

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
BUDAPESTI CORVINUS EGYETEM

Hegel Patrik

**HALANDÓSÁG BECSLÉSE AZ EGÉSZSÉGÁLLAPOT
FÜGGVÉNY SEGÍTSÉGÉVEL**

MSc Szakdolgozat

Témavezető:

Arató Miklós

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2020

NYILATKOZAT

Név: Hegel Patrik

ELTE Természettudományi Kar, szak: Biztosítási és pénzügyi matematika MSc

NEPTUN azonosító: FZDHJC

Szakdolgozat címe:

Halandóság becslése az egészségfüggvény segítségével

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

Budapest, 2020 május 29.

Hegel Patrik

a hallgató aláírása

Tartalomjegyzék

1. Bevezetés	4
2. Várható halandóság becslése	6
3. Egészségállapot függvény	10
3.1. Az egészségi állapot alakulásának modellezése	10
3.2. A haláleset bekövetkezési idejének eloszlása	11
3.3. Kiegészített modellek	14
4. A Skiadas-féle modell	16
4.1. Az egészségállapot függvény paraméterezése	16
4.2. Paraméterbecslés	20
4.3. A modell illesztése valós adatokra	21
5. Az egészségállapot alakulása	27
5.1. Előrejelzés	32
6. Összefoglalás	39

1. fejezet

Bevezetés

Az utóbbi két évszázad folyamán a Föld lakosságának halandósága fokozatosan javul, amelynek eredményeként a várható élettartam körülbelül kétszeresére nőtt az időszak alatt. Az emberek növekvő élethosszának köszönhetően bekövetkező gazdasági és szociális hatások vizsgálata több területen is kiemelkedő fontosságú. Például a nyugdíjrendszerrel, nyugdíjbiztosításokkal és életbiztosításokkal kapcsolatos számítások jelentős része halálozási arányszámokon alapul.

Ezeknek az intézményeknek elengedhetetlen a hosszú élet kockázatát megfelelően felmérni és kezelni, hiszen a lakosság egyre nagyobb hányada él kiemelkedően magas, akár 110 év feletti korig. Aktuáriusi feladat egy populáció mortalitásának előrejelzése is, amely halandósági modellek használatával történik.

A szakdolgozatban egy olyan modell kerül bemutatásra, amely képes egy lakosság halandóságának dinamikáját jól leírni és előrejelezni azt múltbeli adatok alapján.

A szakdolgozat második fejezetében egy egyszerű modellt adunk a várható halandóság becslésére. Az elmúlt néhány év átlagos halandóságát használjuk előrejelzésként a jövőbeli halandóságra és néhány példán keresztül megmutatjuk ennek a módszernek a hibáit és hiányosságait, igazolva ezzel a további részletesebb elemzés szükségességét.

A harmadik fejezetben bemutatjuk a szakdolgozatban később használt és részletezett modell elméleti alapját és feltevéseit. A cél jól megfogni az emberi élettartam során az egyén egészségi állapotának alakulását és annak segítségével sztochasztikusan modellezni a halál bekövetkeztének időpontját. Ennek érdekében definiáljuk az "egészségállapot függvényt" és becslést adunk az élettartam eloszlására. Ennek a fejezetnek az alapjául C. H. Skiadas és J. Janssen cikkei szolgáltak ([1], [2], [3], [4]) és számolásokkal, rövid bizonyítással egészítettem ki őket.

A negyedik fejezetben az előbb említett források alapján több lehetséges paraméterezését megadjuk az egészségállapot függvénynek. A Skiadas és Janssen által alkalmazott paraméterbecslési módszer mellett maximum likelihood módszer segítségével is elvégezzük a becslést. A használt adatok kizárólag különböző országok elérhető halandósági táblái voltak és nem vettünk figyelem-

be az adott országokra jellemző adatokat paraméterként (például GDP, földrajzi elhelyezkedés, egyéb gazdasági mutatók). A bemutatott példában a brit nők és férfiak adatain modelleztünk és elég pontos illeszkedést sikerült kapnunk.

A szakdolgozat ötödik fejezetében alakítunk a modellen, mégpedig úgy, hogy több év halandóságának leírására és előrejelzésre is alkalmazható legyen. Ennek érdekében kicsit módosítunk az egészségfüggvény paraméterezésén: naptári évekre és kohort évekre vonatkozó paramétereket vezetünk be. Tehát egy x éves személy egészségi állapotát Y naptári évben az Y -hoz és a születési évéhez ($Y - x$) tartozó paraméter segítségével becsüljük. Az előrejelzéshez pedig idősoros modellekkel jelezzük előre először a paraméterek alakulását az elkövetkezendő évek során, majd kalkuláljuk belőlük az élettartamok eloszlását. Az előrejelzést svéd, amerikai és brit nők adatain végezzük el, a cél ugyanis részben az angolszász országokban az utóbbi években tapasztalt halandósági romlás magyarázata.

2. fejezet

Várható halandóság becslése

Az életbiztosítások és nyugdíjbiztosítások esetén az egyik legfontosabb téma a lakosság halandóságának becslése és előrejelzése. A megfelelő halandósági ráták használata elengedhetetlen a pontos díjkalkulációhoz és a várható kifizetések meghatározásához is. Ebben a fejezetben egy egyszerű becslést végzünk az elmúlt évek halandósági adatai alapján több ország mortalitásának alakulására.

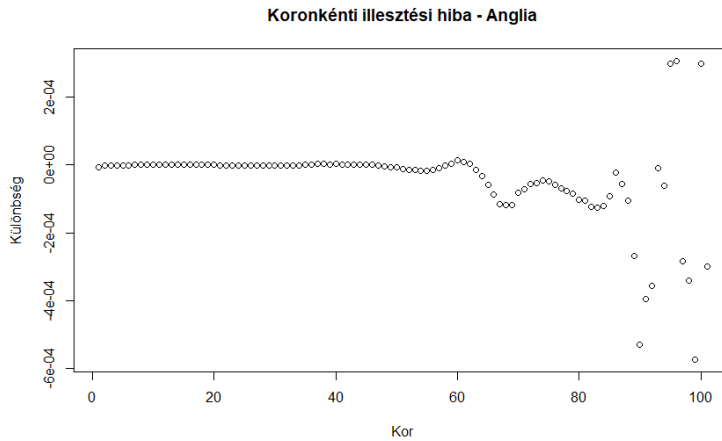
Tegyük fel, hogy rendelkezésünkre állnak adatok az elmúlt n évből az elhunytak és megfigyelték létszámára minden életkorra 1-100 éves korig. Jelölje rendre N_i^j és D_i^j a j -edik évben és i -edik életkorban a megfigyelték és a halottak számát. A tényleges halálozási rátákat egy adott életkorra és évre a kettő hányadosaként kaphatjuk meg: $r_i^j = \frac{D_i^j}{N_i^j}$.

Az adataink alapján határozzuk meg n évre az átlagos halandósági rátákat koronként és azok segítségével jelezzük előre a következő évek halandóságát. Tehát legyen $\bar{r}_i = \frac{\sum_{j=1}^n \frac{D_i^j}{N_i^j}}{n}$.

Ekkor, ha az előrejelzendő évben N_i a megfigyelték létszáma az i -edik életkorban, akkor a becsült száma az elhunytaknak: $\hat{D}_i = \bar{r}_i \times N_i$.

A becsléshez az átlagos ráta számolásához 10 év adatait tekintettük. Az adatok forrása a Human Mortality Database ^[8] volt, ahonnan a Magyarországra, Svédországra, az USA-ra, Japánra és az Egyesült Királyságra vonatkozó adatokat kértük le a 2006-2015 közti évekre. Az előrejelzéseket 2016 és 2017-es évre végeztük (illetve ahol rendelkezésre állt 2018-as adat, ott arra is).

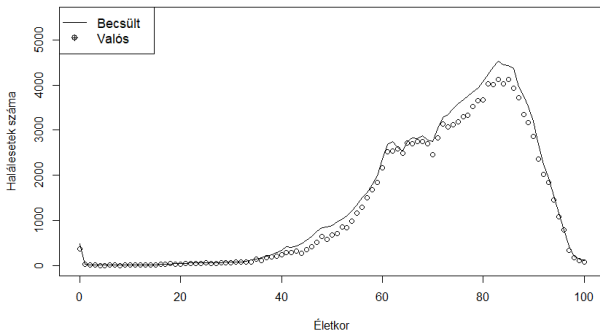
Azt tapasztaltuk, hogy az átlagolt 10 év összesített haláleseteinek a számára viszonylag pontosan illeszkedik a becsült rátával számolt, de ez kevésbé meglepő, hiszen ezek alapján állítottuk elő a becslést. A következő ábrán életkoronként látjuk a különbségeket a valós és illesztett halandósági ráták nagysága között Anglia esetén:



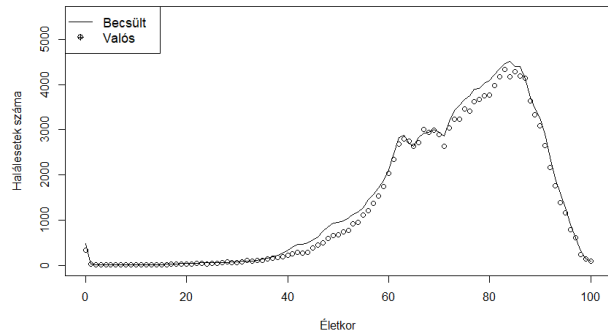
Idősebb korokban ez a becslés pontatlanabb, ennek oka lehet a kevesebb megfigyelés miatti változékonysága a halandósági rátának.

A korábban említett országokra elvégeztük a 2006-2015-ös adatok alapján az előrejelzést 2016-ra és 2017-re (kivéve az Egyesült Királyság esetén, ahol 2017-es adatok nem álltak rendelkezésre). Jellemzően az elhunytak valós számát felülbecsültük ezzel a becsléssel. Ezt a következő ábrákon szemléltetjük: az elhunytak becslött és valós számát rajzoltuk ki a különböző országok esetén (USA esetén külön női és férfi halandóságokra is ábrázoltuk, a többi esetén összesített).

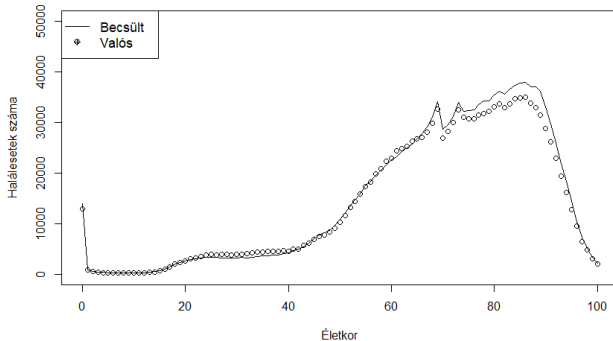
Előrejelzés 2016-ra - Magyarország



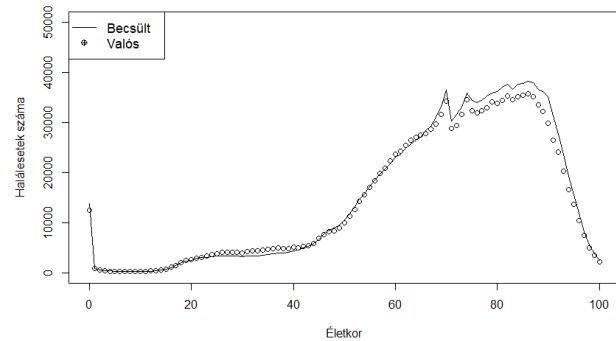
Előrejelzés 2017-re - Magyarország

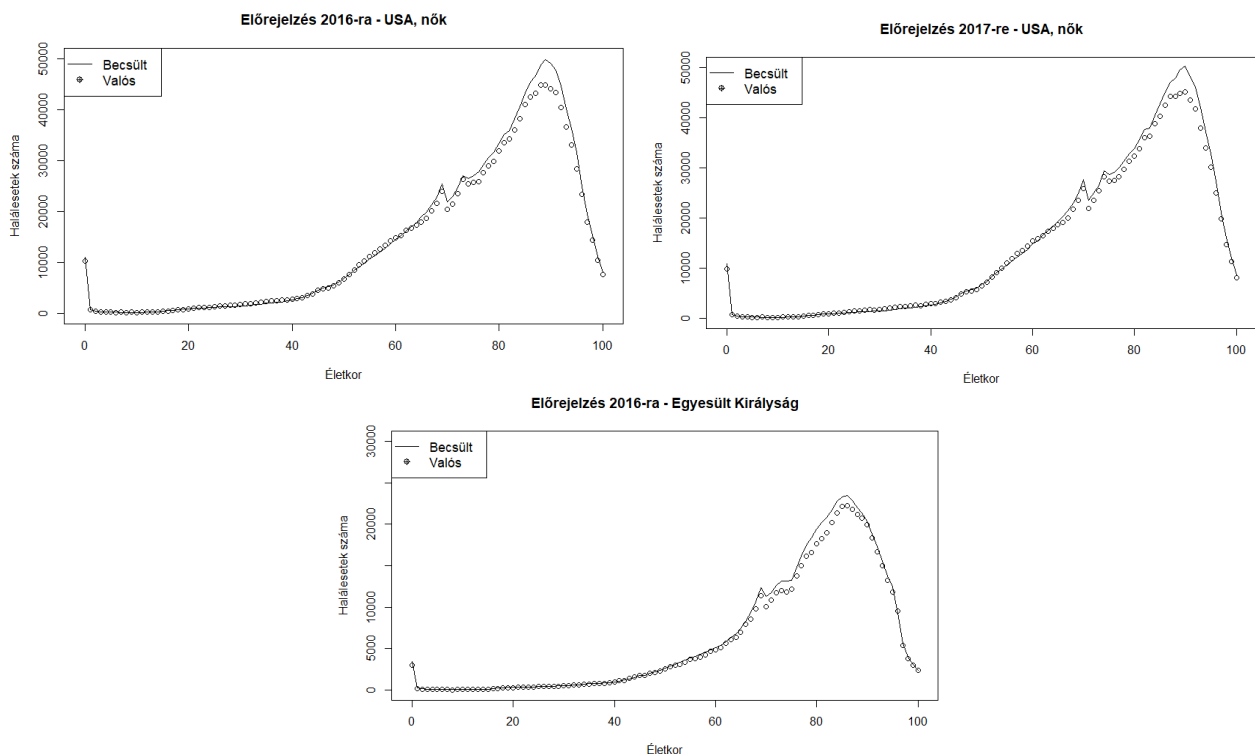


Előrejelzés 2016-ra - USA, férfiak



Előrejelzés 2017-re - USA, férfiak





A magyarázata a túlbecslésnek a halandóság javulása lehet: arra számítunk, hogy a vizsgált évek, illetve évtized alatt az egészségügy általános javulása, a lakosság tudatosabb és egészségesebb életvitele jobb mortalitást eredményez, bármely országot vizsgáljuk.

Az egyetlen kivétel, amikor nem becsültük fölé a halandóságot, az az USA esetén a középkorú lakosság körében történt. Az utóbbi időben több helyen lehetett olvasni, hogy az Egyesült Államokban és az Egyesült Királyságban is megállt vagy lelassult a halandóság javulása az utóbbi években ([10][11]). Az alábbi táblázatban 10 éves korosztályok halandósági rátáinak alakulását foglaltam össze az utóbbi években az Egyesült Királyságban:

	0-9 éves	10-19 éves	20-29 éves	30-39 éves	40-49 éves	50-59 éves	60-69 éves	70-79 éves	80-89 éves	90-99 éves
2006	0.0006494778	0.0002484089	0.0005373801	0.0008952954	0.001950217	0.004781399	0.01208825	0.03329472	0.09828351	0.2696547
2007	0.0006305793	0.0002392723	0.0005192626	0.0008989431	0.001901161	0.004715147	0.01185524	0.03236831	0.09765433	0.2754831
2008	0.0006143925	0.0002228750	0.0005276000	0.0009139731	0.001922058	0.004633496	0.01159828	0.03176996	0.09755028	0.2818079
2009	0.0005878262	0.0002182344	0.0005007006	0.0008747248	0.001883908	0.004504993	0.01104575	0.03000705	0.09274081	0.2620185
2010	0.0005618958	0.0001922388	0.0004629180	0.0008474894	0.001822644	0.004443949	0.01080810	0.02944648	0.09050199	0.2631372
2011	0.0005486189	0.0001686251	0.0004380727	0.0008154834	0.001775309	0.004267020	0.01037153	0.02829811	0.08747491	0.2555161
2012	0.0005239726	0.0001686342	0.0004084201	0.0007687661	0.001727171	0.004139145	0.01023238	0.02821058	0.08958393	0.2676912
2013	0.0004934014	0.0001665682	0.0004024349	0.0007997653	0.001750891	0.004135032	0.01016099	0.02784614	0.08906312	0.2658411
2014	0.0004912467	0.0001707950	0.0004172915	0.0007870193	0.001762425	0.004035181	0.00995088	0.02716372	0.08595690	0.2550737
2015	0.0004916843	0.0001694135	0.0004131526	0.0008065699	0.001775402	0.004077888	0.01007531	0.02762520	0.09021964	0.2720601
2016	0.0004846422	0.0001640003	0.0004426512	0.0008112127	0.001831761	0.004130856	0.01011644	0.02710522	0.08686472	0.2581992

Illetve az USA-ban a nők esetén:

	0-9 éves	10-19 éves	20-29 éves	30-39 éves	40-49 éves	50-59 éves	60-69 éves	70-79 éves	80-89 éves	90-99 éves
2006	0.0008027894	0.0002461082	0.0005303917	0.0009087018	0.002184217	0.004635181	0.010863977	0.02736489	0.08208391	0.2489202
2007	0.0008032340	0.0002416829	0.0005255973	0.0008937706	0.002123891	0.004560074	0.010590481	0.02670515	0.07993356	0.2449139
2008	0.0007739511	0.0002216682	0.0005072845	0.0008700917	0.002128545	0.004533321	0.010581193	0.02674421	0.08007122	0.2495221
2009	0.0007441257	0.0002174379	0.0005165354	0.0008896932	0.002118010	0.004557200	0.010231452	0.02575626	0.07616455	0.2330829
2010	0.0007071900	0.0002006738	0.0005008505	0.0008750462	0.002018547	0.004494286	0.010037140	0.02547099	0.07553545	0.2367001
2011	0.0006990644	0.0001971695	0.0004992312	0.0008931934	0.002036374	0.004543493	0.009930321	0.02528300	0.07545693	0.2386064
2012	0.0006903592	0.0001930656	0.0005070493	0.0008830441	0.001997764	0.004564404	0.009818157	0.02493474	0.07421556	0.2344371
2013	0.0006778582	0.0001913417	0.0005034443	0.0008988507	0.001991194	0.004603765	0.009825405	0.02487971	0.07372663	0.2332672
2014	0.0006692618	0.0001889600	0.0005124463	0.0009248298	0.002010076	0.004687043	0.009762037	0.02457583	0.07156246	0.2252572
2015	0.0006755802	0.0002064694	0.0005383779	0.0009714453	0.002013679	0.004712086	0.009786881	0.02472448	0.07277133	0.2280410
2016	0.0006807752	0.0002118258	0.0005864209	0.0010542754	0.002035452	0.004737291	0.009785232	0.02428819	0.07102379	0.2177809
2017	0.0006557384	0.0002082251	0.0005999709	0.0010734404	0.002029133	0.004700061	0.009812291	0.02430208	0.07147836	0.2176417

A táblázatok alapján megfigyelhető, hogy az aktív korú lakosság esetén a 2010-es éves elejéig a halandóság fokozatosan javult, majd utána elkezdett romlani vagy szinten maradni mindkét ország esetén.

Az előbbieken alkalmazott egyszerű becslés egyik hibája az, hogy a mortalitás időbeli alakulását gyakorlatilag nem tudja kezelni és előrejelezni. Ez jellemzően a pontos ráták felülbecslését eredményezi javuló halandóságú országok esetén, azonban előfordulhat ennek ellentéte is (például USA és Egyesült Királyság). Továbbá elképzelhető az is, hogy egy-egy évben kiugrásszerűen változnak a halandósági mutatói egy országnak. Ehhez vezető példa lehet akár egy a megszokottnál súlyosabb járvány vagy egy a nép egészségügyét és egészségi helyzetét érintő kormányzati intézkedés. Ha egy ilyen évet számításba veszünk, akkor az a figyelembe vett évek számától függően torzíthat az átlagos halandósági ráták meghatározásakor, ezáltal pontatlan becslésre vezetve.

Felmerül tehát az igény egy olyan modell alkalmazására a mortalitásnak és annak több év távlatában való alakulásának modellezésére, amely az előbbi hibákat hatékonyabban kezeli. Az ilyen célokra alkalmazott modellek közül megemlítenődnek a Lee-Carter [1992] és a Cairns-Blake-Dowd [2006] modellek. A szakdolgozat további fejezeteiben azonban egy ezektől különböző modell kerül bemutatásra és kiegészítésre, amely alkalmas a halandóság és változásának leírására.

3. fejezet

Egészségállapot függvény

3.1. Az egészségi állapot alakulásának modellezése

Egy populáció mortalitásával szorosan összefüggő paraméter az adott csoport egészségügyi állapota. Ugyanakkor az egészségi állapot mérése ritkábban alkalmazott eszköz a biztosításban, egészség-tudományokban és a demográfiában. Ennek oka lehet, hogy a mortalitás egy jól meghatározott fogalom, amely precízen számolható adott korcsoportokra az elhunytak és élők számából egy bizonyos populációban. Ezzel szemben a lakosság vagy akár az egyén egészségügyi állapotának számszerű meghatározása nem mérhető ennyire egyszerűen. Bár a mindennapi köznyelv része, a felmérésekkel és kérdőívekkel szerzett adatok bizonytalansága és inkonzisztenciája nem igazán ad lehetőséget modellezésre az egyén szintjén. Ugyanakkor ez nem is szükségszerű, hiszen az elsődleges célunk a lakosság egészére definiálni és meghatározni egy aggregált egészségügyi állapotot. A dolgozatban egy lehetséges modellt és annak továbbgondolását mutatom be a halandóság becslésére az egészségi állapot használatával.

Az egyén egészségügyi állapotának alakulására az évek alatt az jellemző, hogy időben csökken és gyakori, hirtelen változások vannak benne. Emiatt tekintsük egy időben folytonos sztochasztikus folyamatnak, amelyet jelöljön $S = (S(t), t \geq 0)$. Azt feltételezzük, hogy a halál akkor következik be, ha az egészségi állapot nagy mértékben lecsökken, ellenben ha az magasabb szinten van, akkor életben van az egyén. A halál bekövetkeztének idejét definiáljuk úgy, hogy az a pillanat, amikor az egészségállapot függvény egy bizonyos α szint alá csökken először. Tehát tulajdonképpen a sztochasztikus folyamat egy bizonyos szintelérési ideje.

Tegyük fel, hogy születéskor az egyén $S(0)$ "kezdeti értékkel" kezd és az évek során az egészségi állapotot leíró folyamat fejlődése megadható a következő sztochasztikus differenciál egyenlettel:

$$dS(t) = \mu(S, t)dt + \sigma(S, t)dW(t) \quad (3.1)$$

Tételezzük fel továbbá azokat a standard feltevéseket a kifejezésben szereplő függvényekről, amelyek biztosítják, hogy a differenciál egyenletnek egyértelmű erős megoldása létezik adott $S(0)$ mellett.

Janssen és Skiadas modelljében^[1] a következő egyszerűsítéseket választja:

$$\mu(S, t) = \mu(t) \quad , \quad \sigma(S, t) = \sigma(t),$$

ahol tehát $S(t)$ a sztochasztikus folyamat értéke a t idő függvényében, $\mu(t)$ a drift együttható, $\sigma(t)$ a szórásparaméter és $W(t)$ a standard Wiener-folyamat. Vagyis a folyamat lényegében egy Ito-folyamat, adott kezdeti értékkel:

$$S(t) = S(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s)$$

Megemlítendő a C-C-J-P modell^[1], amelyet a következő választással kapunk:

$$\mu(S, t) = \mu S \quad , \quad \sigma(S, t) = \sigma S,$$

ahol $\mu < 0$ és $\sigma > 0$ konstansok. Ez a modell esetén az S folyamat geometriai Brown-mozgás. Az [1] forrás szerint ez a modell nem illeszkedett jól a halandósági tábla adataira.

3.2. A haláleset bekövetkezési idejének eloszlása

Ebben az alfejezetben a [1],[3] és [4] források alapján levezetünk egy közelítést a haláleset idejének eloszlásának sűrűségfüggvényére. Továbbá definiálásra kerül az "egészségállapot függvény", amely populációs szinten hivatott jellemezni az egészségi állapotot.

Visszatérve Janssen és Skiadas modelljére, az S -hez tartozó "átmenet" eloszlásfüggvényt jelölje P , azaz

$$P(x, t; z, \tau) = P[S(t) \leq x | S(\tau) = z], \quad (3.2)$$

ahol $\tau \leq t, x, z > 0$.

Továbbá legyen az átmeneti sűrűségfüggvény (megmutatható, hogy bizonyos regularitási feltételek mellett ilyen létezik):

$$p(x, t; z, \tau) = \frac{\partial}{\partial x} P(x, t; z, \tau).$$

Erre teljesül a Chapman-Kolmogorov egyenlet:

$$p(S_0, t_0; S, t) = \int_{-\infty}^{\infty} p(S_0, t_0; y, s) p(y, s; S(t), t) dy \quad (t_0 < s < t), \quad (3.3)$$

ahol legyen $p(S_0, t; S(t), t) = \delta(S(t) - S_0)$ és δ a Dirac-delta függvény.

A Kolmogorov-Chapman egyenlőség lényegében azt fogja meg, hogy a kiinduló t_0 időből és S_0 értékből a folyamat a t, S_t kettősbe a lehetséges y értékek egyikén keresztül kerül. Részletesebb levezetés a [5] forrásban található.

Folytonos idejű sztochasztikus folyamatok esetén a Kolmogorov-Chapman egyenlet a Fokker-Planck egyenletre redukálódik (ennek bizonyítása a [6]forrásban található):

$$\frac{\partial p(S, t)}{\partial t} = -\mu(t) \frac{\partial p(S, t)}{\partial S} + \frac{1}{2} [\sigma(t)]^2 \frac{\partial^2 p(S, t)}{\partial S^2}. \quad (3.4)$$

A megfelelő határfeltételek mellett mutatunk egy megoldást az előbbi Fokker-Planck egyenletre.

A határfeltételek legyenek a következők:

$$p(S(t), t_0; S_0, t_0) = \delta(S(t) - S_0), \quad \frac{\partial p(S(t), t_0; S_0, t)}{\partial S(t)} \rightarrow 0, \text{ ha } S(t) \rightarrow \pm\infty \quad (3.5)$$

A megoldás karakterisztikus függvények segítségével történik. Legyen

$$\Phi(s, t) = \int_{-\infty}^{\infty} p(S, t; S_0, t_0) \exp(isS) dS.$$

Állítás:

$$\frac{\partial \Phi}{\partial t}(s, t) = is\mu(t)\Phi(s, t) + \frac{1}{2}\sigma(t)^2 s^2 \Phi(s, t). \quad (3.6)$$

Bizonyítás: Parciális integrálások segítségével:

$$\begin{aligned} \frac{\partial \Phi}{\partial t}(s, t) &= \int_{-\infty}^{\infty} \frac{\partial p(S, t; S_0, t_0)}{\partial t} \exp(isS) dS = \\ &= -\mu(t) \int_{-\infty}^{\infty} \frac{\partial p(S, t_0; S_0, t_0)}{\partial S} \exp(isS) dS + \frac{1}{2} \sigma(t)^2 \int_{-\infty}^{\infty} \frac{\partial^2 p(S, t; S_0, t_0)}{\partial S^2} \exp(isS) dS = \\ &= -\mu(t) ([p(S, t; S_0, t_0) \exp(isS)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} p(S, t; S_0, t_0) \exp(isS) is dS) + \\ &+ \frac{1}{2} \sigma(t)^2 ([\frac{\partial p(S, t; S_0, t_0)}{\partial S} \exp(isS)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial p(S, t; S_0, t_0)}{\partial S} \exp(isS) is) = \\ &= \mu(t) is\Phi(s, t) + \frac{1}{2} \sigma(t)^2 ([p(S, t; S_0, t_0) \exp(isS) is]_{-\infty}^{\infty} - (-1) s^2 \Phi(s, t)) = \\ &= \mu(t) is\Phi(s, t) - \frac{1}{2} \sigma^2 s^2 \Phi(s, t) \quad \circ \end{aligned}$$

Az állításból és az első határfeltételből következik, hogy

$$\Phi(s, t_0) = \exp(isS_0).$$

Állítás:

$$\Phi(s, t) = \exp(is[S_0 + \int_{t_0}^t \mu(u) du] - \frac{1}{2} s^2 \int_{t_0}^t \sigma(u)^2 du) \quad (3.7)$$

Bizonyítás: Előző állítás alapján:

$$\frac{\partial \Phi}{\partial t} = is\mu(t) + \frac{1}{2}\sigma(t)^2s^2.$$

$$\log(\Phi) = \int_{t_0}^t is\mu(u) - \frac{1}{2}s^2\sigma(u)^2du + isS_0$$

És innen hatványra emelés után kapjuk az állítást. \circ

Ez éppen a karakterisztikus függvénye egy $S_0 + \int_{t_0}^t is\mu(u)du$ várható értékű és $\int_{t_0}^t \sigma(u)^2du$ szórású normális eloszlásnak. Következésképpen a (4) megoldása:

$$p(S, t; S_0, t_0) = \frac{1}{[2\pi \int_{t_0}^t \sigma(s)^2ds]^{\frac{1}{2}}} \exp\left[-\frac{[S - S_0 - \int_{t_0}^t \mu(s)ds]^2}{2 \int_{t_0}^t \sigma(s)^2ds}\right] \quad (3.8)$$

Tehát meghatároztuk az egyén egészségügyi állapotának alakulását leíró S folyamat átmenet sűrűségfüggvényét eddig. Ez lesz segítségünkre az eredeti célunk, vagyis a haláleset idejének eloszlásának meghatározásában. Ehhez azonban további egyszerűsítésként tegyük fel előbb, hogy a folyamat szórásparamétere konstans, azaz $\sigma(t) = \sigma$. Ekkor az átmeneti sűrűségfüggvény a következőképpen írható:

$$p(S, t; S_0, t_0) = \frac{1}{[2\pi\sigma^2(t - t_0)]^{\frac{1}{2}}} \exp\left[-\frac{[S - S_0 - \int_{t_0}^t \mu(s)ds]^2}{2\sigma^2(t - t_0)}\right] \quad (3.9)$$

A halál bekövetkeztét az S folyamat egy bizonyos α szint alá való csökkenéseként definiáltuk, ennek időpontját jelölje T . Azaz az α szint T elérési ideje: $T = \inf\{t : S(t) \leq \alpha\}$. A feladat T sűrűségfüggvényének meghatározása $t_0 \leq t$ -re:

$$g(\alpha, t; S_0, t_0) = \frac{\partial}{\partial t} Pr\{T \leq t\}$$

Az adott S_0 kiindulási pontból az α szint elérési idejének sűrűségfüggvénye lineáris drift és konstans szórás esetén:

$$g(\alpha, t; S_0, t_0) = \frac{|\alpha - S_0|}{t - t_0} p(S, t; S_0, t_0)$$

Jenssen és Skiadas ennek segítségével közelíti az esetünkben az elérési idő sűrűségfüggvényét, behelyettesítve (3.9)-et az előbbi kifejezésbe:

$$g(\alpha, t; S_0, t_0) = \frac{|\alpha - S_0|}{[2\pi\sigma^2(t - t_0)^3]^{\frac{1}{2}}} \exp\left[-\frac{[\alpha - S_0 - \int_{t_0}^t \mu(s)ds]^2}{2\sigma^2(t - t_0)}\right] \quad (3.10)$$

Az S folyamat az egészségügyi állapotot az egyén szintjén jellemzi. Ahhoz, hogy a teljes populáció egészére jellemző mennyiséget kapjunk, ezeknek egy aggregációjára van szükségünk. Jelöljük

ezt a folyamatot H -val és definiáljuk az S folyamat várható értékeként minden t időpontban:

$$H(t) = E(S(t)) = \int_{t_0}^t \mu(s) ds \quad (3.11)$$

Ezt a $H(t)$ függvényt nevezzük "egészségállapot függvénynek".

Egy másik közelítési módszer, amit használnak az úgynevezett "tangent approximation". Eszerint S_0 kiindulási érték, $\alpha = 0$ elérési szint és $t_0 = 0$ esetén az elérési idő sűrűségfüggvényét a következőképpen közelítjük:

$$\begin{aligned} g(0, t; S_0, 0) &= \frac{|H(t) - tH'(t)|}{t} p(0, t; S_0, 0) = \\ &= \frac{|H(t) - tH'(t)|}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{H(t)^2}{2\sigma^2 t}\right) \end{aligned} \quad (3.12)$$

A $|H(t) - tH'(t)|$ tag használatát indokolhatja, ha felírjuk a következő Taylor-sort:

$$H(t) - tH'(t) = H(0) + \frac{t^2}{2}H'' + \frac{t^3}{3!}H''' + \dots$$

Az eltérés az első megközelítéshez képest tehát H'' nagyságrendű.

3.3. Kiegészített modellek

A szakdolgozat későbbi részeiben az előző alfejezetben felírt modelleket fogjuk használni, azonban Skiadas ezek módosított verzióit is alkalmazza. Ezek kerülnek összefoglalásra röviden ebben a részben.

A (12)-es formula egyszerűsítésének érdekében a számlálóban szereplő $|H(t) - tH'(t)|$ kifejezést helyettesítsük egy k paraméterrel:

$$\begin{aligned} g(0, t; S_0, 0) &= \frac{k}{t} p(0, t; S_0, 0) = \\ &= \frac{k}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{H(t)^2}{2\sigma^2 t}\right) \end{aligned} \quad (3.13)$$

Ekkor ismert g sűrűségfüggvény esetén adott k -ra kifejezhető az ismeretlen $H(t)$ függvény a következő módon:

$$H(t) = \pm \left| \left(-2t \log \frac{g(t) \sqrt{2\pi t^3}}{k} \right)^{\frac{1}{2}} \right|$$

A k paraméter becslésére Skiadas a [4] forrásban mutat be egy lehetséges módszert. További kiegészítésként szolgálhat a (12) és (13) formula közti eltérés kezelésére a következő $f(t)$ korrekciós taggal felírt modell:

$$g(0, t; S_0, 0) = \frac{k}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(H(t) + f(t))^2}{2\sigma^2 t}\right)$$

Az $f(t)$ korrekciós tag alakjára kényelmesnek bizonyult a következő használata:

$$f(t) = at^b$$

ahol az a és b paraméterek meghatározása szimulációs módszerekkel történik.

A szakdolgozatban ezeknek a kiegészítéseknek az alkalmazását nem tartottuk szükségesnek, ugyanis nélkülük is sikerült kellően pontos illeszkedést elérni.

4. fejezet

A Skiadas-féle modell

4.1. Az egészségállapot függvény paraméterezése

Az előző fejezetben bevezettük az egyén egészségi állapotának alakulását leíró S sztochasztikus folyamatot a következőképpen:

$$dS(t) = \mu(t)dt + \sigma dW(t)$$

ahol μ és σ fejezi ki az infinitezimális átlagos változását és varianciáját az egészségállapotnak. A σ variancia paramétert konstansnak tekintettük. A $H(t)$ egészségállapot függvényt az S folyamat várható értékeként definiáltuk:

$$H(t) = ES(t) = S(0) + \int_{t_0}^t \mu(s)ds$$

Az egészségfüggvény $H(t)$ és $\mu(t)$ közti összefüggés másképp felírva:

$$\mu(t) = \frac{\partial H(t)}{\partial t}$$

A fő feladat a $\mu(t)$ függvényt megadni analitikusan. Azt feltételezzük, hogy az egészségállapot jellemzően úgy alakul, hogy születés után gyorsan javul, majd enyhül ez a javulás. Egy bizonyos kor után az egészségi állapot elkezd elromlani. Ennek megfelelően $\mu(t)$ -nek az emberi élettartam két elkülönülő időszakát kell jól jellemeznie. Az első időszakot jellemezze az $u = u(1/\sqrt{t})$ függvény, amelynek gyorsan csökkenőnek és pozitívnak kell lennie az időben, utalva a születés utáni lassuló ütemű fejlődésre. A második időszakot leíró $v = v(t)$ függvény pedig az először fokozatosan javuló, majd romló egészségi állapotot kell megfogja a teljes élettartamra.

Ezen megfontolások nyomán Janssen és Skiadas a következőképpen parametrizálja $\mu(t)$ -t:

$$\mu(t) = u\left(\frac{1}{\sqrt{t}}\right) + v(t), \tag{4.1}$$

ahol

$$u\left(\frac{1}{\sqrt{t}}\right) = b_0 + \frac{1}{2} \frac{b_1}{\sqrt{t}}$$

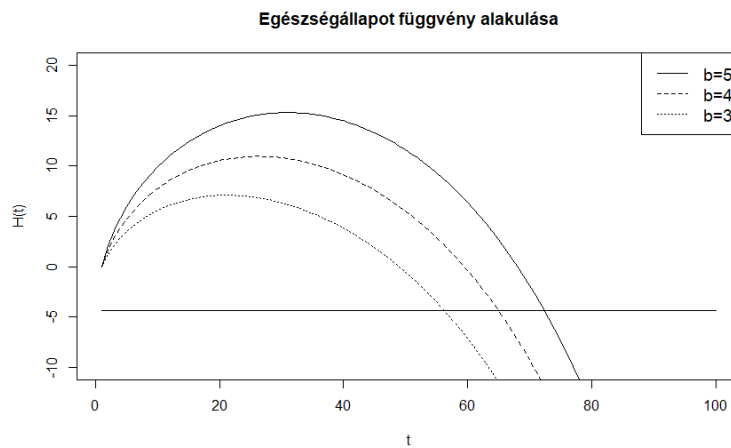
$$v(t) = a_0 + 2a_1t + 3a_2t^2 + 4a_3t^4$$

Továbbá feltesszük, hogy a_0 és b_0 egyenlőek nullával, mert $t = 0$ körül nem adnak hozzá sokat μ értékéhez, mivel az $\frac{1}{\sqrt{t}}$ -s tag miatt az nagyon magas. Tehát ezzel a közelítéssel élve és feltéve, hogy $t_0 = 1$ és $S_0 = 0$, kapjuk a következőt a (11) egyenletből:

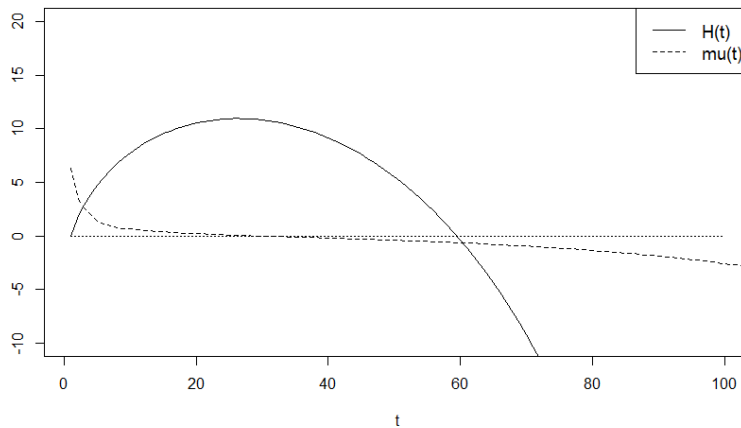
$$H(t) = E[S(T)] = \int_1^t \mu(s)ds = a_1t^2 + a_2t^3 + a_3t^4 + b_1\sqrt{t} - c \quad (4.2)$$

ahol $c = a_1 + a_2 + a_3 + b_1$.

A következő ábrán rögzített a_i és α paraméterek mellett, különböző b -k esetén ábrázoltuk a $H(t)$ függvényt. Tetszőlegesen lettek megválasztva a paraméterek: $a_1 = -10^{-2}$, $a_2 = 10^{-3}$, $a_3 = -10^{-4}$ és $\alpha = -4.4$.



A következő ábrán pedig a $b = 5$ esetben $H(t)$ és $\mu(t)$ látható:



Egy személy halálának bekövetkeztét úgy definiáltuk, hogy az S folyamat értéke valamelyik időpontban az α szint alá csökken. A halál idejének eloszlására kapott (3.10)-es formulába behelyettesítve (4.2)-t:

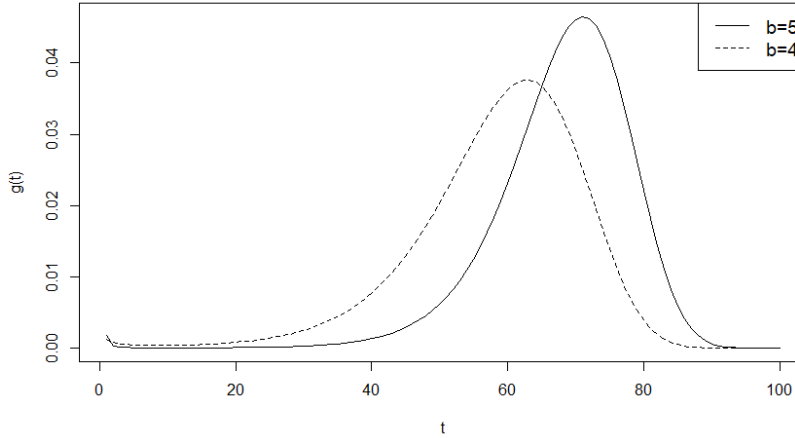
$$g(0, t; S_0, 0) = k \frac{|\alpha|}{\sqrt{2\pi t^3 \sigma^2}} \exp\left(-\frac{[\frac{\alpha+c}{\sqrt{t}} - (b_1 + a_1 t \sqrt{t} + a_2 t^2 \sqrt{t} + a_3 t^3 \sqrt{t})]^2}{2\sigma^2}\right) \quad (4.3)$$

ahol $c = a_1 + a_2 + a_3 + b_1$ továbbra is és k egy normalizáló konstans, amely az alábbi formulából kapható meg:

$$\int_0^\infty g(s) ds = 1 \quad (4.4)$$

Látható, hogy ha minden paramétert elosztunk σ -val, akkor ugyanazt a kifejezést kapjuk a sűrűségfüggvényre, ezért feltehetjük, hogy $\sigma = 1$.

A következő ábrán a korábban ábrázolt egészségfüggvényekhez tartozó sűrűségfüggvényeket ábrázoltuk (csak a $b = 4$ és $b = 5$ eset):



A modell valós adatokra történő illesztéséről és a paraméterek meghatározásának módjáról egy későbbi fejezetben lesz szó részletesebben. Előljáróban annyit említenénk meg, hogy az illeszkedés pontos volt, azonban célunk, hogy csökkentjük a paraméterek számát. Ezáltal inkább összehasonlítható lenne másik egyszerűbb modellel, amelyek paraméterszáma nem különbözik nagy mértékben. Másfelől ez a modell kezeli a teljes emberi élettartamot, beleértve a csecsemőkori halandóságot is, amit az egyszerűbb modellek használatával kevésbé lehet.

Mutatunk egy másik lehetséges paraméterezését a $H(t)$ egészségállapot függvénynek, amelyben paraméterek száma már kevesebb lesz, ugyanakkor nem lesz alkalmas a fiatalok egészségi állapot és mortalitás jellemzésére. Az észrevétel az lenne, hogy a (15) kifejezésben a $H(t)$ függvény csökkenését leginkább a nagyobb kitevős tagok befolyásolják, ezért egy pontos közelítés lehetne a következő paraméterezés:

$$H(t) = c - (lt)^b \quad (4.5)$$

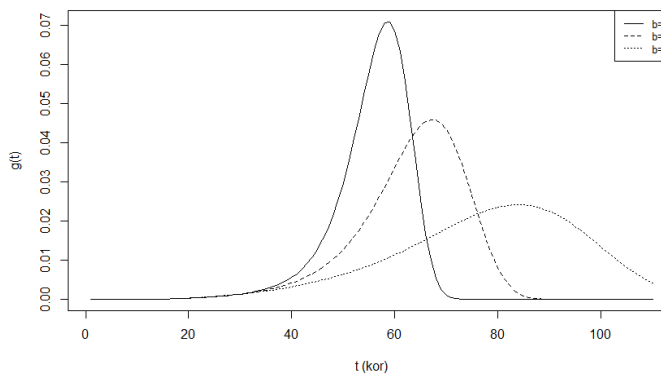
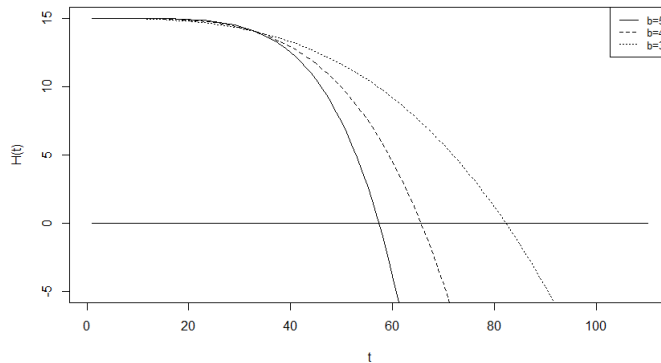
ahol l, c és b pozitív paraméterek.

A b paraméter pozitivitása garantálja, hogy $H(t)$ egy csökkenő függvény (és b -ben csökkenő). A c paraméter az egészségállapot függvény kiindulási szintje $t = 0$ -ban. A halál bekövetkeztére definiált α szintet tekinthetjük 0-nak, ugyanis α közvetetten benne van c -ben, hiszen legyen a paraméterek közti összefüggés $c = S(0) - \alpha$. Azaz ugyanaz, ha $S(0)$ -ból indulva α -t kell elérni, mintha a $c = |S(0) - \alpha|$ -ből indulva a 0 szintet.

Ezzel a (4.5)-ban definiált $H(t)$ egészségállapot függvénnyel kifejezve a haláleset, azaz a 0 szint elérési idejének eloszlásának a sűrűségfüggvénye (3.12) alapján (és feltételezve, hogy $\sigma = 1$):

$$\begin{aligned} g(0, t; c, 0) &= \frac{|c - (lt)^b + tlb(lt)^{b-1}|}{t} p(0, t; c, 0) = \\ &= \frac{|c + (b-1)(lt)^b|}{t} p(0, t; c, 0) = \\ &= \frac{|c + (b-1)(lt)^b|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(c - (lt)^b)^2}{2t}\right) \end{aligned} \quad (4.6)$$

A következő ábrák közül az elsőn a (18) paraméterezéssel felírt egészségállapot függvényeket ábrázoljuk különböző b paraméterek esetén, rögzített c és l paraméter mellett ($c = 15$ és $l = 0.03$). A másodikon pedig az ezekhez tartozó g sűrűségfüggvényei a 0 szint elérési idejének eloszlásaihoz.



4.2. Paraméterbecslés

Miután az előző alfejezetben több lehetséges paraméterezés esetén is kifejeztük az életkorok eloszlásának sűrűségfüggvényét, most rátérnénk a paraméterek becslésének lehetséges módszereire. Tegyük fel, hogy a rendelkezésünkre álló halandósági táblában szerepelnek minden életkorra (0-110 év), hogy hányan hunytak el és mennyien voltak megfigyelve. Jelölje a j -edik életkor esetén d_j a halottak, l_j pedig a megfigyelt leélt évek számát. Legyen továbbá

$$y_j = \frac{d_j}{\sum_{i=0}^{110} d_i}$$

azaz a haláleset bekövetkezési idejének tapasztalati eloszlása: $P(x$ évesen hal meg valaki) $= y_x$. Janssen és Skiadas a következőképpen becsüli a paramétereket. A j éves kori haláleset valószínűségét közelítsük úgy, hogy a g sűrűségfüggvényt tekintsük állandónak az adott életkornak megfelelő szakaszon, így $P(x$ évesen hal meg valaki) $\approx g(j)$.

Ekkor pedig keressük azokat a paramétereket, amelyekre a tapasztalati és a becsült valószínűségek közti négyzetes eltérés minimális, azaz a második paraméterezést tekintve:

$$\min(y_j - g(j))^2 = \min\left(y_j - \frac{|c + (b-1)(lt)^b|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(c - (lt)^b)^2}{2t}\right)\right)^2$$

Az optimális paraméterek meghatározása iteratív módszerrel történt.

Egy másik lehetőség, hogy ha vannak megfigyeléseink az adott életkorban megfigyelt leélt évek számát és az adott életkorban meghaltak számát illetően, akkor maximum-likelihood becslést adunk a paraméterekre:

$$L = \prod_{j=j_0}^{j_{max}} \binom{l_j}{d_j} p_j^{d_j} (1 - p_j)^{l_j - d_j} \quad , \quad (4.7)$$

ahol l_j a megfigyelt leélt évek számát a j -edik életkorban, d_j a megfigyelt halálesetek számát a j -edik évben és p_j a halálozási valószínűség a j -edik évben. A loglikelihood tehát logaritmálás után így fejezhető ki:

$$l = \sum_{j=j_0}^{j_{max}} konst + \log(p_j) d_j + \log(1 - p_j) (l_j - d_j) \quad .$$

A p_j valószínűségek becslése viszont nem triviális. Tegyük fel, hogy j_0 az első vizsgált életkor (nem feltétlen célunk születéstől modellezni a halandóságot). Ekkor annak a valószínűsége, hogy j_0 évesen veszti valaki életét közelíthető g_{j_0} -vel. A $j_0 + 1$ éves korban történő haláleset valószínűsége:

$$p_{j_0+1} = P(j_0 + 1 \text{ évesen hal meg}) / P(j_0 \text{ évesen él}) = \frac{g_{j_0+1}}{(1-g_{j_0})}$$

Rekurzívan minden j -re meg lehet határozni p_j -ket:

$$p_{j+1} = \frac{g_{j+1}}{\prod_{i=j_0}^j (1 - p_i)}$$

Ez a rekurzív felírás sokat lassíthat a likelihood minimalizálásakor, ezért számoljuk máshogy a likelihoodot. Az eddigi jelöléseknek megfelelően legyen a kezdeti j_0 korban a megfigyelt emberek száma legyen L_{j_0} . Annak a valószínűségét, hogy j éves korban fog elhunyni valaki, közelítsük $g(j)$ -vel megint. Ekkor a likelihood a következő formában is felírható:

$$L = l_{j_0}! \prod_{j=j_0}^{j_{max}} \frac{g(j)^{d_j}}{d_j!} \times \frac{(1 - g(j_0) - \dots - g(j_{max}))^{(l_{j_0} - d_{j_0} - \dots - d_{j_{max}})}}{(l_{j_0} - d_{j_0} - \dots - d_{j_{max}})!}$$

ahol d_j továbbra is a j évesen elhunytak száma. A likelihoodban szereplő szorzó tag a jobb oldalon azokra vonatkozik, akik nem halnak meg j_{max} kor előtt, hanem tovább élnek.

Ekkor a loglikelihood:

$$l = konst + \sum_{j=j_0}^{j_{max}} d_j \log(g(j)) + (l_{j_0} - d_{j_0} - \dots - d_{j_{max}}) \log(1 - g(j_0) - \dots - g(j_{max}))$$

Ezt maximalizálva kaphatunk becsléseket a paraméterekre. A maximalizálás ebben az esetben is iteratív módszerekkel történhet.

4.3. A modell illesztése valós adatokra

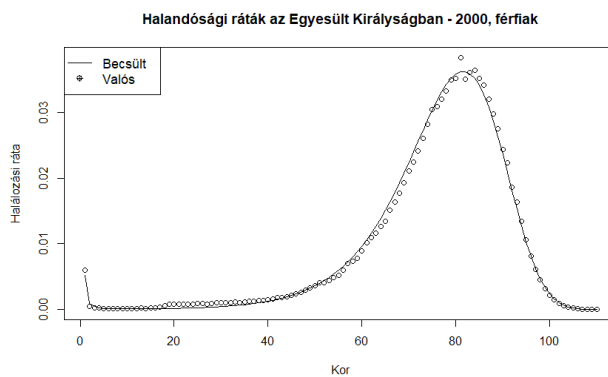
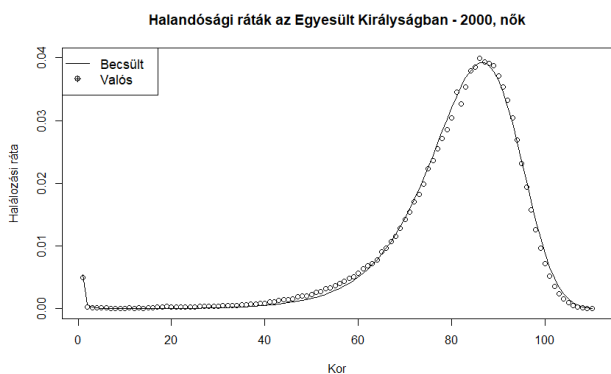
Ebben az alfejezetben a korábban taglalt lehetséges modellek és becslések illesztését végezzük el valós adatokra. Az adatok forrása a Human Mortality Database [8], ahol több országra érhetőek el halandósági adatok évtizedekre visszamenőlegesen. Ebben a részben a brit lakosság 2000-es adatait használjuk, külön becsüljük a női és férfi halandóságot és összehasonlítjuk azokat.

Először tekintsük a (4.2) egyenlet szerint parametrizált egészségfüggvényt, azaz amikor

$$H(t) = a_1 t^2 + a_2 t^3 + a_3 t^4 + b_1 \sqrt{t} - c$$

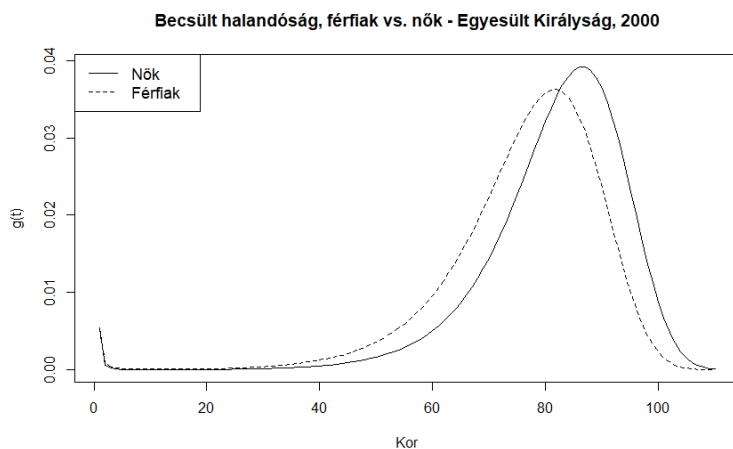
ahol $c = a_1 + a_2 + a_3 + b_1$.

Továbbá a paraméterek becslését most az előző fejezetben taglalt Skiadas és Janssen féle négyzetes eltérés minimalizálásával tesszük meg. A következő két ábrán látható a valós halandósági rátákra a becslések illeszkedése.

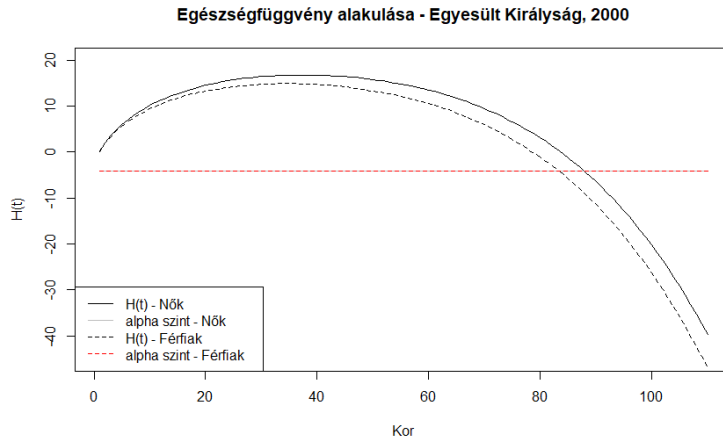


Észrevehető, hogy a modell képes kezelni a csecsemőkori halandóságot valamilyen szinten. A paraméterek meghatározására és a négyzetes eltérés minimalizálására az R programcsomag beépített iteratív optimalizáló eszközét, az *optim* függvényt használtuk. A kapott paraméterek nagy mértékben függtek a kiindulási értékektől, amikkel az iterációt indítottuk. Előfordult olyan kifutás is, amikor az illesztés nem kezelte jól a csecsemőkori halandóságot.

A becsült halandóságokat hasonlítsuk össze a nők és férfiak esetén. Egy képen ábrázolva a kettő:



A nők halandósága köztudottan jobb a férfiakénál az európai országok tekintetében. Ezt a becsült eloszlásból is láthatjuk, hiszen a nők halandósága "jobbra tolódott" a férfiakéhoz képest. Mielőtt megnézzük, hogy ebben az esetben ez melyik paraméterek értékének következménye, ábrázoljuk a $H(t)$ egészségfüggvényt és a haláleset definiálásához használt α szintet a két esetben:



A nők egészségfüggvénye gyakorlatilag az élettartam folyamán mindig magasabb értéket vesz fel, mint a férfiaké és később metszi az elérendő α szintet. Az ábrán talán nehezen vehető ki, hogy a férfiakhoz és nőkhöz különböző α szint tartozik, mert a kettő becslése nagyon közel esik egymáshoz (kerekítve nők esetén -4.17 , férfiaknál -4.14).

A következő táblázatban a becsült paraméterek értékeit és néhány egyéb információt foglalkoz össze az illesztésről.

Egyesült Királyság, 2000		
	Nők	Férfiak
α	-4.168	-4.138
b	5.092	4.735
$a_1 \times 10^2$	-1.001	-1.000
$a_2 \times 10^4$	1.246	1.227
$a_3 \times 10^7$	-9.063	-9.151
$MSE \times 10^{-5}$	5.635	5.492
H_{max}	16.81	14.92
$T : H_T = H_{max}$	37	35
$T : H_T = \alpha$	88	84

Az egyik paraméter, amely értelemszerűen jelezhetné, hogy egy populáció halandósága jobb a másikénál, ha az α szint abszolút értékben nagyobb lenne. Ez most teljesül is a női és férfi lakosság összehasonlításakor, azonban az eltérés mértéke elég kicsi. A kapott paraméterek alapján az egyetlen, ahol nagyobb eltérés van, az a b paraméter. A nők esetében ez nagyobb értéket vesz fel, és az előző fejezetekre hivatkozva ez felel az alacsonyabb becslött rátákért a nők esetében.

A nők esetén az átlagos élettartam a becslött rátákkal 81.1 év, a valós rátákkal 80.2 év volt. A férfiak esetén pedig a becslött 75.67 év, a valós 75.38 év volt.

Alkalmazzuk azt a modellt is, amikor az egészségállapot függvény alakja (4.5) szerinti, azaz

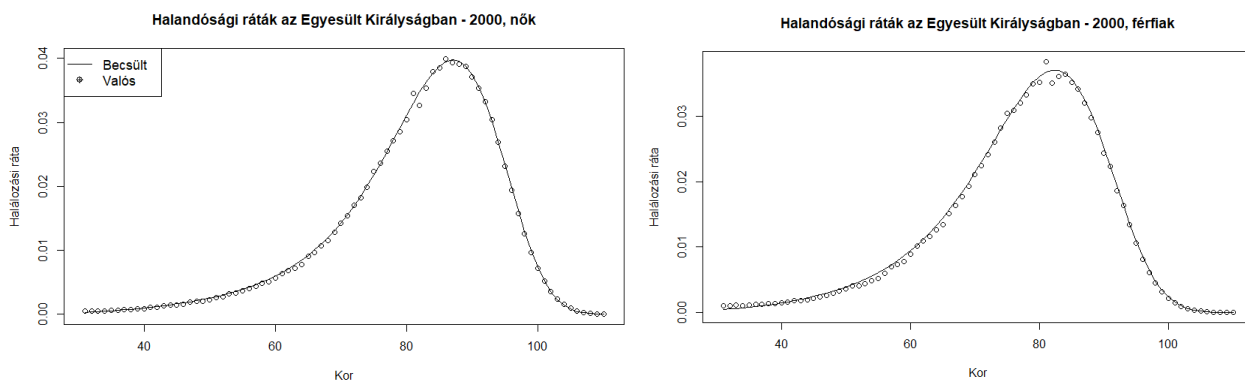
$$H(t) = c - (lt)^b \quad (4.8)$$

ahol c, b és l a becsülendő paraméterek. Ekkor az élettartam eloszlásának sűrűségfüggvénye a következőképpen volt megadható:

$$g(0, t, c, 0) = \frac{|c + (b - 1)(lt)^b|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(c - (lt)^b)^2}{2t}\right) \quad (4.9)$$

Ellentétben az előző modellel, ez nem képes kezelni a csecsemőkori halandóságot. Emiatt a modell illesztését nem a teljes élettartamra végezzük el, hanem a kezdeti életkort 30 évnél választjuk meg. Ennek oka a csecsemőkori halandóságon kívül a 17-25 éves kor körüli lakosság kiugróan rosszabb halandósága, amely többek között a friss jogosítvánnyal rendelkezők magasabb balesetszámának lehet eredménye.

Most is a brit lakosság adataira végezzük el az illesztést. Az alábbi két ábrán láthatóak külön a férfiak és nők esetén a becsült és valós ráták:



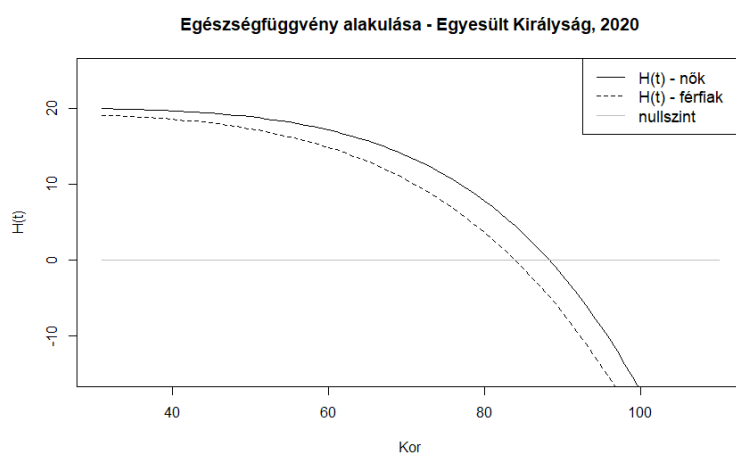
Az ábrák alapján most is egészen pontos illeszkedést sikerült elérni és a következő táblázatban összefoglaljuk az illesztéssel kapcsolatos többi információt:

Egyesült Királyság, 2000		
	Nők	Férfiak
c	20.115	19.373
b	4.987	4.381
$l \times 10^2$	2.067	2.343
$MSE \times 10^{-5}$	1.360	2.317
$T : H(T) = 0$	89	84

A nők lényegesen jobb halandóságát ez a modell is jelzi. A paraméterek szintjén ez az egészségállapot függvény c kiindulási szintjének magasabb értékéből és a b hatványparaméter értékéből látszik.

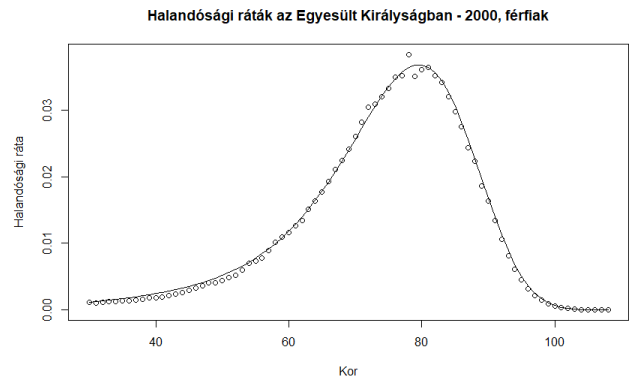
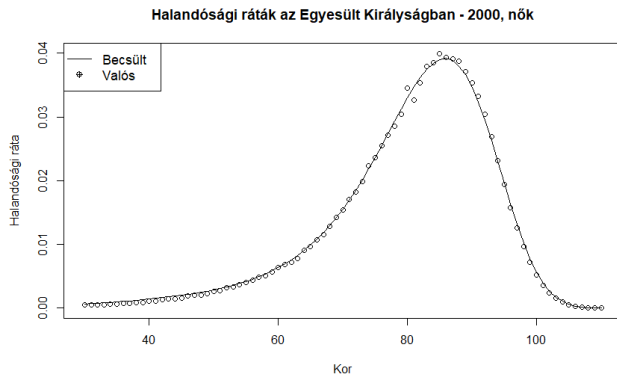
Az előző modell esetén a halál bekövetkezteként a definíciója az α szint elérése volt. A mostaninál $\alpha = 0$. Összehasonlítva, hogy a két modell esetén a $H(t)$ függvény az α szintet melyik kornál éri el először, akkor közel ugyanazt kapjuk: a nők esetén a második modell egyetlen életévvel később lépi át a szintet.

A teljesség kedvéért ábrázoljuk a második modell esetén a két egészségállapot függvény alakulását a különböző életkorokra.



A következőekben az előző, három paraméteres modellt alkalmazzuk, azonban a Skiadas és Janssen által végzett négyzetes eltérés minimalizálása helyett a maximum-likelihood alapú paraméterbecslést végezzük el. A célunk tehát a (4.7) kifejezés szerinti likelihood maximalizálása a c , b és l paraméterek szerint.

Mivel az idősebb korokban viszonylag kevés megfigyelés és magas halálozási valószínűségek vannak, ezért ezekre az illesztést a 30-90 éves korosztályon végeztük. Tekintsük most is a korábbiaknak megfelelő ábrákat és táblázatot:



Egyesült Királyság, 2000		
	Nők	Férfiak
c	15.13	14.717
b	4.742	4.285
$l \times 10^2$	2.148	2.429
$MSE \times 10^{-5}$	2.019	2.944

A paraméterek értékei nagyobb mértékben változtak azokhoz képest, mint amit a négyzetes eltérés minimalizálásával kaptunk. Azonban a kiindulási c szint és a b paraméter a nőknél továbbra is magasabb értéket vesz fel, amelynek eredménye, hogy a becsült halandóság most is jobb a nők esetén, mint a férfiaknál.

A kapott paraméterekkel számolt becsült és valós halandósági ráták négyzetes eltérését ebben az esetben is megnéztük (konkrét értékeket lásd a táblázatban). A négyzetes eltérés a férfiaknál és a nőknél is nagyobb volt, mint a másik módszer esetén - de ez kevésbé meglepő, mivel ott a minimalizálandó célfüggvény pont az volt.

Összességében sikerült pontos illeszkedést a modelljeink és a rendelkezésre által adott adatok között az Egyesült Királyság 2000. évi lakosságát illetően. A dolgozat további részében az lesz a célunk, hogy a mortalitás változását az évek folyamán modellezzük az eddig bemutatott modellek segítségével.

5. fejezet

Az egészségállapot alakulása

Az előző fejezetekben több modellt felírtunk és alkalmaztunk egy adott populáció élettartamának leírására. Ezek a modellek jól használhatóak voltak, amikor egy év adatait tekintettük. A következőekben a fő célunk az lesz, hogy a megfigyelt ország halandóságának évtizedek alatti alakulását modellezzük.

Adatainkat most is a Human Mortality Database-ről kérjük le, mégpedig most a svéd lakosság halandóságát fogjuk vizsgálni. Ennek oka az, hogy Svédországból az 1750-es évekig visszamenőlegesen jó minőségű halandósági adatok állnak rendelkezésre, amelyen jól lehet modellezni a mortalitás hossz távú alakulását.

A korábban ismertetett modellek közül azt fogjuk alkalmazni, amelyik esetén az egészségállapot függvény paraméterezése a következő:

$$H(t) = c - (lt)^b$$

ahol c , l és b paraméterek.

Tegyük fel, hogy rendelkezésre állnak adatok egymást követő naptári évek egy I és életkorok egy J halmazára. Annak érdekében, hogy a halandóság és az egészségállapot függvény változását modellezzük az évek alatt, a paramétereket naptári évekhez és kohortokhoz rendeljük. A vizsgált I naptári évekhez tartozzon külön-külön más kiindulási szint, azaz egy c_i paraméter ($i \in I$). A kohortokhoz pedig l_{i-j} legyen a vonatkozó paraméter ($j \in J$). Vagyis az i -edik naptári évben egy j éves ember egészségállapotát a $c_i - (l_{i-j}j)^b$ kifejezés írja le.

A c paraméterek tehát egy adott évre jellemző hatásokat hivatottak jellemezni, például ha volt háború vagy komolyabb járvány, akkor az évhez tartozó paraméter esetén alacsonyabb értéket várnánk. A kohort paraméterekben való kiugró változás pedig jelezhet egy-egy generációs hatást akár, vagy hogy a kohorthoz tartozó születési évben például olcsóbbak lettek a dohánytermékek vagy éppen bevezették az alkoholtilalmat.

Azokat a kohortokat tekintjük csak, amelyekhez tartozó születési évek halmaza megegyezik a vizsgált naptári évekkel. Tehát a legkisebb naptári év esetén egyetlen megfigyelés áll rendelkezésre.

zésre valójában, mégpedig a legfiatalabb korhoz tartozó, ugyanis az annál idősebbek egy nem vizsgált kohorthoz tartoznak. Hasonló a helyzet az l_k paraméterek esetén, csak ott a legnagyobb vizsgált évnél rendelkezünk egyetlen megfigyeléssel.

A paramétereket a 4. fejezetben leírtakhoz hasonlóan becsüljük maximum-likelihood módszerrel. A maximalizálandó likelihood ekkor:

$$L^* = \prod_{k \in I} N_k! \prod_{i,j:i-j=k} \frac{g_k(j)^{d_{ij}}}{d_{ij}!} \times \frac{(1 - g_k(j_0) - \dots - g_k(j_{max}))^{N_k - d_{ij_0} - \dots - d_{ij_{max}}}}{(N_k - d_{ij_0} - \dots - d_{ij_{max}})!} .$$

ahol k a vizsgált kohortokat jelenti, N_k a k -adik kohort létszáma j_0 éves korban, j_0 és j_{max} a vizsgált minimális és maximális életkor, $g_k(j)$ az a valószínűség, hogy a k -adik kohortból egy személy j évesen hal meg és d_{ij} az i -edik évben a j éves korúak halálozási száma.

A loglikelihoodot az előbbi kifejezés logaritmálása után a következőképpen írhatjuk fel:

$$l^* = \sum_{k \in I} \sum_{i,j:i-j=k} \log(g_k(j))d_{ij} + \log(1 - g_k(j_0) - \dots - g_k(j_{max}))(N_k - d_{ij_0} - \dots - d_{ij_{max}}) + konst . \quad (5.1)$$

A $g_k(j)$ valószínűségeket közelítjük a $H(t)$ egészségállapot függvényből adott b, c és l paraméter esetén becsült értékéből:

$$g_k(j) \approx g(0, t, c, 0) = \frac{|c_i + (b-1)(l_k t)^b|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(c_i - (l_k t)^b)^2}{2t}\right) \quad (5.2)$$

ahol $k = i - j$ továbbra is.

Megjegyezzük, hogy egy kohort népességszáma nem kizárólag csak a halálozások miatt változhat. Ennek oka lehet az adott országból történő kivándorlás illetve bevándorlás. Emiatt nem teljesen pontos a fenti likelihood, azonban ezt a hatást kellően kicsinek véljük.

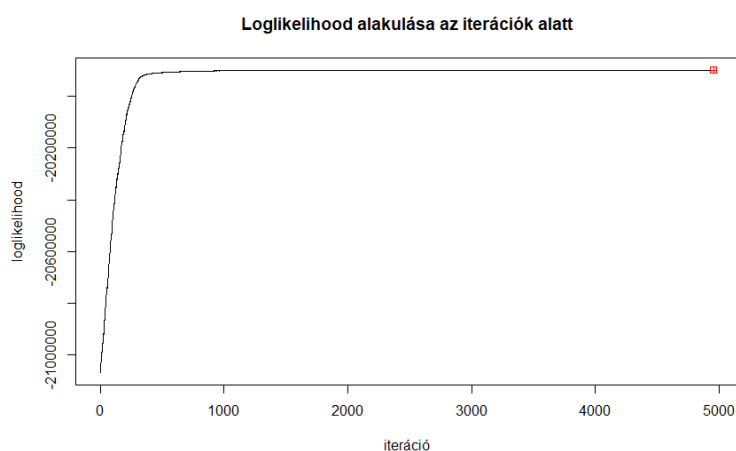
A b, c_i és l_k paraméterek maximum-likelihood becslését zárt alakban bonyolult lenne felírni, ezért meghatározásuk sztochasztikus optimalizálással történt a következőképpen. A paramétereknek olyan kezdeti értékeket adunk, amelyek nagyságrendileg körülbelül összhangban vannak azokkal az illesztésekkel, amikor egyetlen naptári évet tekintettünk (lásd korábbi fejezetek). Több futtatást végeztünk másik kezdeti értékekkel, de az optimalizálás végeredménye közel ugyanaz lett. Egy iteráción belül először egy adott paraméterhez hozzáadunk egy normális eloszlású változást. Ez a normális eloszlás mindig 0 várható értékű, a szórásnégyzete pedig a paramétertől függött. Megnézzük, hogy ez a paraméterváltozás mennyire mozdítja el a loglikelihoodot: ha "megfelelő" mértékben és irányban, akkor felülírjuk a paraméter korábbi értékét, ha nem, akkor pedig lépünk a következő paraméterre. Ezt konkrétan az alábbi módon formalizáljuk:

Legyen $U \sim E[0, 1]$ egyenletes eloszlás, l_0 a loglikelihood értéke a paraméter változtatása előtt, l_1

a loglikelihood értéke miután a normális változást hozzáadtuk az eredeti paraméterértékhez. Az adott paraméter értékét akkor írjuk felül, ha $e^{l_1-l_0} > U$. Másképpen fogalmazva $\max(1, e^{l_1-l_0})$ valószínűséggel fogadjuk az új paramétert. Miután az összes paramétert esetlegesen változtattuk, a következő iterációban ugyanezt megejtjük, és egészen addig iterálunk, amíg a likelihood értéke egy nagyjából állandó szinten beáll. Ezután végzünk még további N darab iterációt és azokra átlagoljuk a paraméterek értékét, azt elfogadva maximum likelihood becslésnek.

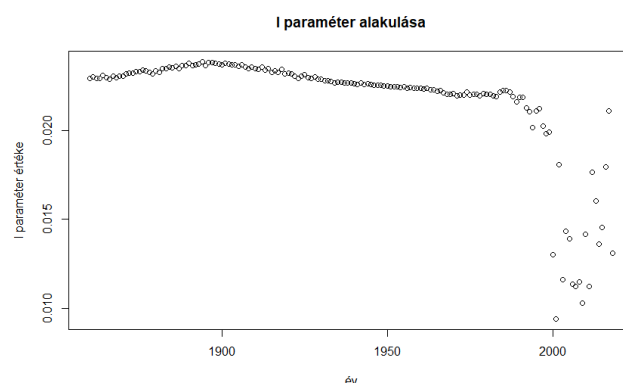
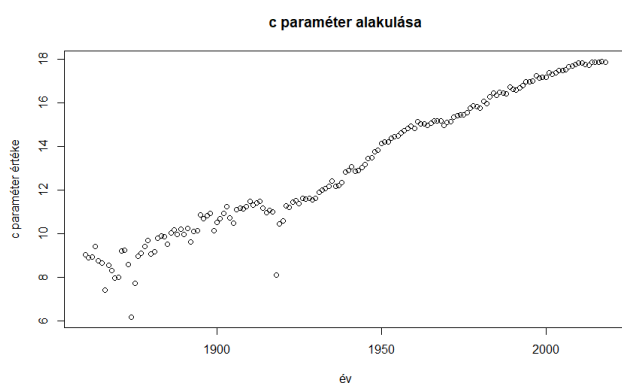
Először a svéd nők adataira végeztem el illesztést. Az 1860-2018 közti éveket vettem figyelembe. A vizsgált életkorok megint 30-tól 90 éves korig voltak.

Az (5.1) kifejezés iterációs optimalizálása során a likelihood értékének alakulása a következő ábrán van szemléltetve:



Az jobb felső sarokban szereplő piros jelzés jelöli a végső becslés likelihoodjának értékét. Az optimalizálás során 1650 iterációt végeztünk és az utolsó 300 iterációban szereplő paraméterértékek átlagaként adtuk meg a végső becslést (az ábrán azért látható 4950 iteráció, mert ott minden iterációhoz 3 változás van ábrázolva: a b , c és l paraméterek frissítése után a likelihood változása).

A becsült b paraméter értéke 4.306, a c kiindulási szint és l paraméterek alakulása az évek során a következő két ábrán látható:



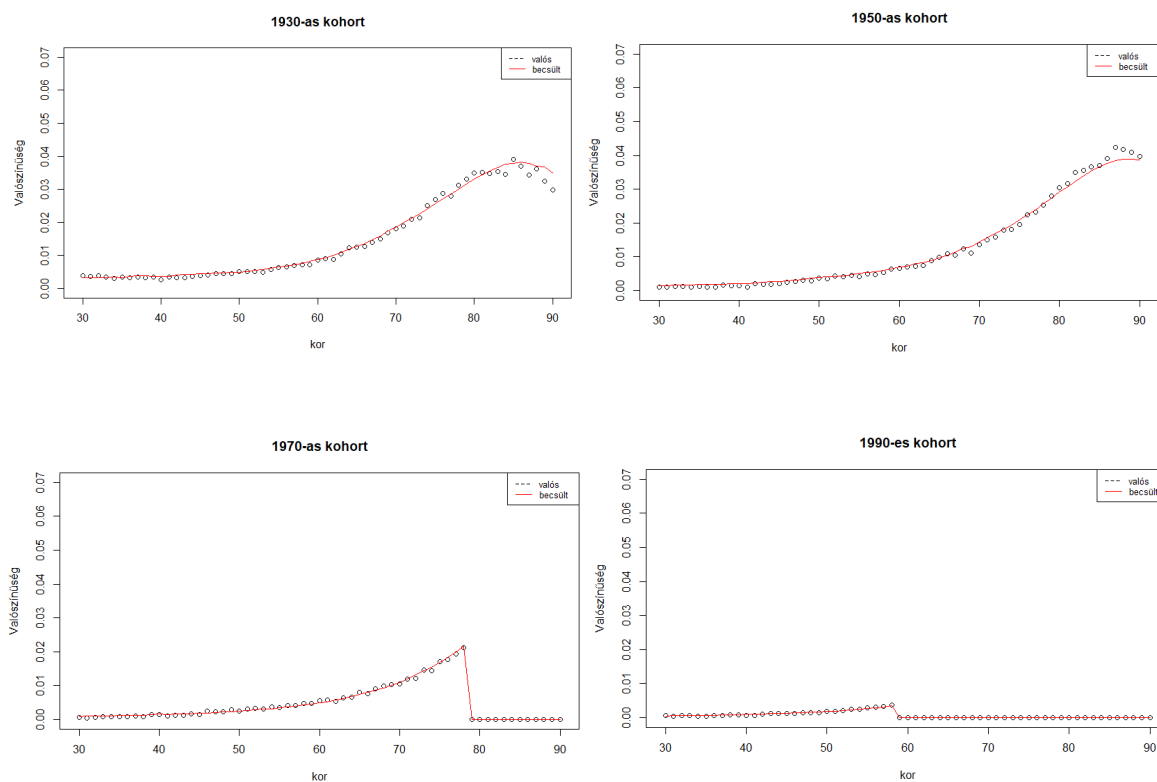
Az ábra alapján is látható, hogy a kezdeti évekhez tartozó c és az utolsó évekhez tartozó l paraméterek szóródása nagy. Ennek oka az lehet, amit már korábban is említettünk, miszerint ezekre a paraméterekre lényegesen kevesebb megfigyelés jut, amelyet a becsléskor figyelembe veszünk. Tehát például a 2018-as kohort esetén csak 2018-ra, azaz a 0-1 éves korosztályra van megfigyelésünk.

Az egészségállapot függvény kiindulási szintjében, azaz a c paraméter értékében megfigyelhető egy trendszerű emelkedés. Ez megfelelhet annak, hogy az évek során a halandóság jellemzően javul a folyamatosan fejlődő egészségügynek és az emberek általános egészségesebb életmódjának köszönhetően. A c paraméterek esetén megfigyelhető az 1910-es években egy nagyobb csökkenés, ami az első világháborúra utalhat. Van egy kiugróan alacsony érték is, ami meglepő lehet, mivel egyrészt női adatokat tekintettünk, másrészt Svédországnak sikerült katonailag semlegesnek maradnia a háború jelentős részében.

A szakdolgozat elején említettük, hogy az angolszász országokban a 2010-es években a halandóság javulása megállt és kis mértékben romlani kezdett. A modellünk becsült c paramétereit alapul véve, a svéd nők esetén az mondható el, hogy a halandóság javulásának üteme itt is lassulni kezd.

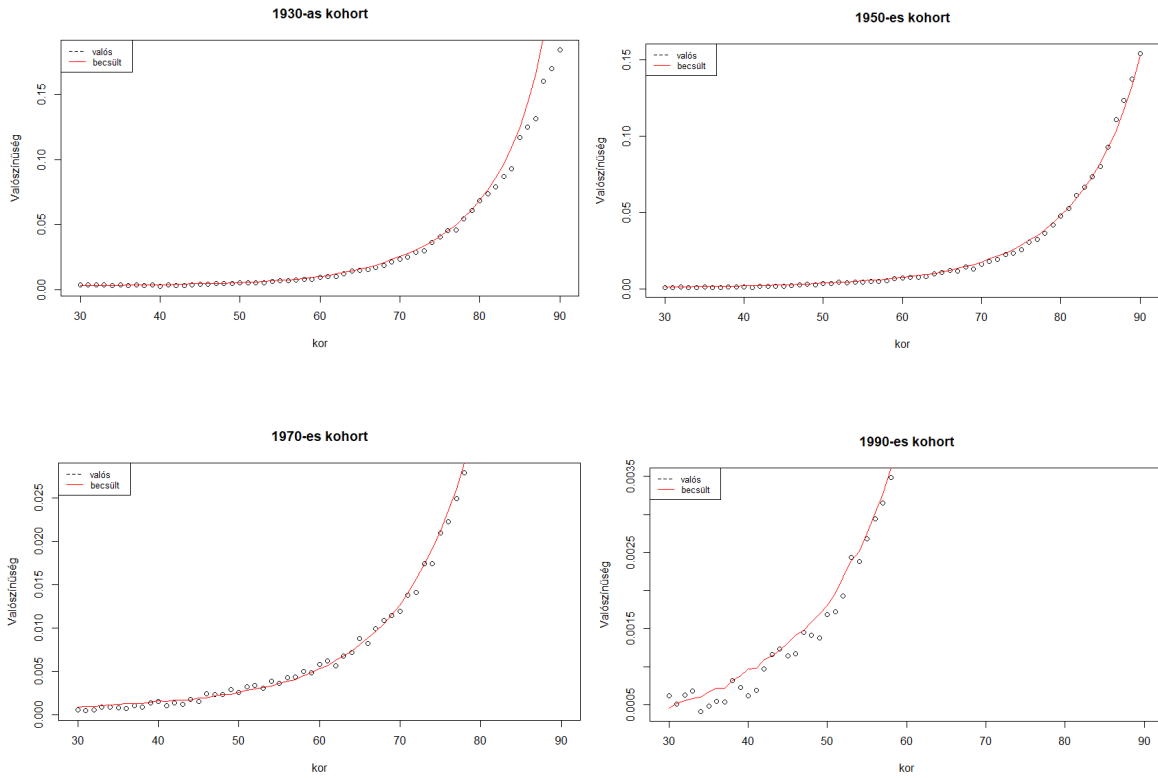
A kohortokra jellemző l paraméterek esetén 1900-1980-ig egy trendszerű csökkenést figyelhetünk meg, de kiugró értéket nem látunk.

A következő ábrákon négy kohortra ábrázoljuk, hogy a becsült életkoreloszlása a kohortoknak mennyire illeszkedik jól a valós adatokra.

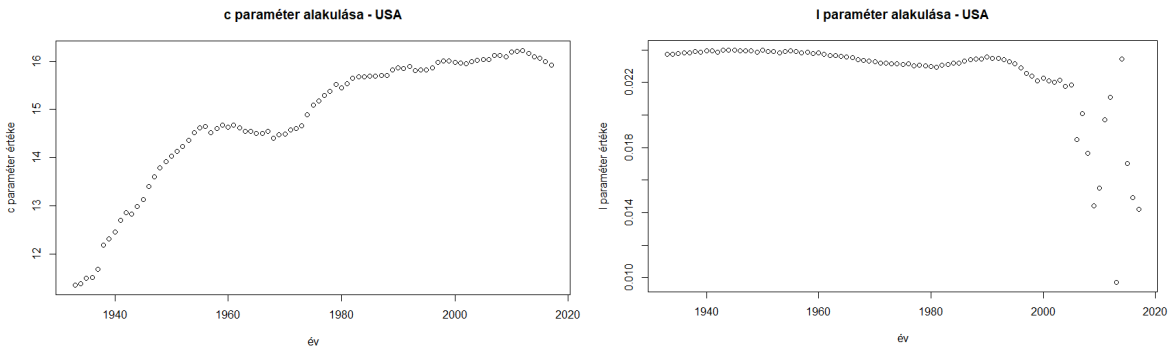


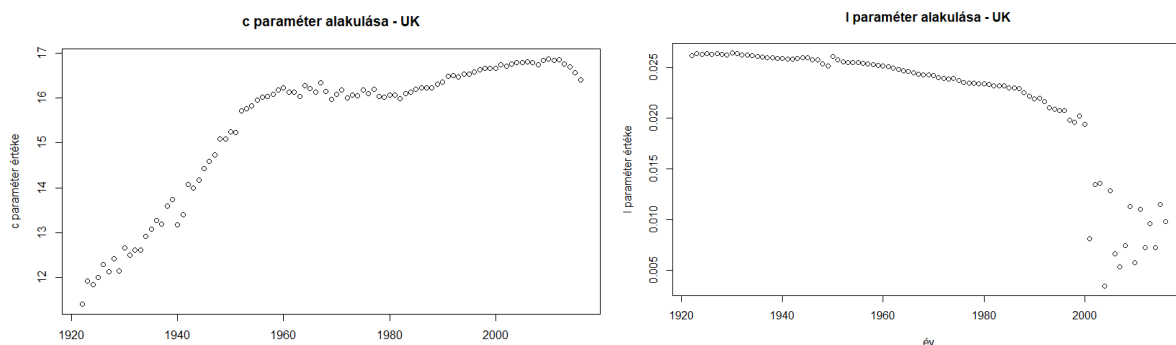
Ahol egy idő után 0 lesz a becsült ráta, ott az annyit jelent, hogy az az életkor még nem fordul elő a kohortban.

Egy másik ellenőrzés az illeszkedés minőségére, ha összehasonlítjuk egy adott évben minden életkorra az egy éven belüli halálozási valószínűségeket (q_x -eknek megfelelő ráták) a becsült és valós esetben. Ezek szemléltetésére készültek az alábbi ábrák:



A svéd adatokon kívül elvégeztük az illesztést amerikai (1933-2017 évekre áll rendelkezésre) és brit (1922-2016) női adatokra is. A célunk a lenne, hogy kimutassuk a 2010-es években jelentkező enyhe romlást a halandóságban a két ország esetén. A módszer ugyanaz, mint a svéd adatok esetén, a paraméterek alakulását a következő ábrákon szemléltetjük:





Az l paraméterek utolsó kohort évekhez tartozó értékeinél a svéd esethez hasonlóan lényegesen nagyobb a szórás, aminek az oka a kevés megfigyelés lehet, ami alapján végeztük a becslést.

A c paraméterek értékeiben minkét ország esetében összességében emelkedés figyelhető meg. Az utolsó évtized folyamán azonban rendre csökken a c kiindulási szintje az egészségfüggvénynek. Az alacsonyabb kiindulási szint rosszabb halandóságot eredményez, tehát ilyen módon megjelenik a modellben az említett halandóság romlás a két országot illetően. Az Egyesült Királysághoz tartozó c paraméterek esetén látható, hogy az 1940-hez és 1941-hez tartozó értékek kiugróan alacsonyabbak, mint a környező években. Ezt azzal lehet magyarázni, hogy a második világháborúban Anglia területe ezekben az években volt a leginkább érintett. További észrevétel a c paramétereket illetően, hogy az 1950-es évek közepétől a 60-as évek végéig enyhén csökken vagy stagnál az értéke. Ennek okairól részletesebb elemzés a [12] forrásban található.

5.1. Előrejelzés

Az utóbbi évtizedekre a halandóság javulása jellemző a Föld országainak többségében. Az életbiztosításban és a nyugdíjbiztosításban is egyre jelentősebb a hosszú élet kockázata, azaz hogy az emberek egyre nagyobb valószínűséggel élnek kimagaslóan hosszú évekig.

A jelen részben a célunk az, hogy az előzőekben ismertetett modell segítségével előre jelezzük a halandóság alakulását. Példaként megint a svéd nők adatait tekintjük.

Az alapvető koncepció az, hogy a paraméterek időbeli alakulását jelezzük előre, majd azok segítségével számoljuk a különböző kohortok mortalitásának alakulását az évek során. A korábbiakhoz hasonlóan az egészségfüggvény a következő alakban írható fel:

$$H(t) = c - (lt)^b$$

ahol a b paraméterert továbbra is tekintjük állandónak az évek alatt, a c kiindulási szint egy adott naptári évre jellemző, az l paraméterek pedig egy-egy kohorthoz vannak rendelve.

Az előző részben maximum likelihood módszerrel becsültük a paramétereket, a c paraméterek az 1860-2018-as naptári évekhez, az l paraméterek az 1840-1998-as születési évekhez tartoztak. Ezen paraméterek becslése későbbi évekre úgy történjen, hogy tekintsünk rájuk, mint idősorok és illesszünk rájuk ARIMA modellt.

A c paraméterek esetén először ahelyett, hogy 1860-tól tekintettük volna az idősort, csak az elmúlt 100 évre illesztettünk modellt. Ennek oka egyrészt az, hogy a korábbi évekhez tartozó becslések relatíve kevés megfigyelés alapján történtek, másrészt kellően régen voltak ahhoz, hogy ne vegyük figyelembe őket 2015 utáni évek becsléséhez. Végeztünk becslést úgy is, hogy csak az utolsó 35 év adatai alapján jeleztünk előre. Az l paraméterek esetén pedig az utolsó, későbbi évekhez tartozó paraméterek becslése történt kevés megfigyelés alapján és eléggé ingadozó becsléseket kaptunk. Ezért az 1960-1998-as születési éveknek megfelelő l paramétereket nem tekintettük az ARIMA modell illesztésekor. Továbbá az előző alfejezetben azt láttuk, hogy körülbelül az 1880-as születési évek legelejétől folyamatosan csökken a paraméter értéke időben, előtte pedig jellemzően nőtt, ezért illesztettünk úgy is idősoros modellt, hogy az első 40 évet belevettük és úgy is, hogy kihagytuk.

Az ARIMA modellek és paramétereiknek meghatározása az R programcsomag segítségével, a *forecast* package *auto.arima* beépített függvényével történt: a modell kiválasztása az Akaiké információs kritérium minimalizálásával történt. A c paraméterek utolsó 100 évre vonatkozó értékei alapján kapott modell egy ARIMA(0,2,2) lett, azaz a második differenciákon egy kéttagú MA folyamat:

$$c_t^{**} = -1.142\epsilon_{t-2} + 0.181\epsilon_{t-1} + \epsilon_t$$

ahol c^{**} a második differenciáját jelenti a c paramétereknek, ϵ pedig egy $\sigma^2 = 0.0186$ szórásnégyzetű fehér zaj.

Ha csak az utolsó 35 év adataira illesztettünk, akkor az optimális modell az Akaiké információs kritérium alapján egy driftes ARIMA(0,1,0) modellnek adódott:

$$c_t = c_{t-1} + 0.051 + \epsilon_t$$

ahol ϵ_t szórásnégyzete $\sigma^2 = 0.0066$.

Az l paraméterek esetén, amikor 1960-ig minden születési évet figyelembe vettünk, akkor a következő ARIMA(0,2,3) modellt kaptuk:

$$l_t^{**} = -1.570\epsilon_{t-1} + 0.873\epsilon_{t-2} - 0.226\epsilon_{t-3} + \epsilon_t$$

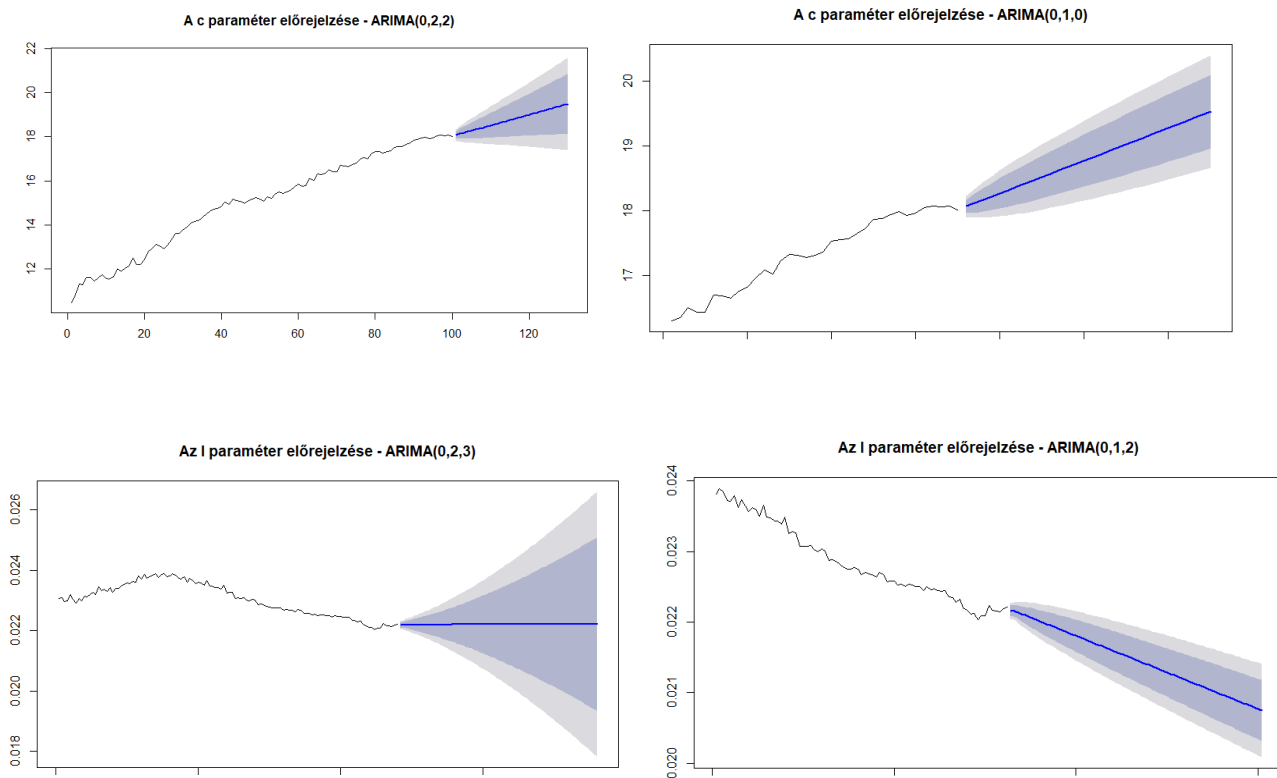
ahol l^{**} a második differenciája az l paramétereknek és ϵ_t szórásnégyzete $5.4 * 10^{-9}$. A második esetben, amikor az első éveket nem vettük figyelembe, csak mióta csökken a paraméter értéke,

egy ARIMA(0,1,2) modellt kaptunk:

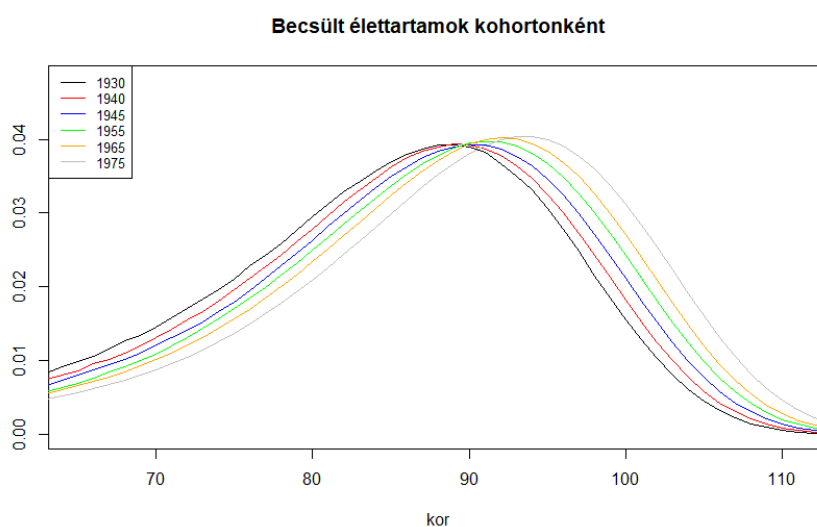
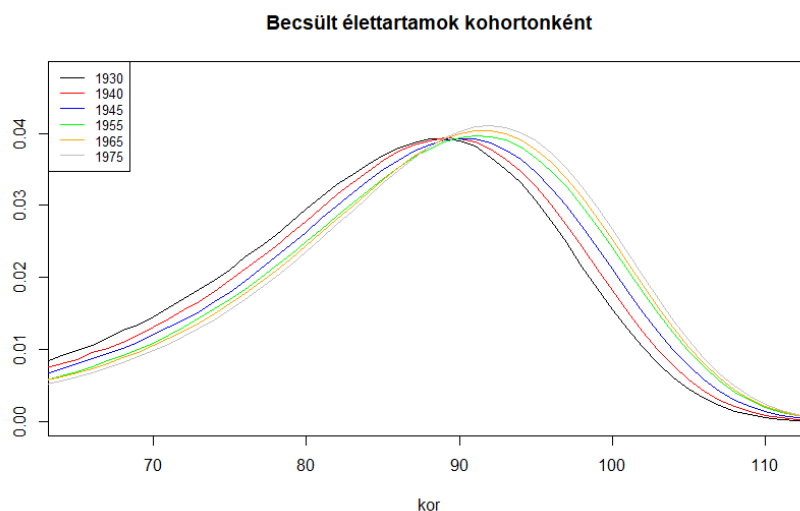
$$c_t^* = 0.22\epsilon_{t-2} - 0.616\epsilon_{t-1} + \epsilon_t$$

ahol ϵ_t szórásnégyzete $3.6 * 10^{-9}$.

A c és l paraméterek alakulásának becslése ezeknek az idősoroknak az előrejelzésével történik. Az előrejelzést, a 80% és 90% konfidencia intervallumokat a következő négy ábrán szemléltetjük:



A következőekben nézzük meg, hogy néhány kohort esetén ezek az előrejelzett paraméterek milyen "életkoreloszlást" eredményeznek. Az első esetben tekintsük azt, amikor a c és az l paraméterek becslése is a tágabb idősor alapján történik, a második esetben pedig amikor kevesebb becslést paraméterre illesztünk. Az eredményeket a következő ábrákon szemléltetjük (az első ábrán az első esethez tartozó, a másodikon a másodikhoz tartozó becslés szerepel):



Az idő előrehaladtával a kohortok mortalitása fokozatosan javul az ábra alapján, ugyanis a későbbi kohortok esetén egyre nagyobb valószínűséggel fordulnak elő idősebb korok. Ez a halandóság általános javulásának megfelelő eredmény, és erre is számítottunk, lévén hogy a c paraméterek folyamatos emelkedését jeleztük előre.

Az ábrákon csak a 65 évesnél idősebb személyekre ábráztuk a becslést, mert ezeknél a korosztályoknál szemretűnőbb volt az eltérés. Nyugdíjjáruélékok esetén például fontos mutató szám lehet a nyugdíjban töltött évek várható száma. A születéskori várható élettartam meghatározása a kohortra jellemző g élettartam "eloszlás" sűrűségfüggvényével felírva:

$$\int_0^{110} sg(s)ds \approx \sum_{s=0}^{110} g(s)s$$

A 65 év feletti várható élettartama pedig:

$$\int_{65}^{110} sg(s)/p_{65} ds \approx \sum_{s=65}^{110} g(s)s/p_{65}$$

ahol p_{65} jelöli azt a valószínűséget, hogy egy személy megéli a 65 éves kort.
Az első becslés alapján a várható élettartama egy 65 évet megélt svéd nőnek:

Születési év	1930	1940	1945	1955	1965	1975
Várható élettartam	86.5	87.3	88.1	88.9	89.2	89.7

A második becslés alapján:

Születési év	1930	1940	1945	1955	1965	1975
Várható élettartam	86.5	87.3	88.1	89.0	89.7	91.0

A becsült várható élettartamok a későbbi kohortok esetén nagyobb eltérést mutatnak a két becslés esetén.

Az előző esetekben a c paramétereket az elmúlt 100 és 35 év alapján jeleztük előre. Azonban az utolsó 10 évhez tartozó értéke a paramétereknek közel állandó volt, nem volt jellemző a korábbi fokozatos növekedés. Készítettünk úgy is egy előrejelzést, hogy a c paraméter értéke a 2018 utáni években konstans és az utolsó 10 megfigyelt évhez tartozó átlagos paraméterértéket veszi fel (az l paraméter pedig a hosszabb időtartam alapján előrejelzett értékeket vegye fel). Az így kapott várható élettartamot a 65 évet megélt svéd nők esetén a vizsgált kohortok esetén a következő táblázat foglalja össze:

Születési év	1930	1940	1945	1955	1965	1975
Várható élettartam	86.3	86.9	87.5	88.1	88.0	87.9

A várható élettartam tehát a későbbi kohortok esetén közel állandó a becslések alapján, sőt enyhén csökken.

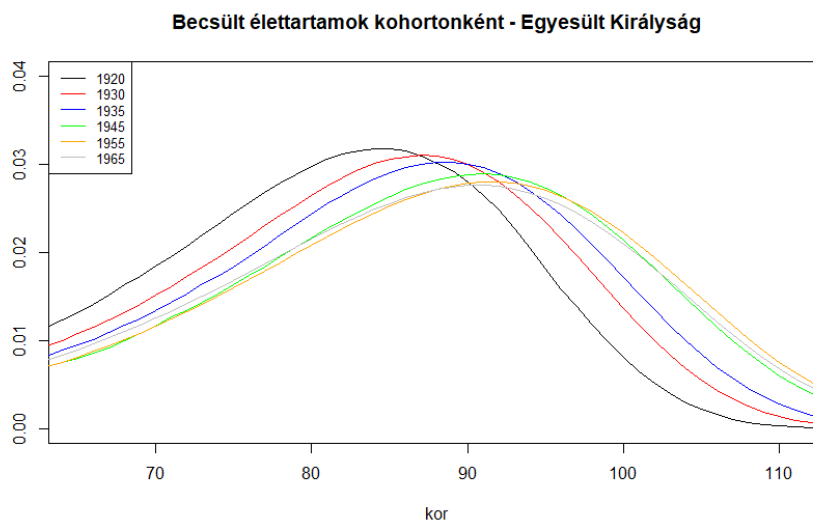
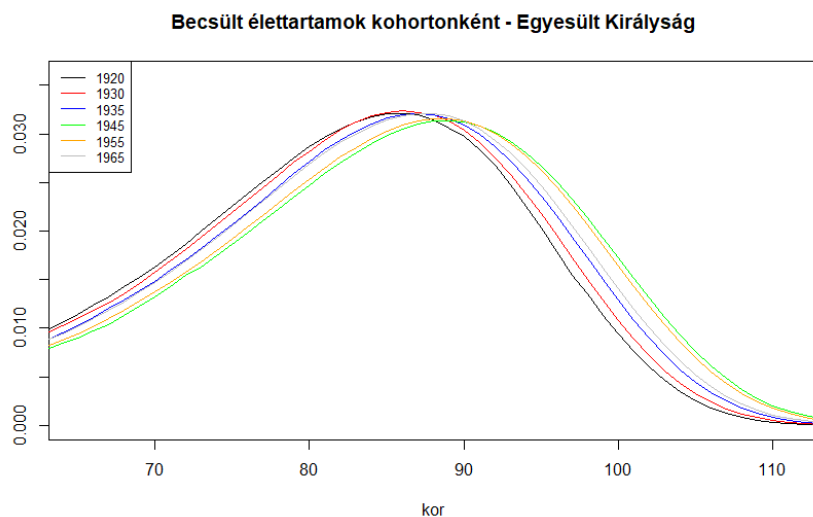
Összehasonlításképpen megnéztem a hivatalos svéd statisztikai hivatal (Statistiska centralbyran, SCB, [7]) által közzétett várható élettartamot 65 éves korban a vizsgált kohortok esetén:

Születési év	1930	1940	1945	1955	1965	1975
Várható élettartam	84.9	85.6	86.1	87	-	-

Ezek alapján a modelljeinkkel felül becsüljük a 65 éves kori várható élettartamot. Ennek oka lehet, hogy modelljeink esetén pozitív a valószínűsége annak is, hogy egy személy akár 115 éves korig éljen, míg más modellek esetén ez nem feltétlen van így és egy maximális élettartamot határoz meg 100 vagy 110 éves kornál.

Az előzőekhez hasonlóan amerikai és brit nők esetén is készítettünk halandósági előrejelzést. Ezekben az esetekben lényegesen kevesebb adat állt rendelkezésre, mint a svédek esetén (USA: 1933-2017, UK: 1922-2016-ig van elérhető adat), ezért az ARIMA modellek illesztésekor minden elérhető év adatait felhasználtuk.

A következő ábrákon néhány vizsgált kohort becsült élettartam eloszlásait ábrázoltuk a két ország esetén:



Mindkét ország esetén az látszik az ábrákon, hogy a legutolsó vizsgált kohortok esetén az eloszlás egy kicsivel kisebb valószínűségekkel vesz fel nagy értékeket és balra tolódott. Kiszámoltuk a várható élettartamát a 65 év felettieknek a vizsgált kohortokra, amelyet a következő táblázatban foglalunk össze. Ezek alapján szintén romlás várható a későbbi kohortok halandóságában.

Születési év	1920	1930	1935	1945	1955	1965
Várható élettartam (UK)	83.8	86.0	87.4	89.3	89.5	89.2
Várható élettartam (USA)	84.6	85.1	85.8	87.3	87.0	86.1

6. fejezet

Összefoglalás

A szakdolgozat célja egy adott populáció halandóságának dinamikus modellezése volt. Bemutatásra került egy Skiadas és Janssen nevéhez köthető sztochasztikus halandósági modell. A modell alapvető szemlélete az, hogy az egyének egészségügyi állapotának alakulását és halandóságát kapcsolja össze. Bevezetésre került az "egészségállapot függvény", amely egy lakosság átlagos egészségügyi helyzetét hivatott jellemezni különböző korcsoportokra.

Megmutattuk egy lehetséges átalakítását a modellnek, amelynek segítségével a halandóság alakulását tudjuk modellezni több egymást követő naptári év folyamán, figyelembe véve az egyes kohortokra jellemző hatásokat.

Több ország valós mortalitási adataira illesztettük a modelljeinket. Először brit nők és férfiak élettartam eloszlásait hasonlítottuk össze 30 éves kor felett (2000-es év adatai alapján). A pontos illeszkedés mellett visszakaptuk azt a tényt, hogy a női lakosság halandósága (és egészségügyi állapota) lényegesen jobb a férfi lakosságénál.

Svédország, az Egyesült Államok és az Egyesült Királyság női lakosságának esetén modelleztük a halandóság alakulását. A Human Mortality Database adatbázisból több évtizedre visszamenőleg érhetőek el ezekre az országokra mortalitási adatok (a svédek esetén egészen a 18. század második feléig), amelyek alapján hosszabb távú előrejelzést is adtunk. Az előrejelzés ARIMA idősoros modellekkel történt.

Az angolszász országok esetében az utóbbi évtizedben jelentkező halandóságromlást sikerült modellezni az egészségfüggvény becsült paraméterei révén.

A bemutatott modellek felépítése nem akadályozza, hogy ne a lakosság teljes egészére, hanem egy szubpopulációra végezzünk halandóság becslését. Megfelelő adatok rendelkezésre állása esetén például az országon belüli régiónkénti halandóság összehasonlítására is alkalmazható a modell. Továbbá a mortalitás mellett akár más kockázat becslésére is használható a módszertan: valamilyen gyógyíthatatlan betegséggel rendelkező lakosság számának becslése szintén történhet hasonló sztochasztikus modellek segítségével.

A modell nem vesz figyelembe az adott lakosságra jellemző adatokból semmi egyebet, csak a halandósági táblák alapján végezzük a becslést. Az egészségállapot függvény paraméterezésekor célszerű lehet a vizsgált ország különböző gazdasági mutatóit vagy az egészségügyére jellemző információkat felhasználni (például GDP, kórházi ágyak száma).

Ha külső információk segítségével sikerül jól becsülni az "egészségállapotot" az egyén szintjén (például jövedelme, lakhelye alapján), akkor az segíthet olyan döntések meghozatalában, mint korai nyugdíjba menetel vagy járadékszolgáltatási kérdésekben.

Természetesen nagyon sok mindent nem végeztünk el a szakdolgozatban, de az mindenképpen túllépett volna az adott munka keretein. A teljesség igénye nélkül felsorolunk néhány témát amit a későbbiekben vizsgálni lehetne.

- A becslési eljárások összehasonlítása más nevezetes modellekkel;
- A paraméterek összefüggései az országra jellemző indikátorokkal;
- A javasolt előrejelzések összehasonlítása általánosan használt előrejelzési modellekkel (például Lee-Carter).

Munkánkban mi „csak” egy konkrét modellt vizsgáltunk meg és azt próbáltuk továbbfejleszteni. Skiadas és szerzőtársai munkáihoz képest legalább két területen sikerült új eredményt elérni. Egyrészt a paraméterek becslését maximum likelihood módszerrel is elvégeztük, másrészt a sokkal természetesebb kohorsz alapú modellt is felépítettük. Ez a későbbiekben felhasználható lenne előrejelzésekhez is (ennek kezdeti lépéseit be is mutattuk).

Irodalomjegyzék

- [1] J. Janssen, C. H. Skaidas, *Dynamic modelling of life table data*, Applied Stochastic Models and Data Analysis,11, 1995.
- [2] C.H. Skiadas, C. Skiadas, *Exploring the health state of a population by dynamic modeling methods*, The Springer Series on Demographic Methods and Population Analysis, 2018.
- [3] C.H. Skiadas, *A modeling approach to life table datasets*, Recent Advances in Stochastic Modeling and Data Analysis, 2007.
- [4] C.H. Skiadas, C. Skiadas, *Exploring the state of a stochastic system via stochastic Simulations: An interesting inversion problem and the health state function*, Springer Science, 2014.
- [5] N. G. van Kampen, *Stochastic processes in physics and chemistry*, North-Holland Personal Library, 2007, 73-95
- [6] C. W. Gardiner, *Handbook of Stochastic Methods*, Springer, 2009.
- [7] *Svéd Statisztikai Hivatal adatai*, 2020, <https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/yearly-statistics-the-whole-country/life-expectancy/>
- [8] *Human Mortality Database adatai*, 2020, www.mortality.org
- [9] C. J. Geyer, *Markov Chain Monte Carlo Maximum likelihood*, School of Statistics, 1991.
- [10] L. Carroll, *U.S. life expectancy declining due to more deaths in middle age*, Reuters, 2019.
- [11] N. Ouellette, M. Barbieri, J. R. Wilmoth, *Mortality and life expectancy trends in the UK: stalling progress*, The Health Foundation, 2019.
- [12] L. Marshall, D. Finch, L. Cairncross, J. Bibby, *Period-Based Mortality Change: Turning Points in Trends since 1950*, Population and Development Review, 40, 2015.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Arató Miklósnak, hogy ismertette velem a témát, rengeteg hasznos tanáccsal és észrevétellel látott el, valamint bármikor alapos magyarázatot adott kérdéseimre. Sokat tanultam az utóbbi évek konzultációiból.

Szeretném továbbá megköszönni Hegel Istvánnak és Nagy Fanninak a szakdolgozat elkészítésével kapcsolatos segítségüket.