levonható néhány következtetés:

- A kár bekövetkezése után rövid a bejelentési késlekedés, azaz várhatóan a károk nagy részét 0 vagy 1 periódus késlekedéssel jelentik be.
- Várhatóan a bejelentést követő első vagy második periódusban megtörténik az első kifizetés.
- Az első kifizetés utáni tovább kifizetésre illeszkedő geometriai eloszlás paraméterei az mutatják, hogy nagy valószínűséggel az első kifizetést már nem követik továbbiak. Ez összhangban van azzal, hogy a mintában 29 749 kárra volt csak egy kifizetés, míg 211 darabra legalább kettő.
- Ha az első kifizetést követi további, akkor várhatóan az azt követő 1-2 periódusban megtörténik a következő is.

A károkat jellemző változók ismeretlen paramétereinek becslése után szükséges, hogy megbecsüljük az IBNR károk várható darabszámát is. A vizsgálatok során sajnos az állomány mérete és a kockázatban álló szerződések száma nem állt a rendelkezésemre, csak a káradatok. Ebből kifolyólag az IBNR károk becslését nem tudom elvégezni a Poisson-eloszlás alapján, így a lánc-létra módszert használom a még be nem jelentett károk darabszámának becslésére, amelyre 6 315,201 adódik.

A függőkárok tartalékának becslése, összehasonlítás

Miután a maximum likelihood módszerrel megbecsültem az ismeretlen paramétereket, minden adott ahhoz, hogy kiszámítsam a függőkárok tartalékszükségletét is. Az alábbiakban az analitikus eredmények felhasználásával megbecsülöm a függőkárok tartalékszükségletét, és a kapott eredményeket összevetem a lánc-létra módszer által becsült tartalékokkal, valamint a ténykárkifizetéssel. Ezek alapján vizsgálom a modellem jóságát és illeszkedését az adatokra, és ahol lehet, az eltérések okát igyekszem feltárni.

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 az I. 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 a lezárt károk átlagának (1,235), de az eltérés csak 5 százalék körüli.

Az IBNR károk várható darabszáma az előző fejezet alapján 6 315201, míg az RBNP károk ismert darabszáma 6 639. Ezek alapján az IBNR károk tartalékszükséglete 7 458,253, míg az RBNP károké 7 843,197.

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 értéket a II. 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 össztartalék-szükségletére 333,764 adódott.

Így a modellben a függőkárok teljes tartalékszükséglete 15 635,21.

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 jelentették be őket a biztosítótársaságnak. Így a kifizetések egy teljes kárkifutási négyzetet alkottak, amelyet a Függelék 3. táblázata foglal össze.

Az ebben látható adatok kumulált kárkifizetéseket tartalmaznak, a szürke hátterű értékek mutatják a még nem ismert kifizetéseket.

Ezekből a vizsgálat során csak azokat vettem figyelembe, amelyeknek a kárbejelentési periódusa nem nagyobb 6-nál. Ezáltal egy felső kárkifutási 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 Függelék 4. táblázata. A szürke hátterű értékek adják meg a becsült kumulált kárkifizetéseket a bejelentési késés függvényében.

A táblázatokból kiszámolható, hogy a tény tartalékszükséglet 20 475,165 lenne, míg a lánc-létra módszer által becsült érték, 18 844,689 némileg alatta marad ennek (az eltérés kb. 8% a tényadatokhoz viszonyítva).

Összehasonlítás, az eltérések magyarázata

Az általam vizsgált modellben az össztartalék-szükséglet (az IBNR, RBNP és RBNS károk tartalékának összegzése után) 15 635,21 adódott, míg a lánc-létra módszer által becsült érték 18 844,689, a tényadatokból kiolvasott tartalék pedig 20 475,165 lenne. Látható, hogy a modell által becsült érték alatta marad mind a lánc-létra becslésnek, mind a tényeredményeknek. Az eltérés a tényadatoktól 24 százalék körüli, míg a lánc-létra módszer eltérésre 8 százalék körüli. Ezért szükségesnek látom, hogy megvizsgáljam az eltérések okát, illetve a modellem jóságát.

A legnagyobb eltérést az okozza, hogy az RBNP státuszú károk (azaz a 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 RBNP státuszú károk átlagkifizetése 1,5-szer nagyobb a lezárt károk átlagkifizetésénél.

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 összes kárkifizetést, 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.

Összegzés, további vizsgálati lehetőségek

A szakdolgozatomban bemutattam egy lehetséges, diszkrét idejű tartalékolási modellt. Leírtam a vizsgált modellt, annak változóit, majd önálló eredményként analitikus formulákat adtam a várható összkárkifizetés becslésére, ha a kárfejlődési vektor eloszlását többdimenziós Pareto-eloszlásnak feltételezzük. Végül egy biztosító nem-életbiztosítási adataira illesztettem a modellem, valamint kiszámítottam a tartalékszükségletet. Az eredményeket összevetettem 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 százalék körüli, ennél jobb eredményt ad a lánc-létra módszer, amelynek hibája csak 8 százalék (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, egyben legfőbb hátrányaként annak részletességét emelném ki. Azáltal, hogy egyéni szinten vizsgáljuk a bekövetkezett károkat és azok fejlődését, 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. Másrészt viszont meg kell említeni, hogy a részletesség, a viszonylag sok paraméter becslése rengeteg apró hibát von maga után, amely pontatlanabbá teheti az eredményeket.

Modellemben a paraméterek becsléséből is származik hiba, hiszen ezeket az eddig ismert adatok, főként a lezárt károk alapján végeztem el. Érdekes képet kaphatnék akkor, ha a becsült paraméterekből (kárkifizetési vektor, illetve időváltozók) károkat szimulálnék, majd összevetném, hogy a kapott véletlen minta eloszlása mennyire hasonlít a tényadatok eloszlására. A szimulációs módszer arra is lehetőséget ad, hogy előrejelezzük az összkifizetés eloszlását, előrejelzési intervallumokat adjunk a késői kárdarabszámra és kifizetésekre. Mindennek előfeltétele az, hogy tudjunk többdimenziós Pareto-eloszlású véletlen vektorokat generálni, ami további külön fejlesztést igényel.

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 figyelembevétele 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 szélesebb körben elterjedhet a jövőben, 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 pontosabb, jobban illeszkedő előrejelzést alkothatunk meg a jövőben.

IRODALOMJEGYZÉK

Knódel Máté János (2016): Egyéni modellek a tartalékolásban (szakdolgozat), Eötvös Loránd Tudományegyetem, Természettudományi Kar

E. Arjas (1989): The claim reserving problem in non-life insurance: some structural ideas, ASTIN Bulletin, Volume 19, Issue 2. pp. 139–152 http://dx.doi.org/10.2143/ast.19.2.2014905

M. Pigeon et al (2013): Individual Loss Reserving with the Multivariate Skew Normal Framework, ASTIN Bulletin, Volume 43, Issue 03. pp. 399–428 http://dx.doi.org/10.1017/asb.2013.20

K. V. Mardia (1962): Multivariate Pareto Distributions, The Annals of Mathematical Statistics, Volume 33, Issue 3. pp. 1008–1015 http://dx.doi.org/10.1214/aoms/1177704468

R. Norberg (1993): Prediction of Outstanding Liabilities in Non-Life Insurance, ASTIN Bulletin, Volume 23, Issue 1. pp. 95–115 http://dx.doi.org/10.2143/ast.23.1.2005103

FÜGGELÉK

1. táblázat: Leíró statisztika a kárfejlődési vektorról

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

Forrás: a felhasznált adatok alapján saját számítás

2. táblázat: Az eloszlások ismeretlen paramétereinek maximum likelihood becslése

Változó neve	ML becslés	Eloszlás
Pareto-eloszlás első elhelyezkedés paramétere	$\widehat{a_1} = 1{,}000592$	-
Pareto-eloszlás második elhelyezkedés paramétere	$\widehat{a_2} = 1{,}120636$	-
Pareto-eloszlás harmadik elhelyezkedés paramétere	$\widehat{a_3} = 1{,}497919$	-
Pareto-eloszlás lecsengés paramétere	$\hat{p} = 6{,}738437$	-
bejelentési késlekedés	$\widehat{r_T} = 4{,}4993076$ $\widehat{q_T} = 0{,}8855756$	negatív binomiális
első fizetési késlekedés	$\widehat{\lambda_Q} = 1{,}703278$	Poisson
első kifizetést követő további kifizetések darabszáma	$\widehat{p_U} = 0{,}9838264$	geometriai
második fizetési késlekedés	$\widehat{r_{N_1}} = 17{,}870478$ $\widehat{q_{N_1}} = 0{,}966587$	negatív binomiális
harmadik fizetési késlekedés	$N_2 = 2$	-

Forrás: a felhasznált adatok alapján saját számítás

3. táblázat: A kumulált ténykárkifizetések a bejelentési év és a bejelentési késés szerint

Claim\Payment lag	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	10 332,82
5	546,209	5345,121	7710,437	9286,894	11 461,25	11 954,06
6	832,193	5534,142	8290,022	9328,821	10 064,73	10 485,07

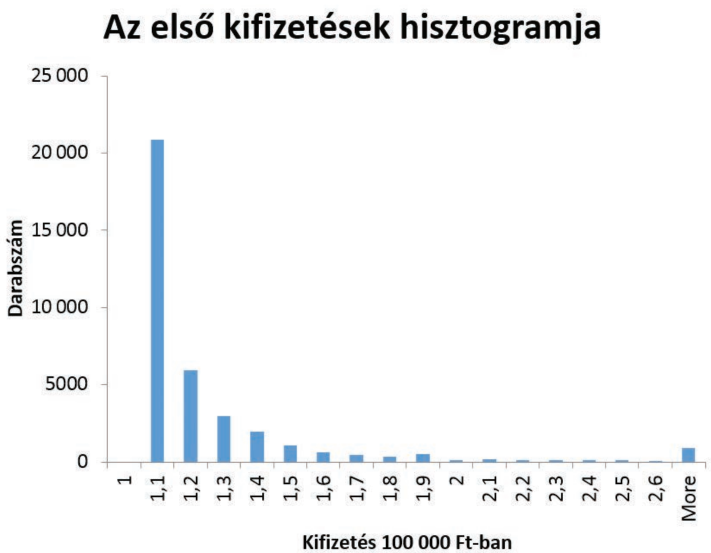
Forrás: a felhasznált adatok alapján saját számítás

4. táblázat: A lánc-létra módszerrel becsült, kumulált kifizetések a bejelentési év és a bejelentési késés szerint

Claim\Payment lag	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	10 377,88
5	546,209	5345,121	7642,577	8687,483	9426,501	10 160,2
6	832,193	5516,481	7887,591	8965,996	9728,706	10 485,92

Forrás: a felhasznált adatok alapján saját számítás

1. ábra: Az első kifizetések hisztogramja 100 000 Ft-ban



Forrás: a felhasznált adatok alapján saját számítás