SZAKDOLGOZAT

Kössünk biztosítást!

Czimer Máté

Eötvös Loránd Tudományegyetem

Természettudományi Kar

Matematika Bsc

Témavezető: Zempléni András egyetemi docens

Valószínűségelméleti és Statisztika Tanszék

Budapest, 2019.

Tartalomjegyzék

1. Bevezetés **1**

2. Elméleti háttér **4**

2.1. Díjkalkulációs elvek 4

2.1.1. Alapfogalmak 4

2.1.2. A várható érték elve 5

2.1.3. A maximális veszteség elve 7

2.1.4. A kvantilis elv 7

2.1.5. A szórásnégyzet elv 8

2.1.6. A szórás elv 8

2.1.7. A féloldali szórás elv 8

2.2. Megoldás ötlet: Markov-láncok 10

2.2.1. A cél meghatározása 10

2.2.2. Markov-láncok ismertetése 10

3. A matematikus túlcsúszás-biztosítás kidolgozása **13**

3.1. Adatfeldolgozás 14

3.1.1. Források 14

3.1.2. Tisztítás, konverzió 14

3.2. Előkalkulációk 16

3.2.1. Tapasztalati eloszlás 17

3.2.2. Besorolás, érettségi pontszámok normálása 18

3.2.3. Első féléves eredmények 20

3.2.4. Átmenetvalószínűségek simítása, váratlan értékek simítása 22

3.2.5. Képzésből kilépő hallgatók esete 24

3.3. Szimuláció 26

3.3.1. Adatgenerálás kezdeti állapottal 27

3.3.2. A Markov-folyamat állapotainak létrehozása 29

3.3.3. A végbizonyítványt szerzők kiléptetése 30

3.3.4. Ellenőrzések 31

3.3.5. Díjak meghatározása 33

3.3.6. Cashflow 35

4. Konklúziók **38**

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Zempléni András Tanár úrnak a hihetetlenül sok segítségért, jó tanácsért és bizalomért, amivel támogatott a szakdolgozatom elkészítésében.

Szeretném továbbá megköszönni Édesanyámnak, Húgomnak és Páromnak, hogy mindvégig támogattak és bíztattak engem. Nélkülük nem sikerült volna.

1. Bevezetés

Szakdolgozatom témája egy fiktív biztosítási modell bemutatása, amely az Eötvös Loránd Tudományegyetem Matematika Bsc szakon tanuló hallgatóknak nyújtana támogatást. A biztosítás célja, hogy az állami képzésből – valamilyen oknál fogva – átsorolt diákok támogatást kapjanak a biztosítótól az önköltséges féléveik finanszírozására.

A modell felépítéséhez, a kapott eredmények értelmezéséhez szükséges némi biztosítási ismeret. A számítások és szimulációk elkészítéséhez elengedhetetlen a meglévő adatok elemzése, továbbá fontos részét képezi a feladatnak az elméleti háttér megértése, és alkalmazása. A konklúziók levonásához pedig szükség lesz az eredmények vizuális megjelenítésére, amelyek szemléltetik a modell egyes ismérveit.

Ezen lépéseket szeretném bemutatni a dolgozatomban részletekbe menően. Az adatfeldolgozás, a gyakorlati számítások és vizualizációk az *RStudio* szoftverrel készültek *R* programnyelven.

A fejezetek egymásra épülve vezetik végig az olvasót a modell konstruálásában. Elsőként arra a kérdésre térek ki, hogy miért is lehetne fontos a diákok ilyen módon vett biztosítása a tanulmányaik során, illetve igyekszem megfoghatóvá tenni a biztosítás fogalmát. Ezek után az elméleti háttér ismertetése következik, mely a biztosítási díjak számításához lényeges matematikai tételeket, állításokat, bizonyításokat és definíciókat tartalmazza. Ezt követően bemutatásra kerül a számításokhoz használt adathalmaz. Természetesen elengedhetetlen a bemutatott elvek gyakorlatba ültetése, illetve a válaszokhoz vezető elemzések, vizualizációk prezentálása. Mindezek eredményeképp a felépített modell szimulációja következik, amely igyekszik előrejelzésekkel választ biztosítani a felmerülő kérdések minél bővebb részhalmazára. Az utolsó fejezetben a konklúziók levonásával kiderül, hogy várhatóan egy sikeres modellt tudtunk-e kidolgozni, avagy sem.

Miért kössünk biztosítást?

Mi a biztosítás? Véletlenszerűen bekövetkező, előre nem látható események anyagi kárai elleni védelem. És miért kötnek az emberek biztosítást? Mert bizonyos – számukra negatív – esemény bekövetkezése esetén pénzre lesz szükségük, melyet önerőből nem, vagy csak nagy nehézségek árán tudnak kifizetni. Tehát védelmet, biztonságot várnak a biztosításuktól. És mi a célja a biztosító társaságoknak? Természetesen a profit elérése.

Azt gondolnánk, hogy a két fél – a biztosító és a biztosítást igénybe vevő – két ellentétes oldalt képvisel, de végeredményben mindkettőjük „nyer az üzleten", hiszen a biztosított biztonságot, a biztosító társaság pedig profitot realizál. A mai modern biztosítás nem más, mint a kockázat felosztása pénzalap képzésével a kockázatközösség (biztosított) tagok jövőbeni, esetleges és felmérhető szükségleteinek kielégítése céljából.

A családok számára nagy terhet jelent gyermekük, gyermekeik felsőfokú tanulmányainak támogatása még akkor is, ha a diák államilag finanszírozott formában kerül be az intézménybe. A hallgatók nagy többsége lakóhelyétől távol, az ország egy más pontján végzi a főiskolát, egyetemet, amiből következik, hogy a családnak a gyermek megélhetését, lakhatását is támogatnia kell. Könnyen előfordulhat, hogy a diák valamely oknál fogva kiesik az állami támogatói rendszeréből, amikortól a tandíj is nagyobb kiadást jelent.

Dolgozatomban erre az esetre vonatkozóan próbálok felállítani egy biztosítási modellt amelynek az a célja, hogy anyagiak hiányában ne vesszen kárba a diák több éves felsőfokú oktatása, és ne kelljen diploma nélkül félbehagynia tanulmányait. Az általam kidolgozott biztosítás esetében a hallgató tanulmányai megkezdésekor köthet biztosítást a biztosító társasággal. A biztosítás – a támogatott rendszerből történő kiesés miatt – a fizetendő tandíjra vonatkozik.

A díjakat a biztosított havi részletekben fizeti, a részletek pedig az első félévre megállapított díjrészletekhez képest a féléves eredmények függvényében módosulnak. A biztosító társaságok profit fejében szerződnek, nyereségük pedig a bevételükkel szembeállított kiadásaik különbözete. Ennek részletesebb elemzése nem képezi jelen dolgozat témáját, ezért szeretnék néhány alapvetést lefektetni, melyeket a számításaim során figyelembe veszek:

- a biztosító társaság 20%-os profittal dolgozik,

- a biztosító társaság káreseményen felüli kiadásaival nem számolok,
- a biztosítási szerződésben foglalt díj a következő aktív tanulmányi félévre szól, és minden szemeszter végén teljesítmény alapon felülvizsgálásra kerül,
- figyelmen kívül hagyjuk a befizetett biztosítási díjak banki kamatozását,
- nem számolunk azokkal az esetekkel, amikor a biztosított biztosítási díj nem fizetése miatt kiesik a rendszerből,
- korlátozásra kerül a maximálisan fizethető károsszeg. Mivel az önköltség díjában bekövetkező változások felmenő rendszerben történnek, így emelés esetén az újonnan felvett hallgatóknál egyszerűen indexáljuk a díjakat. A jelenlegi önköltség értéke a szakot a 2019/20-as őszi félévben megkezdők esetén 500 000 Ft félévente a TTK Tanulmányi osztály adatai alapján [8]. Ennek legfeljebb kétszeresét, azaz 1 000 000 Ft-ot fizet a biztosító társaság károsszegként. Ezen felül nem vállal több kötelezettséget még abban az esetben sem, ha a hallgató költségtérítési összege meghaladja ezt az értéket a tanulmányi évei során. Tehát a biztosító az aktuális tanévre vonatkozó önköltség értékének kétszeresét szabja ki felső korlátnak.

Ezen alapvetések tudatában történik a modell felépítése, és a biztosítási díjak meghatározásához felhasznált elméleti háttér gyakorlatba ültetése is.

2. Elméleti háttér

Az alábbi fejezetben tárgyaljuk a modell felépítéséhez lényeges elméleti fogalmakat, tételeket, definíciókat. Az itt megfogalmazott elvek a [1] és [2] könyvek alapján kerülnek bemutatásra. Ezen ismereteket szeretnénk a biztosítási feladat példáján keresztül a gyakorlatban alkalmazni.

2.1. Díjkalkulációs elvek

2.1.1. Alapfogalmak

A biztosító társaságok feladata, hogy minél pontosabb előrejelzést készítsének a biztosítottakról, amelyek alapján biztosítási díjat határozhatnak meg. Természetesen ez nem egyszerű. Az esetek nagy többségében kevés információ áll rendelkezésre, így nehéz meghatározni egy olyan összeget, amely fedezi az esetleges kár költségét, a biztosított megfizethetőnek tekinti, és még hasznot is hoz a biztosítónak. Ezt a pénzösszeget biztosítási díjnak, a meghatározására szolgáló módszert pedig biztosítási díjkalkulációnak nevezzük.

A biztosítási díjért cserébe a biztosító kockázatot vállal, hogy káresemény bekövetkezése esetén fizet. Ezt egy (nemnegatív) valószínűségi változóval írjuk le. A díjkalkulációban valamilyen szabály alapján egyértelműen rendelünk értéket a kockázat eloszlásához, ezért az eloszlás ismerete elengedhetetlen.

A továbbiakban térjünk át a matematika nyelvére, hogy a megfelelő definíciók bevezetésével pontosabb képet kapjunk a díjkalkulációs elvekről és tulajdonságaikról.

2.1. Definíció. Legyen $\mathcal{H}$ azon eloszlások halmaza, melyek a nemnegatív félegyenesre koncentrálódnak. Legyen továbbá $X : \Omega \to \Re_0^+$ valószínűségi változó a kockázat, melynek eloszlásfüggvényét jelölje $F_{\mathrm{x}} \in \mathcal{H}$. A Π függvényt *díjkalkulációs elvnek* vagy *díjelvnek* nevezzük, ha

$$\Pi : \mathcal{H}_\Pi \subseteq \mathcal{H} \longrightarrow \Re_0^+ \cup \{\infty\}$$

leképezés, tehát az eloszláshoz hozzárendelt érték az eloszlás díja.

Az elveket egyaránt értelmezhetjük valószínűségi változóra, és eloszlásfüggvényre is. Amennyiben előbbit számítjuk, akkor indexként egy v betűt használunk a *variable*, azaz változó

jelölésére, míg eloszlásfüggvény esetén F jelzi a függvényre vett értelmezést. A kockázatra, mint valószínűségi változóra azonban nem tudunk visszakövetkeztetni egyértelműen az eloszlásfüggvényéből, ugyanis több olyan valószínűségi változó is létezhet, melyeknek megegyezik az eloszlásfüggvényük. Fordított esetben azonban a valószínűségi változó meghatározza eloszlásfüggvényét.

A díj a nemnegatív valós számok halmazán veszi fel értékeit. Végtelenként értelmezzük abban az esetben, ha a kockázat nem biztosítható az adott díjelvre nézve.

2.2. Definíció. Ha a $\Pi(X)$ díjelv, X tetszőleges (nemnegatív) valószínűségi változó és F nemnegatív félegyenesen értelmezett eloszlásfüggvény, akkor

$$\Pi_v(X) := \Pi(\mathcal{Q}_X), \text{ és } \Pi_F(F) := \Pi(\mathcal{Q}_F),$$

ahol $\mathcal{Q}_{\mathrm{X}}$ az X valószínűségi változó eloszlása, $\mathcal{Q}_{\mathrm{F}}$ pedig az F eloszlásfüggvénnyel meghatározott eloszlás.

Tehát azt mondhatjuk, hogy a kockázat megadható valószínűségi változóként, és eloszlásfüggvénnyel egyaránt. A díjakat az eloszlás alapján számítjuk, mely mindkét alakból meghatározható.

A továbbiakban vezessük be az $\Pi(X) := \Pi(F_X)$ jelölést, ahol X kockázat, mint nemnegatív valószínűségi változó, F_X pedig X eloszlásfüggvénye.

A fenti definíciók segítségével már értelmezhetőek a különböző díjelvek, amelyeket részletesebben szeretnék a továbbiakban bemutatni. Az ismertetni kívánt díjelvek mindegyike a klasszikus díjelvek csoportjába tartozik. Előfordulnak még elméleti díjelvek (pl. svájci díjkalkulációs elv, zéró hasznossági elv). Ezeket azonban a gyakorlatban kevésbé használják, így a dolgozatban felépülő biztosítási modell esetében sem alkalmazom. Lássuk tehát, milyen díjelveket értünk klasszikus díjelveken.

2.1.2. A várható érték elve

A gyakorlatban talán legtöbbet használt elv a várható érték elve. Amennyiben megfelelő számításokkal és információkkal rendelkezünk a kockázatról, akkor könnyen alkalmazhatjuk ezt a díjkalkulációt.

2.3. Definíció. A Π $\lambda \geq 0$ paraméterű díjelv várható érték elv, ha

$$\Pi_v(X) = (1+\lambda)EX$$

minden nemnegatív X valószínűségi változóra.

A $\lambda = 0$ paraméterérték megválasztásával a várható érték elv egy speciális tagját, a nettó várható érték elvek kapjuk eredményül. Definiáljuk továbbá, hogy mely tulajdonságok fennállása esetén nevezhetünk egy Π díjelvet várható érték elvnek. Ehhez először szükségünk lesz a folytonos díjelv fogalmának bevezetésére.

2.4. Definíció. A Π díjelv folytonos, ha

$$\left(\mathcal{Q}_X, \mathcal{Q}_{X_n} \in \mathcal{H}_\Pi \text{ és } E|X_n - X| \xrightarrow[n\to\infty]{} 0\right) \Longrightarrow \left(\Pi(\mathcal{Q}_n) \xrightarrow[n\to\infty]{} \Pi(\mathcal{Q})\right). \tag{1}$$

Az (1) definíció mellett a következő jelölést szükséges még bevezetnünk:

- $\delta_C := C$ konstans eloszlása.

2.1. Állítás. *A Π díjkalkulációs elv pontosan akkor várható érték elv, ha*

a) folytonos,

b) $\Pi(\delta_C) = (1+\lambda)C$ minden $C \in \Re_+$ esetén,

c) $\Pi(t\mathcal{Q} + (1-t)S) = t\Pi(\mathcal{Q}) + (1-t)\Pi(S)$ $\forall t \in [0;1]$ és $\mathcal{Q}$, $S \in \mathcal{H}$ esetén.

Bizonyítás. A várható érték elvre triviális mind a három tulajdonság, tehát csak a másik irányt kell belátnunk. Ehhez vegyünk egy X nemnegatív valószínűségi változót, melyre $EX < \infty$. Léteznek olyan $0 \leq X_n \leq X$ változók, melyek véges sok értéket vesznek fel, és $E|X_n - X| \xrightarrow[n\to\infty]{} 0$. Mivel *b)* és *c)* alapján $\Pi_v(X_n) = (1+\lambda)EX_n$, illetve *a)* szerint $\Pi_v(X_n) \xrightarrow[n\to\infty]{} \Pi_v(X)$, ezért $\Pi_v(X_n) = (1+\lambda)EX$.

Ha $EX < \infty$ nem áll fenn, tehát nem véges várható értékű a változónk, akkor *c)* és az előzőek alapján

$$\Pi_v(X) = \Pi_v(\chi_{(X\leq n)}X + \chi_{(X>n)}X) = P(X \leq n)\Pi(\mathcal{Q}_{X|X\leq n}) + P(X < n)\Pi(\mathcal{Q}_{X|X>n}) \geq$$
$$P(X \leq n)\Pi(\mathcal{Q}_{X|X\leq n}) = P(X \leq n)(1+\lambda)E(X|X \leq n) = (1+\lambda)E(\chi_{(X\leq n)}X).$$

Ekkor az utolsó tag, vagyis $(1+\lambda)E(X)$ végtelenhez tart a monoton konvergencia tétel szerint, így bizonyítottuk a nem véges várható értékű esetet is. $\square$

2.1.3. A maximális veszteség elve

A várható érték elv nem alkalmas az esetlegesen fellépő nagy károk által okozott pénzügyi kockázatok kezelésére. Ennek megoldására bevezetünk egy p paramétert, mely segítségével külön is súlyozzuk a legnagyobb kárt.

2.5. Definíció. A $\Pi(X)$ biztosítási díjelvet maximális veszteség elvének nevezzük, ha

$$\Pi_v(X) = pEX + (1-p)\sup\{x : P(X > x) > 0\}, \text{ ahol } p \in (0,1),$$

és X nemnegatív valószínűségi változó.

A maximális veszteség elve nem korlátos károk esetén végtelent ad, tehát nem biztosíthatónak tekinti az ügyletet.

2.1.4. A kvantilis elv

A kvantilis elvet abban az esetben szokás használni, mikor a kockázatunk nem korlátos. Ekkor a várható érték elv alkalmazható, azonban azt láttuk, hogy ilyen esetben nem biztonságos. A kvantilis elv azonban kezeli az ilyen típusú kockázatokat. Az előző maximális veszteség elvhez hasonlóan itt is a veszélyes, nagy károkat büntetjük.

2.6. Definíció. A $\Pi(X)$ biztosítási díjelvet kvantilis elvnek nevezzük $(1-\epsilon, p)$ paraméterekkel, ha

$$\Pi_v(X) = pEX + (1-p)\inf\{x : P(X < x) \geq 1-\epsilon\}, \text{ ahol } p, \epsilon \in (0,1),$$

X nemnegatív valószínűségi változó, ϵ a kvantilis meghaladásának valószínűségét jelöli.

2.2. Állítás. *Ha $\epsilon > 0$ megfelelően kicsi, akkor X az $1-\epsilon$ kvantilisnél kis valószínűséggel vesz fel nagyobb értéket. $\epsilon = 0$ esetén pedig*

$$1 - 0 = \inf(x : F(x) = 1) = \sup\{x : P(X > x) > 0\}$$

Ekkor tehát a maximális várható érték elvet kapjuk.

2.1.5. A szórásnégyzet elv

A eddig bemutatott díjelvek képletében nem került felhasználásra a kockázat szórása, azonban a következő három díjkalkuláció ezen statisztikai ismérv alkalmazásával kapható meg. Ennek legfőbb jelentősége, hogy egy könnyebben becsülhető paraméter révén bizonyos biztonsági költségpótlékot termeljen a biztosító számára, amivel finanszírozni tudja azokat az időszakokat, mikor a károk összege a várható érték elvvel meghatározott biztosítási díjak összértéke fölé csúszik.

2.7. Definíció. A $\Pi(X)$ biztosítási díjelv a szórásnégyzet elv, ha

$$\Pi_v(X) = EX + \beta D^2 X, \text{ ahol } \beta \geq 0,$$

X nemnegatív valószínűségi változó, és a $\beta D^2 X$ tag az úgy nevezett biztonsági költségpótlék.

A szórásnégyzet elve hasznos lehet a biztosító társaság számára, mikor a kitűzött célok között prioritást élvez a biztonsági tartalék halmozása. Amennyiben $D^2 X = \infty$, a kiszabott díj is végtelen.

2.1.6. A szórás elv

A (2.1.5)-höz hasonlóan a szórás elv is egy bizonyos paraméterrel beszorzott szórás tagot alkalmaz a képletben. Itt azonban a szórásnégyzet gyökét számítjuk, és ennek felhasználásával kapjuk a biztonsági költségpótlékot.

2.8. Definíció. A $\Pi(X)$ biztosítási díjelvet szórás elvnek nevezzük, ha

$$\Pi_v(X) = EX + \alpha DX, \text{ ahol } \alpha \geq 0,$$

X nemnegatív valószínűségi változó, továbbá az αDX tag értéke a biztonsági költségpótlék.

2.1.7. A féloldali szórás elv

Bár ez a díjelv is a szórást használja a kalkuláció során, mégis kicsit eltérően funkcionál a korábbi esetekhez képest. A (2.1.5) és (2.1.6)-ban említett díjelvek számításánál figyelbe

vettük azt az esetet is, mikor a kockázatunk szórása nagy és ebben jelentős súllyal szerepelnek olyan esetek, amikor ez az eltérés negatív irányú. A féloldali szórás elve pontosan erre a problémára koncentrál. A díjelv bevezetéséhez azonban először definiáljuk a pozitív oldali szórás fogalmát.

2.9. Definíció. A pozitív féloldali szórásnégyzet a

$$D_+^2 X = E\left(|X - EX|^+\right)^2$$

képlet által meghatározott értéke az X nemnegatív valószínűségi változónak.

2.10. Definíció. A $\Pi(X)$ biztosítási díjelv a féloldali szórásnégyzet elv, ha

$$\Pi_v(X) = EX + \beta D_+^2 X, \text{ ahol } \beta \geq 0,$$

X nemnegatív valószínűségi változó, továbbá az $\beta D_+^2 X$ a biztosítás biztonsági költségpótléka.

A fejezet során definiált díjkalkulációs elvek közül a várható érték elve a leggyakrabban használt módszer, így ennek gyakorlatba ültetését igyekszünk a szimuláció során létrehozott adatokon bemutatni, illetve elemezzük a biztosító társaság pénzügyi helyzetének alakulását, előrejelzését a díjelv alkalmazása esetén.

2.2. Megoldás ötlet: Markov-láncok

2.2.1. A cél meghatározása

A képzeletbeli cég, nevezzük *StudIn Biztosítónak* feladata, hogy kiszámítsa a káresemények bekövetkezési valószínűségét. Ehhez elengedhetetlen, hogy nyomon követhető legyen a hallgatók tanulmányi előrehaladása. Bizonyos időközönként felül kell vizsgálni a biztosítottak teljesítményét, hiszen ennek függvényében változhat a kockázat, és a biztosítási díj is egyben.

A cél tehát az, hogy teljesítmény alapon be sorolja a biztosító a hallgatókat osztályokba, majd adott idő elteltével rendszeresen vizsgálnia kell előre meghatározott kritériumokat, mely alapján átsorolást végezhet.

Az átsorolásnál a hallgató kerülhet felsőbb és alsóbb osztályba is egyaránt (ha van ilyen osztály), vagy maradhat ugyanazon kategóriában, melybe korábban besorolták.

A cél tehát az, hogy egy olyan matematikai elméletet ültessünk át a gyakorlati példába, amelyben az adatpontjainkat csoportokba oszthatjuk, majd bizonyos jellemzők alapján ezeken a csoportokon belül mozoghatnak. Fontos szempont az is, hogy mivel a folyamat időben előrehalad, valahogy modellezhető legyen a kilépés fogalma is, mégpedig többféle végkimenetellel.

2.2.2. Markov-láncok ismertetése

Az ilyen típusú problémákat Markov-folyamattal (vagy Markov-lánccal) írhatjuk le. A Markov-folyamat fogalma Andrej Andrejevics Markov (1856–1922) orosz matematikustól származik.

A Markov-folyamatokat a Markov tulajdonsággal definiálhatjuk, azonban ehhez szükséges a sztochasztikus folyamat, és a diszkrét idejűség jelentésének tisztázása. Az itt megfogalmazott definíciókat az [5] és [6] jegyzetek segítségével írjuk le.

2.11. Definíció. Sztochasztikus folyamat az $(X_t)_{t\in T}$, ahol $T \subset \Re$ paramétertér és $\forall t$-re X_t valószínűségi változó.

2.12. Definíció. A sztochasztikus folyamatot diszkrét idejűnek, vagy diszkrét paramété-

rűnek nevezzük, ha T paramétertér számossága véges, vagy megszámlálhatóan végtelen.

2.13. Definíció. Egy diszkrét idejű sztochasztikus folyamatot Markov-folyamatnak nevezünk, ha $\forall n \in \mathbb{Z}_0^+$ és $\forall t_0 < t_1 < \cdots < t_n$-re:

$$P\left(X\left(t_n\right) = k_n | X\left(t_{n-1}\right) = k_{n-1}, \ldots, X\left(t_0\right) = k_0\right) = P\left(X\left(t_n\right) = k_n | X\left(t_{n-1}\right) = k_{n-1}\right)$$

2.3. Állítás. *A Markov-tulajdonság (2.13) felírható a várható érték segítségével:*

$$E\left(f\left(X\left(t_n\right)\right) | X\left(t_{n-1}\right), \ldots, X\left(t_0\right)\right) = E\left(f\left(X\left(t_n\right)\right) | X\left(t_{n-1}\right)\right)$$

$\forall f$ *folytonos és korlátos függvény esetén.*

Tehát a (2.13) definíció alapján tudjuk, hogy az ilyen tulajdonságú folyamatoknál elég a jelenlegi állapottal számolni, mikor a jövővel kapcsolatban szeretnénk információt adni.

A modellünk szempontjából meg kell vizsgálnunk, hogy az egyes állapotok közötti átmenetvalószínűségek milyen tulajdonságúak. Ezért néhány jelölést és definíciót szükséges még bevezetnünk.

2.14. Definíció. Azt mondjuk, hogy a Markov-lánc homogén (vagy stacionárius), ha a $P(X(t_n) = l | X(t_{n-1}) = k)$ feltételes valószínűség független n-től, azaz $\forall n$ esetén megegyezik.

A fenti elméleti fogalmakat szeretnénk a továbbiakban felhasználni, alkalmazni. Ehhez szükségünk lesz néhány fontos jelölésre, melyek az alábboiak:

- S jelölje a homogén Markov-lánc állapotterét,
- $q^T = (q_1, q_2, \ldots)^T$, ahol $q_i = P(X_0 = k)$, és $k \in S$,
- az egylépéses átmenetvalószínűségek:

$$p_{k,l} := P(X_n = l | X_{n-1} = k),\ k, l \in S$$

- $P = (p_{k,l})$ az átmenetvalószínűség mátrix, ha $k, l \in S$,
- az n lépéses átmenetvalószínűségek:

$$p_{k,l}^{(n)} := P(X_n = l | X_0 = k),\ k, l \in S$$

- $P^{(n)} = (p_{k,l}^{(n)})$ az n lépéses átmenetvalószínűség mátrix, ha $k, l \in S$.

Ezekkel a jelölésekkel, és a korábban leírt definíciók segítségével nekiláthatunk a számítási feladatok konstruálásának és megoldásának.

A következő fejezetekben folyamatosan építünk a Markov-láncokkal kapcsolatos fogalmakra, gyakran anélkül, hogy ezt hangsúlyoznánk. Ezzel biztosítjuk az elméleti hátteret azokra a helyzetekre, mikor egy bizonyos problémára megoldást szeretnénk találni. Azonban a Markovitás feltételezése nem mindig helytálló, de úgy gondoljuk, hogy a mi esetünkben reális és egyszerű modellt ad.

3. A matematikus túlcsúszás-biztosítás kidolgozása

Képzeljük magunkat a *StudIn Biztosító* elemzői helyébe, és a korábbi elméleti háttérrel felfegyverkezve építsünk fel egy biztosítási modellt, amelyet frissen induló biztosító társaságunk szeretne felhasználni a jövőben. Ahhoz, hogy sikeres és hosszú életű karriert biztosítsunk a cégnek, minél pontosabb számításokra és alkalmazásokra lesz szükségünk.

Mielőtt elemznénk a rendelkezésre álló adatokat, fektessünk le néhány irányelvet a modell struktúráját illetően. A (2.2) alfejezetben megismert módszer szerint szeretnénk kezelni a diákok előrehaladását. Állapotokat vezetünk be a félévekre, melyekben a diákokat teljesítményük függvényében kategorizálunk. A folyamatban, akárcsak a gépjárműbiztosításoknál használt bonus-malus rendszernél [9], itt is osztályok között léptetjük a biztosítottakat, és a már korábban megszerzett *bonus* pontokat (esetünkben például egy tárgy elbukását) feljegyezzük és lesorolással büntetjük. Ennek használata lényeges a modellünk konstruálásakor, mivel a bukások okozta kárt, tehát az esetleges csúszást nem lehet visszafordítani. A modellben éppen ezért a teljesítmény értékelésekor szigorú megszorításokat kell eszközölnünk egy első féléves tárgy bukásánál, hiszen az kihat a teljes tanulmányi időszakra.

Az alábbi fejezet három alfejezetre bomlik, melyek mindegyike a korábbi évek statisztikáinak feldolgozásával foglalkozik.

A (3.1)-ben megismerkedünk a rendelkezésre álló információkkal, adattáblákkal. A meglévő adathalmazokat felkészítjük az elemzésekre és számításokra megfelelő konverziókkal.

A (3.2) alfejezetben törekszünk a leghasznosabb információkat kinyerni a kapott adatokból, és ezek alapján elvégezni a modellezéshez szükséges kalkulációkat. Kiszámítjuk azokat a tapasztalati ismérveket, melyek segítségével következtethetünk, és alkalmazhatjuk az elméleti háttérként összefoglalt tudásunkat.

Végül a (3.3)-ban konstruálunk egy szimulátor programot, mely segítségével megbecsülhetjük a várható kárkifizetést az egyes esetekben, majd ebből kiszámítva a díjakat előrejelzést adhatunk a *StudIn Biztosító Cashflow*-járól, tehát a pénzének áramlásáról a szimulált adatok fényében.

3.1. Adatfeldolgozás

3.1.1. Források

A számításokhoz a TTK Tanulmányi Osztályáról kapott három különböző táblázatot dolgozunk fel, amelyek a következő adatokat tartalmazzák:

- Az első forrásban a 2013-tól a képzésre beiratkozott és 2018-ig a szakot sikeresen végzett diákok jogviszonyának kezdeti és megszűnési időpontja, az egyetemen töltött aktív féléveik száma, illetve a státusz (diplomát szerzett / abszolvált) és a jogviszony megszűnésének oka szerepel (például sikeres kimeneti vizsga).
- A második forrásban a matematika érettségi eredmények, első félévre vonatkozó gyakorlati jegyek és vizsgaeredmények, illetve a kritérium tárgy pontszáma szerepel valamely 2013 után kezdett évfolyamra nézve.
- A harmadik táblázat a 2014-től a szakra bejelentkezett hallgatók státuszát, jogviszony kezdetét és végét jelző dátumot, az aktív félévek számát, és - amennyiben ez már megtörtént - a jogviszony megszűnésének okát tartalmazzák.

Természetesen az adatok név és egyéb azonosító nélkül kerültek feldolgozásra, így a statisztikákban szereplő hallgatók személyi jogai nem sérültek.

A táblázatokat az *R*-ban gyakran használatos *data.frame* objektumként tároljuk el különálló adattáblákként (Tábla 1: *Jogviszony*, Tábla 2: *Vizsga*, Tábla 3: *Kilepo*). Mivel nem áll fenn kapcsolat a forrástáblák között, így az adathalmazokat egymástól függetlenül elemezzük. Itt fontos megjegyezni, hogy a *Jogviszony data.frame* tartalmaz általános információkat a diákok végzési adatairól. Az innen kinyert statisztikákat a *Vizsga* adathalmazból készített kalkulációk, illetve a későbbi számítások és a szimuláció ellenőrzésére tudjuk majd felhasználni. A *Kilepo* táblát a szakról végbizonyítvány megszerzése nélkül távozó hallgatók modellezésére alkalmazzuk.

3.1.2. Tisztítás, konverzió

Ez a részfolyamat kizárólag a *Vizsga* eredményeket tartalmazó tábla esetén szükséges, mivel a másik két forrásban megfelelően fel vannak töltve az értékmezők.

A tábla átdolgozásához kiválogatjuk azokat a sorokat, melyek hiányosak egy, vagy akár több jellemzőjükben is. Ennek eszközeként létrehozunk egy érvényesség jelző oszlopot, amivel kiszűrhetőek a problémás esetek ('Y' - Yes, 'N' - No). Előfordulnak olyan sorok, ahol az első féléves gyakorlati-, és vizsgajegyek mindegyike üres értéket tartalmaz. Ezeket a rekordokat nem lehet felhasználni az elemzéshez, így inaktív adatként tároljuk őket.

Következő lépésként az adathalmazok oszlopait alakítjuk át a könnyebb kezelhetőség érdekében. A félévek száma, a vizsgák és gyakorlatok eredményei, illetve az érettségi és kritérium dolgozat pontszámának formátuma szám, a jogviszony kezdete és vége dátum, míg a leíró adatok (például: jogviszony megszűnésének oka) szöveg. Ezeket a megfelelő típusra konvertáljuk, amennyiben szükséges. A gyakorlati jegyek esetében előfordulnak olyan adatpontok, melyek nem szám formátumban lettek rögzítve, így ezeket javítjuk.

A konverziók elvégzése után a *Vizsga data.frame* objektumban a típusokat meghatározó adatok, mint az érettségi, vagy az egyes tárgyak szintje mind *factor*ként lettek tárolva, míg a jegyek *numeric* adattípusú listákként kerültek be az adathalmazba. Ezt a *str()* függvény segítségével ellenőrizhetjük, amely a paraméterként beadott halmaz struktúráját jeleníti meg. A teljes *Vizsga* adattábla felépítését az alábbi kódrészlet mutatja:

```
> str(Vizsga)
'data.frame': 67 obs. of  10 variables:
 $ ExamResult     : num  100 100 100 98 98 98 98 97 97 96 ...
 $ ExamType       : Factor w/ 2 levels "emelt","kozep": 1 1 1 1 1 1 1 1 1 1
   ...
 $ AlgebraGyak    : num  5 5 5 5 5 5 4 5 2 3 ...
 $ AlgebraKoll    : num  5 5 5 5 5 5 5 5 5 4 ...
 $ AnalResult     : num  5 5 5 5 5 5 5 5 2 3 ...
 $ AnalType       : Factor w/ 2 levels "analizis","kalkulus": 1 1 1 1 1 1 1
   1 1 1 ...
 $ VegesMatType   : Factor w/ 3 levels "halado","intenziv",..: 2 1 1 2 2 1 1
    2 2 2 ...
 $ VegesMatResult: num  5 5 5 5 5 5 5 5 3 4 ...
 $ SzamelmGyak    : num  5 4 5 5 5 4 5 5 4 4 ...
 $ SzamelmKoll    : num  5 5 5 5 5 5 5 5 3 3 ...
>
```

3.2. Előkalkulációk

A modellben a biztosítási díjakat hallgatók egyes csoportjaira külön-külön számítjuk. A kapott értékeket a szimuláció során összegezzük. Ez teszi ki a *StudIn Biztosító* várható bevételét egy adott időintervallumra előrevetítve. Ahhoz, hogy elkészíthessük és kiértékelhessük a szimulációt, létre kell hoznunk a program számára a futáshoz szükséges változókat, adatstruktúrákat és egyéb paramétereket. Jelen fejezet tehát a szimuláció előkészítésének és kiértékelésének lépéseit írja le.

A (3.2.1) alfejezetben elemezzük a félévekre vonatkozó *Jogviszony* adatok segítségével a kockázatot. A modellünkben a kockázat két értéket vehet fel, és egyedüli paramétere a $6+2$ féléven túlcsúszás valószínűsége. A kockázat értékei a túlcsúszás, és az időben végzés állapota, melyből az előbbi akkor a nyolc féléven túli végzést jelöli, az utóbbi pedig a nyolc féléven belüli eseteket.

Az érettségi pontszámokat transzformáljuk, majd előállítjuk a Markov-folyamat kezdeti állapotát a (3.2.2) alfejezetben. Ehhez a *Vizsga* táblázatot használjuk fel.

Ezt követően készítünk egy algoritmust a (3.2.3) alfejezetben, mely segítségével értékeljük az első féléves teljesítményeket, és elvégezhetjük az átsorolást. Ezt követően kiszámítjuk az átmenetvalószínűségeket az első két állapot között, és a szokatlan adatok miatti torzításokat kiszűrve adunk egy becslést a további állapotok közti átmenetetekre a (3.2.4) alfejezetben.

Lényeges kérdés továbbá az is, hogy milyen számban hagyják el a diákok az egyetemet végbizonyítvány megszerzése nélkül. Előfordulhat, hogy a hallgató képzést vált, nem éri el a minimális kreditértéket, vagy személyes okból kifolyólag otthagyja a szakot. Ezeket az eseteket a *Kilepo* adatok segítségével elemezzük és építjük be a szimulációba a (3.2.5) alfejezetben.

A feladatunk tehát az, hogy a rendelkezésre álló információk felhasználásával feltérképezzük a diákok előrehaladását, a statisztikákban megbúvó mintázatokat, trendeket.

3.2.1. Tapasztalati eloszlás

A *Jogviszony* táblázat tartalmazza a diákok aktív féléveinek számát, amely alapján megfigyelhető az egyetemen töltött szemeszterek tapasztalati eloszlása. Azonban az adattábla éves bontásban csak és kizárólag a 2013-ban felvételt nyert hallgatókról ad teljeskörű információt, mivel az utána következő évfolyamok esetében nincs még meg a végleges eloszlás az időbeli előrehaladás miatt. Ennek eredményeként egyetlen mintából dolgozhatunk a számítások során.

Figyeljük meg a mintánkat az (1) hisztogram segítségével. A hallgatók abszolválási arányát olvashatjuk le féléves bontásban. Az ábrából kivehető, hogy a diákok nagyobb arányban csúsznak egy teljes évet, mint felet. Különösen fontos az utolsó két oszlop, ugyanis ezeknél az eseteknél következik be a káresemény, tehát az önköltséges képzésbe történő átsorolás. A látottak alapján készítsünk tapasztalati eloszlást a végzett hallgatók aktív féléveinek

1. ábra. A félévek számának eloszlása a 2013-ban felvett, és a szakot azóta elvégzett hallgatók statisztikái alapján.

modellezésére.

Jelölje a továbbiakban $Y \in \mathbb{Z}^+$ az egyes diákok matematika Bsc szakon töltött aktív féléveinek számát a végbizonyítvány megszerzéséig. A mintából felírható Y eloszlása:

$$P(Y = y) = \begin{cases} 0, & \text{ha } y < 6 \\ 30/64 \approx 46.9\%, & \text{ha } y = 6 \\ 9/64 \approx 14\%, & \text{ha } y = 7, \\ 20/64 \approx 31.3\%, & \text{ha } y = 8, \\ 5/64 \approx 7.8\%, & \text{ha } y > 8 \end{cases}$$

A modellben szereplő káresemények tehát $P(Y > 8)$ valószínűséggel következnek be, ez pedig közel 8% a tapasztalati minta szerint a 2013-as évfolyam esetében. A szimuláció felépítésében, és a díjkalkulációban is jelentős szerepet játszik Y eloszlása, ugyanis a kapott eredmények ellenőrzéséhez tudjuk majd felhasználni, illetve a díj meghatározása esetén a károsszeg várható értéke 1 000 000 $\cdot P(Y > 8)$ lesz. Összevetjük a generált adatok eloszlását a tapasztalati minta eloszlásával, hogy megbizonyosodjunk a szimuláció hitelességéről (3.3.4).

3.2.2. Besorolás, érettségi pontszámok normálása

A hallgatók a szakra történő bejelentkezést követően köthetnek szerződést a biztosítóval. Ekkor még nem tudunk semmit tanulmányi teljesítményükről, azonban nem bízhatjuk teljesen a véletlenre, hogy melyik diák mekkora összeget legyen köteles az első félév során befizetni. Ezen probléma megoldására használjuk a már meglévő információnkat, a matematika érettségi pontszámokat.

Az osztályokba sorolás, melynek velejárója a differenciált díjszabás, a biztosítottak kontraszelektáltságát csökkenti, ami kedvező a biztosítóra nézve [10]. Hiszen a megfelelően alacsony díjösszegekkel olyan diákokat is bevonzunk a biztonságot nyújtó biztosítással, akik jobb képességeik miatt nagyon csekély valószínűséggel csúsznak túl az állami nyolc féléven. Ezzel szemben, ha csak olyan hallgatók kötnek biztosítást, akik kockázata magas, akkor várhatóan a kárkifizetés is magas lesz. Ezt szeretnénk elkerülni a kategorizálással.

A legtöbb hallgató emelt szintű vizsgával érkezik a szakra. Mivel ez nem kötelező a felvételhez (lehet tenni bármely másik természettudományos tárgyból emelt szintű vizsgát a 2019. évi általános felvételi követelmények alapján), így vegyesen előfordulnak közép és emelt érettségit tett diákok. A szintkülönbség kiküszöbölését valamilyen korrekció segítségével végezhetjük el.

A korrigálásnál figyelembe vesszük az emelt és középszintű dolgozatok közti nehézségi különbséget, és ezek alapján normáljuk az értékeket. Köztudottan az emelt szintű érettségivel a diákok 50 plusz pontot szerezheznek a felvételi eljárás során, amennyiben elérnek legalább 45%-ot. Az emelt eredményekhez emiatt 50 pontot adunk hozzá (amennyiben elérik a 45%-os alsó határt), míg a közép szintű érettségik esetében megtartjuk az eredeti pontszámot. Ezek után végezzük el a normálást, tehát a maximálisan elérhető 150 pontszámmal leosztunk, majd szorzunk 100-al. Ettől nem változik a skála terjedelme, a pontszámok a $[0, 100]$ intervallumból kerülnek ki.

A *data.frame* objektumot, amelyben a *Vizsga* adatok szerepelnek, bővítjük egy *CorrExamResult* oszloppal. A továbbiakban az érettségi pontszámokkal történő kalkulációk mindegyikében ezt az oszlopot használjuk.

A Markov-folyamat (2.2.2) kezdeti állapotaként a hallgatókat csoportokra bontjuk. A korrigált és normált pontszámok alapján kerülnek besorolásra négy osztályba:

- *A osztály* - akik a korrigált pontszámok esetében a harmadik kvartilis, azaz 96 pont feletti értékkel rendelkeznek,
- *B osztály* - azon hallgatók, akik a mediánnál, azaz 88.14 pontnál jobb eredményt érnek el, de legfeljebb a harmadik kvartilissal megegyező pontszámot,
- *C osztály* - akik az első kvartilisnál kisebb pontszámmal rendelkeznek, ami 84 pont, de legfeljebb a mediánnal megegyező eredményt érnek el,
- *D osztály* - az első kvartilis alatt teljesítő hallgatók.

Vizsgáljuk meg a besorolás utáni csoportbontásokat. Az így kialakult biztosítási osztályok (A,B,C és D *osztály*) eloszlása a következő:

$$A \approx 24\% + B \approx 24\% + C \approx 27\% + D \approx 25\% = 1$$

A *data.frame* objektumot bővítjük az *Allapot0* oszloppal, ahol a fenti kategorizálás eredményét tároljuk adatsoronként, azaz hallgatónként. Ezzel a lépéssel elkészült a Markov lánc kezdeti állapota.

3.2.3. Első féléves eredmények

Az első félév végére már élesebb képet kapunk a hallgatók képességeiről. Ismerjük az alapozó tárgyakon szerzett jegyeiket, tanulmányi átlagukat. Ezekkel az információkkal szeretnénk a továbbiakban dolgozni. A célunk, hogy kidolgozzunk egy algoritmust, mely segítségével elemezzük a diákok jegyekben mért teljesítményét, és besoroljuk őket a négy biztosítási osztály valamelyikébe.

Mint ismeretes, tantervi háló segítségével vizsgálható [4], hogy az első féléves tárgyak bukása esetén mely előfeltételek nem fognak teljesülni a későbbi tárgyaknál. Az elméleti, alkalmazott és elemző matematikus tantervet folytató hallgatókra más-más hatással lehet egy alap tárgy (pl. Algebra1) sikertelensége. A második féléven és a szakirányokon szerzett jegyekkel azonban nem tudunk számolni, mert nem rendelkezünk információval róluk.

A kategorizálás során elsődlegesen a tanulmányi átlagot vesszük figyelembe, de a felsőbb osztályokba kerüléshez vizsgáljuk, hogy a hallgató bukott-e valamiből az első félév során.

Az első félév során az alapvető kötelező tárgyak az Analízis1/Kalkulus1, Algebra1, Véges Matematika1, illetve a Számelmélet. Ezek mindegyikéhez tartozik gyakorlati óra és előadás is. A további választott és kötelezően választott kurzusokról nincs információnk, de azok hatása a tanulmányokra figyelmen kívül hagyható az elemzés ezen szakaszában.

Az algebra és számelmélet tárgyakhoz az adathalmazban találunk gyakorlati jegyet és kollokvium jegyet tároló mezőt, míg az analízis és a véges matematika tárgyaknál egy összesített érték szerepel, tehát itt nincs bontás a gyakorlaton és a vizsgán szerzett eredmények között. Így a *Vizsga data.frame* objektumban hat oszlopban szerepelnek tárgyi eredmények. Ezeket az értékeket szeretnék soronként (egy sor egy hallgatónak felel meg) átlagolni, illetve eltároljuk az elégtelen jegyek számát, mivel ez a két információ elegendő a további számításokhoz.

Az algoritmusban az alábbi lépéseket végezzük:

1. kiemeljük a *Vizsga data.frame* azon hat oszlopát, melyben a vizsgaeredmények és gyakorlati jegyek találhatóak,
2. a részhalmazt eltároljuk mátrixként, mely a tapasztalati minta esetében 67 sorból és 6 oszlopból áll,
3. kiegészítjük egy oszloppal a mátrixot, melyben tároljuk a bukásokat,
4. kiveszünk egy sort, melyre átlagot számolunk, és eltároljuk az elégtelenek számát,
5. kiértékeljük a sort, azaz besoroljuk valamely osztályba a hallgatót,
6. iterálunk a mátrix sorain.

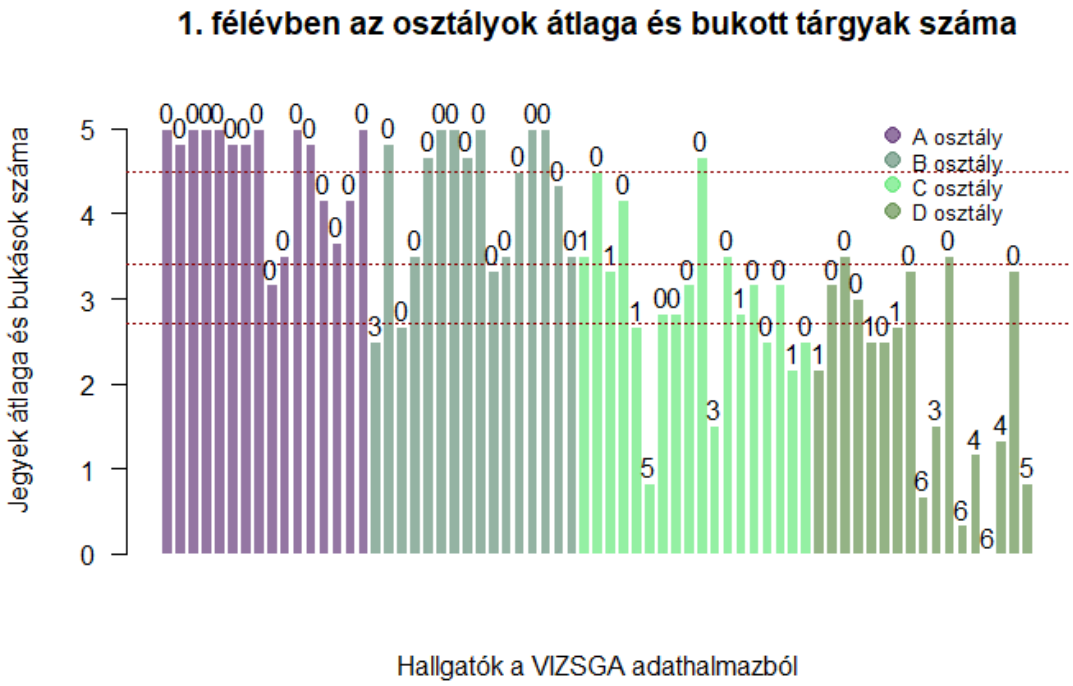
A lényeges lépés a biztosítási modell szempontjából az **5.** pont. Két paramétert használunk, az átlagot és a bukások számát. A két felső, tehát az *A,B* osztályok esetén nem engedjük meg az elégtelen jegyeket. Itt fontos megjegyezni, hogy előfordulnak olyan értékek a *Vizsga* adathalmazban, ahol 0 szerepel a tárgyra kapott értékelés mellett. Ez a valóságban azt jelenti, hogy a diák nem vette fel a kurzust, így jegyet sem kapott rá. Ilyen esetben kivételesen belevesszük a számításba a 0 értéket is (átlagnál nem csökkentjük a tárgyak számát, és bukásnak is beszámítjuk). Tehát a 0-t is 1-nek tekintjük, hiszen ez az eset ugyanúgy okozhat csúszást a későbbiekben, mivel nem teljesítette a hallgató az első félév során az alapozó tárgyat. A határokat igyekszünk úgy kialakítani, hogy az osztályok létszámai megegyező számosságúak maradjanak. Így az alábbi kritériumok fogalmazhatóak meg a második állapotra:

Azok a hallgatók, akik

- nem szereztek elégtelen érdemjegyet, és átlaguk legalább 4.5, ők kerülnek az A osztályba,
- nem szereztek elégtelen érdemjegyet, és több, mint 3.5-ös átlaggal állnak, de kevesebb, mint 4.5-tel, ők kerülnek a B osztályba,
- átlaga legalább 2.7 és legfeljebb 3.4 (de itt és az alsóbb kategóriában már megengedett a bukás), ők kerülnek C osztályba,
- átlaga kisebb, mint 2.7, ők kerülnek a D osztályba.

Ezzel a módszerrel meghatároztuk a második állapotot, melyet *Allapot1* néven eltárolunk az *Adathalmaz data.frame* objektumba.

Az algoritmusban specifikált értékek indoklására szolgál a (2) ábra. Itt láthatjuk, hogy az A osztályba sokan kerülnek a B-ből, illetve a C,D osztályok tagjai néhány diáktól eltekintve maradnak az alsóbb kategóriák valamelyikében. Az ábrán látszik, hogy az A osztályban kezdők nagy része továbbra is ugyanazon kategóriában marad, ha azonban rosszabb besorolást kap, akkor sem bukás miatt. Tehát ilyenkor lecsúszik a B vagy C kategóriába.



2. ábra. Az első féléves tanulmányi átlagok, és bukások száma hallgatónként a *Vizsga data.frame* objektumból. Piros szaggatott vonallal az algoritmusban használt határvonalak láthatóak, a színek a kezdeti besorolás osztályait jelölik.

3.2.4. Átmenetvalószínűségek simítása, váratlan értékek simítása

Az érettségi pontszámok segítségével besoroltuk a diákokat a tanulmányaik kezdetekor a négy biztosítási osztály egyikébe. Egy félév elteltével teljesítmény alapján készítettünk egy algoritmust az átsorolásokra. Most vizsgáljuk meg, hogy milyen átmenetvalószínűségeket kaptunk az állapotváltozás során. Mivel alacsony mintaelemszámmal, és egyetlen

mintával dolgozunk, így előfordulhatnak esetleges értékek, melyek kiszűrése nélkül rossz következtetésekre juthatunk, és nehezíthetik a szimuláció összeállítását.

Adathalmaz$Allapot0 és *Adathalmaz$Allapot1* oszlopok összehasonlításával megkapjuk a tapasztalt átmenetvalószínűségeket. Ezt úgy számítjuk, hogy megnézzük az első állapot alapján mind a négy osztályra, hogy hányan kerültek át egy másik csoportba, és mennyi diák maradt a kezdetben kiosztott kategóriájában. Így kapunk egy tapasztalati valószínűséget az osztályok közti mozgásokra.

Nézzük meg, hogy milyen értékeket kapunk az átmenetvalószínűség mátrixra:

$$P = \begin{pmatrix} 0.6875000 & 0.2500000 & 0.0625000 & 0.0000000 \\ 0.5625000 & 0.2500000 & 0.1250000 & 0.0625000 \\ 0.1111111 & 0.1111111 & 0.5555556 & 0.2222222 \\ 0.0000000 & 0.1176471 & 0.2941176 & 0.5882353 \end{pmatrix}$$

A mátrix sorai fentről lefelé, míg az oszlopai balról jobbra haladva az *A,B,C,D osztályokra* vonatkozó átmenetvalószínűségeket mutatják az oszlop, ahol P mátrix diagonálisa jelöli az egyes csoportokban a maradók valószínűségét.

A kapott értékek esetén elég erős kilengéseket figyelhetünk meg. Például a B osztály esetén azt látjuk, hogy kevés olyan hallgató van, aki maradna a korábban meghatározott kategóriában, nagy részük ugyanis átkerül az A osztályba. Természetesen ez megtörténhet a valóságban, azonban hosszú távon teljesen eltűnhet ez az osztály, így a szimuláció során szükséges az átmenetek korrigálása a további állapotoknál.

A P átmenetvalószínűség mátrix értékei a kilengések ellenére nem feltétlen torzítanak, hiszen a gyengébb kategóriában lévő hallgatók alacsonyabb osztályokban maradnak, és hasonlóképp történik ez a felsőbb osztályok esetében is.

Feltételezzük, hogy az első félév (sok esetben a második is, azonban erről nincs információnk) vízválasztóként szolgál a hallgatók további tanulmányaival kapcsolatban. Korábban már kitárgyaltuk a (3.2.3) alfejezetben az első féléves tárgyak hatását a későbbi tanulmányokra nézve, illetve a következő (3.2.5) alfejezetben megvizsgáljuk, hogy milyen százalékban hagyják ott a diákok a képzést félévekre lebontva. Mind a két elemzésből látszik, hogy az első félévben elért teljesítmény határozza meg leginkább a hallgatók előrehaladását. Ezen feltevésünk alapján kezeljük a további átmeneteket homogén Markov-láncként

(2.2.2), mely értékeit a P mátrix korrigálásával határozzuk meg.

Fontos továbbá megjegyezni, hogy a homogén átmenetek csupán a hatodik félévig tartanak, ott ugyanis már máshogy vizsgálódunk a végbizonyítványt szerzett hallgatók miatt.

A becslésünk a homogén átmenetvalószínűség mátrixra a következő:

$$P_{hom} = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

Ez duplán sztochasztikus, ami biztosítja, hogy a kiinduló egyenletes eloszlás stacionér eloszlás lesz, azaz nem fog a megoszlás lényegesen változni a szimulációk során.

A további féléveknél a szimulációban a P_{hom} mátrix elemeivel határozzuk meg a csoportok közti átmeneteket, azonban az első félév esetében még a korábbi simítatlan P mátrix szerinti átmeneteket alkalmazzuk a szimulációra. Ennek oka, hogy feltételezzük az első szemeszterről, hogy akár váratlan eredményeket is produkálhatnak a diákok, az érettségi pontszámok alapján tett besorolásunkhoz képest, mind pozitív, mind negatív irányban. Ezt a feltevést használjuk a szimuláció során is, azaz az első két állapot közti $p_{i,j}$ átmenetvalószínűségeket a P mátrixból vesszük.

3.2.5. Képzésből kilépő hallgatók esete

Általánosságban elmondható, hogy a egyetemek alapképzésein az első év során a legmagasabb a lemorzsolódási ráta. Ennek különböző okai lehetnek, amelyek pszichológiai vonatkozásait, hátterét nem feltétlen ismerjük. Azonban statisztikai adatokkal alátámaszthatjuk a jelenség meglétét.

A *Kilepo* adathalmazból szeretnénk igazolni a feltevésünket. Vizsgáljuk meg az (1) táblázatot. Az ábrán a 2014-es évfolyam azon hallgatóinak számát láthatjuk félévekre lebontva, akiket elbocsátottak, töröltek, vagy saját döntésből szakot váltottak. Ezen az évfolyamon a kilépők aránya $37/95 \approx 39\%$ volt, tehát a hallgatók közel 2/5-e nem fejezte be a képzést.

Előfordul olyan eset, amikor a diák még a tanulmányok megkezdése előtt megváltoztatja továbbtanulási döntését, és nem kezd bele az egyetemi éveibe. Szerencsére ezen hallgatók

aránya a tapasztalati minta alapján csupán 1%. A diákok többsége bejelentkezik a szakra, azonban az első két félév során mintegy 23.1% félbehagyja tanulmányait. A másod-, illetve harmadévben ez az arány monoton csökkenést mutat. Emellett a standard hat félév utáni időszakban elenyésző számban történik kilépés. A 2014-es évfolyam mintája alapján tehát azt látjuk, hogy a feltevésünk helytálló, így a modell konstruálása során felhasználhatjuk a tapasztalt arányokat.

Félév	0	1	2	3	4	5	7	9	**Össz. kilépő**	**Össz. felvett**
Kilépők	1	12	9	5	5	2	2	1	**37**	**95**

1. táblázat. 2014-es évfolyam kilépői félévekre bontva, és a felvettek száma.

A kilépési adatok nem tartalmaznak információt arról, hogy a szakot elhagyó hallgatók milyen eredményeket értek el tanulmányaikban. A tapasztalat azonban arra enged következtetni, hogy a gyengébb teljesítményt nyújtó diákok döntenek a kilépés mellett. Mivel a felső két osztályban a második állapot kialakítása során nem engedtük meg az elégtelen érdemjegyet, ezért ezekhez a csoportokhoz kisebb értéket rendelünk, míg a másik két osztályhoz nagyobbat.

Készítünk egy becsült arányt, amelyben lefektetjük a biztosítási csoportok közti kilépési különbségeket. Itt megengedjük a felső és alsó két-két kategória között, hogy egyenlő arányban vegyük a kilépőket. Mivel láthattuk korábban a (2) ábrán, hogy az *A,B* és *C,D* osztályok esetében hasonló eredményeket érnek el a hallgatók, így feltételezhetjük a hasonló kilépési arányt is. Az általunk kreált értékek a következők:

- A osztály esetén $2/10 = 20\%$,
- B osztály esetén $2/10 = 20\%$,
- C osztály esetén $3/10 = 30\%$,
- D osztály esetén $3/10 = 30\%$.

Ezeket az arányokat osztjuk el minden egyes félévre az (1) táblázat alapján számolt tapasztalati eloszláshoz igazítva.

3.3. Szimuláció

A számítások során szerzett tapasztalatok alapján szeretnénk előrejelezni a biztosítóhoz befolyó befizetéseket, és a káresemények bekövetkeztéből eredő kifizetéseket. Ahhoz, hogy hiteles eredményeket biztosítsunk, elengedhetetlen a (3.2) fejezetben tapasztalt tulajdonságok, észrevételek és megszorítások beépítése a szimulált adatainkba.

Mint a *StudIn Biztosító* elemzői, a fő feladatunk az, hogy megismerjük a célközönséget, akinek terméket tudunk értékesíteni. Ehhez közel sem elegendő bizonyos tézisek felállítása, szükséges valamilyen módon létrehozni egy adathalmazt, melyből vizuálisan értelmezhetővé tehetjük a vezetőség számára a döntéshozatalhoz szükséges információkat.

A szimuláció hitelességének súlyát mi sem írhatja le jobban, mint hogy a *StudIn Biztosító* vezetői az általunk készített előrejelzés alapján tervezik alakítani a cég jövőjét. Természetesen szeretnék a lehető legjobban és legpontosabban elvégezni a feladatot, így a következő néhány alfejezet során végigmegyünk a szimulálás egyes lépésein, kezdve az első (3.3.1)-gyel, ahol a tapasztalt eloszlásokat felhasználva létrehozzuk a szakra a jövőben bekerülő diákok érettségi eredményeit, és az első féléves jegyeiket. Majd a (3.2.4) alfejezetben konstruált átmenetvalószínűségek segítségével legeneráljuk az egyes diákok elképzelt előrehaladását a tanulmányaik során, egészen a hatodik félévig a (3.3.2)-ben. Ezen a ponton megvizsgáljuk, hogy ki mekkora gyakorisággal került az egyes biztosítási osztályokba az elmúlt évek során. A (3.3.3) alfejezetben készítünk egy algoritmust, melyben definiáljuk a végbizonyítványt szerzett hallgatók kiléptetését. A végzésig a szakon töltött félévek számának arányát összevetjük a (3.2.1)-ben tapasztaltakkal, hogy megbizonyosodjunk a szimuláció hitelességéről.

Ezek után a díjkalkulációs elvek gyakorlati alkalmazásával fogalkozunk a (3.3.5) alfejezetben. Beépítjük a modellbe a várható érték módszert, hogy végül a (3.3.6)-ban megvizsgálhassuk a *StudIn Biztosító* jövőjét.

A szimulációt 100 évfolyamra generáljuk előre. Ebből a halmazból határozunk meg egy tapasztalati eloszlást az egyes biztosítási osztályok kockázatára, amelyek a díjkalkulációhoz elengedhetetlenek. A modellben igyekszünk a meglévő mintánkhoz igazodni, azonban a kevés rendelkezésre álló információmennyiség miatt néhol szükségesek bizonyos korrekciók a szimuláció létrehozásában.

A modell konstruálása során feltesszük továbbá azt, hogy az esetleges káresemény bekövetkezésekor a hallgató él a biztosító által biztosított mindkét finanszírozott félévvel, tehát a károsszegünk minden bekövetkezés esetén a két félévre kiszabott 1 millió Ft lesz. Mivel ezt az értéket szabtuk felső korlátnak, így az ezen összegen felüli károkat nem finanszírozza a biztosító.

3.3.1. Adatgenerálás kezdeti állapottal

A folyamat talán legfontosabb állomása a tapasztalati minta megfelelő közelítése, hiszen a célunk, hogy a előrejelezzük a jövőben valószínűsíthető történéseket. Ehhez kizárólag a jelenlegi ismereteinkre hagyatkozhatunk, így a már meglévő információk alapján állítjuk össze a szimuláció indításához szükséges adathalmazokat.

Elsődlegesen egy olyan helyzetet kell kialakítanunk, melyben a felvételi eljárás már véget ért, és egy új évfolyam érkezik a ELTE matematika Bsc képzésére, ahol ismerjük a hallgatók érettségin szerzett pontszámát matematikából, de további információink nincsenek.

Az érettségi eredmények generálásánál ismerjük a tapasztalati minta statisztikáit a különböző biztosítási osztályokra, és ezekhez igazítjuk a szimulált adatokat. Továbbá szükségünk lesz a (3.2.2) alfejezetben kalkulált biztosítási osztályok közti arányokra is.

A modell egyszerűsítése kedvéért nem generálunk külön emelt és közép érettségi eredményeket, hanem a normált értékeket hozzuk létre (3.2.2). Ennek elkészítéséhez a [3] hivatkozásban található *rtruncnorm()* függvényt alkalmazzuk. A korrigált eredményeket egy olyan normális eloszlásban határozzuk meg, melynek a szokásos μ és σ paramétereein kívül megadható a generálni kívánt véletlen számok minimuma és maximuma, illetve elemszáma.

A generátor folyamat egy 100 fős évfolyam esetén a következő:

1. kiszámítjuk a *Vizsga$CorrExamResult*, azaz a korrigált érettségi pontszámok átlagát, szórását, minimumát és maximumát,
2. a statisztikákat megadjuk az *rtruncnorm()* függvény paramétereiként (elemszám, legkisebb érték, legnagyobb érték, átlag, szórás),
3. véletlen számokat generálunk, és vektorként tároljuk azokat,

4. a generált értékeket a (3.2.2) alfejezetben készített érettségi pontszám alapján besoroló módszerrel osztályozzuk, így kapjuk a kezdeti állapotot.

A módszer **3.** lépésének alapötlete az, hogy a jövőben nem mindig a tapasztalati mintában számolt pontszámeloszlást fogjuk látni a következő évfolyamoknál. Ezzel modellezzük a véletlen ingadozásokat.

Az érettségi pontszámok mellett szeretnénk generálni első féléves jegyeket is, melyeket kiértékelését használjuk a későbbiekben a második állapot létrehozásához. Ehhez a *sample()* függvényt alkalmazzuk, amely egy lista vagy vektor elemeiből a neki megadott számban egy újabb vektort készít, mintavételezést hajt végre. Előnye, hogy vektorként előre megadhatjuk, hogy a lista egyes elemeit milyen valószínűséggel válassza a véletlen generálás során. Mivel a tapasztalati mintánkban ki tudjuk számolni tárgyanként, hogy hány darab elégtelen, elégséges stb. jegyet szereztek a hallgatók az adott biztosítási csoportban, így elő tudjuk állítani a szükséges valószínűségeket.

Az algoritmus lépései az alábbiak:

1. kiszámítjuk a tapasztalati mintában tárgyanként és azon belül jegyenként a bekövetkezési valószínűségeket,
2. a nyers bekövetkezési valószínűségeket eltároljuk egy vektorban,
3. a *sample()* függvénynek megadjuk paraméterbe, hogy mekkora valószínűséggel generáljon értékeléseket, továbbá milyen hosszú eredményvektort szeretnénk kapni,
4. majd generálunk a meghatározott valószínűségekkel mindegyik alapozó tárgyra egy vektort, ami a szerzett jegyeteket tartalmazza.

A jegyek generálását a kezdeti állapot osztályaihoz igazítjuk, azaz előre kiválasztjuk, hogy a kezdetben az A,B,C vagy D osztályba sorolt diákok jegyeit állítjuk elő.

A kapott érettségi pontszámok és jegyek vektorait összefűzzük, majd elvégezzük a fenti folyamatot mind a négy biztosítási osztályra. A (3.2.2) alfejezetben specifikált algoritmus alapján meghatározzuk, hogy a diákok létszáma hogyan oszlik meg az egyes csoportok között.

Nem hagyhatjuk figyelmen kívül a (3.2.5) alfejezetben tárgyaltakat, miszerint már az első félév megkezdése során előfordul olyan diák, aki nem jelentkezik be az egyetemre. Ezt

a kiesést egy újabb véletlen generálással oldjuk meg szintén a *sample()* függvénnyel. Itt viszont nem tudjuk meghatározni, hogy vajon melyik hallgató volt az, aki nem kezdte el tanulmányait, ezért a tapasztalt 1%-os rátával kisorsoljuk jelen esetben azt az egy hallgatót, aki már a kezdeti állapotában kilépőként lesz regisztrálva az adathalmazunkban.

3.3.2. A Markov-folyamat állapotainak létrehozása

Rendelkezésünkre állnak a generált érettségi pontszámok, gyakorlati és vizsgajegyek, és a Markov-lánc kezdeti állapota. Következő lépésként alkalmazzuk a tanulmányi átlag alapján átsoroló algoritmust az adatainkra. Ennek eredményeként megkapjuk az újabb biztosítási csoportbeosztást, és folytathatjuk a további félévek szimulációját a homogén átmenetvalószínűségekkel.

Mivel az első félév végén még rendelkezünk információval a hallgatók teljesítményéről, így a (3.2.4) alfejezetben alkalmazott algoritmust használjuk az átsorolások modellezésére. A másodiktól az ötödik félévig azonban már nincsen adatunk a hallgatók teljesítményéről, ezért a szintén a (3.2.4)-ben konstruált P_{hom} homogén átmenetvalószínűség mátrix segítségével léptetjük tovább a diákokat a további félévekben. Ennek segítségével számoljuk a létszámmozgásokat az egyes biztosítási osztályok között.

Ismét vegyük figyelembe, hogy a (3.2.5) alfejezetben meghatároztuk a szak elvégzése előtt kilépők eloszlását. A korábban megbecsült kilépési rátát alkalmazzuk a félévek során a következőképpen:

1. megnézzük, hogy a szimulációban melyik félévnél járunk, és a rá érvényes elhagyási aránnyal számolunk,
2. beszorozzuk a szimulált évfolyam még versenyben lévő hallgatóinak létszámával, így megkapjuk, hogy az adott félévben hány főt kell kilépőként kezelnünk,
3. a kilépőkre kapott létszámhoz a (3.2.5)-ben specifikált valószínűségekkel kiválasztunk hallgatókat az egyes osztályokból, és kilépőként regisztráljuk őket.

Ezt a módszert iteráljuk egészen a hatodik szemeszterig, ahol már számolnunk kell a szakot sikeresen elvégzett hallgatók kiléptetésével is.

3.3.3. A végbizonyítványt szerzők kiléptetése

A Markov-láncból kétféle módon léptetünk ki hallgatókat. A hatodik félévig csupán az egyik lehetőséget tárgyaltuk, ugyanis azok hagyták el idáig a szakot, akik valamilyen okból kifolyólag megszakították tanulmányaikat. Az első három év elteltével azonban számításba kell vegyük azt az esetet, mikor a diákok az elvártnak megfelelően teljesítettek, és sikeresen elvégezték az alapképzést.

Azt, hogy mely hallgatóknál nem történt csúszás az első hat félév során, a tanulmányi előrehaladásukat tükröző állapotokkal tudjuk modellezni. Az alábbi kritériumot határozzuk meg azokra, akik pontosan hat félév alatt végeztek:

Azok a hallgatók végeznek időben a szakon, akik a hatodik félév végén

- az *A osztályban* tartózkodtak, és legfeljebb háromszor kerültek át a *B osztályba*, de alsóbb kategóriában nem jártak,
- az *A osztályban* tartózkodtak, és legfeljebb egyszer kerültek át a *C osztályba*, ezen kívül legfeljebb egyszer tartózkodtak a *B osztályban*, a többi esetben végig az *A osztály* tagjai voltak,
- a *B osztályban* tartózkodtak, és ezen kívül legfeljebb kétszer voltak ugyanezen csoport tagjai, a többi esetben végig az *A osztály*ban voltak,
- a *B osztályban* tartózkodtak, és legfeljebb egyszer kerültek át a *C osztályba* úgy, hogy a többi esetben az *A osztály* tagjai voltak.

Az első hat félév elteltével nem alkalmazunk további átsorolásokat, mivel az eddigi tanulmányokból szimulálható az egyes hallgatók kilépésének időpontja. Így a továbbiakban a hat félév során végzett módszert alakítjuk át és alkalmazzuk a végbizonyítványt szerzett hallgatók kiléptetésére.

A következő két félév során is azt vizsgáljuk, hogy az egyes diákok milyen biztosítási osztályok között mozogtak, ennek segítségével adunk becslést a végzett hallgatók számára. A fenti módszer átdolgozása a hetedik félév végén a szakott teljesített hallgatókra az alábbi:

Azok a hallgatók végeznek ekkor a szakon, akik a hetedik félév végén

- az *A osztályban* tartózkodtak, és nem jártak még a *D osztályba*,
- a *B osztályban* tartózkodtak, és nem jártak még a *D osztályba*.

Az utolsó államilag támogatott félévben, tehát a nyolcadik szemeszter végén pedig az alábbi algoritmus szerint definiáljuk a végbizonyítványt szerző hallgatókat:

Azok a hallgatók, akik a nyolcadik félév végén

- az *A osztályban* tartózkodtak,
- a *B osztályban* tartózkodtak,
- a *C osztályban* tartózkodtak.

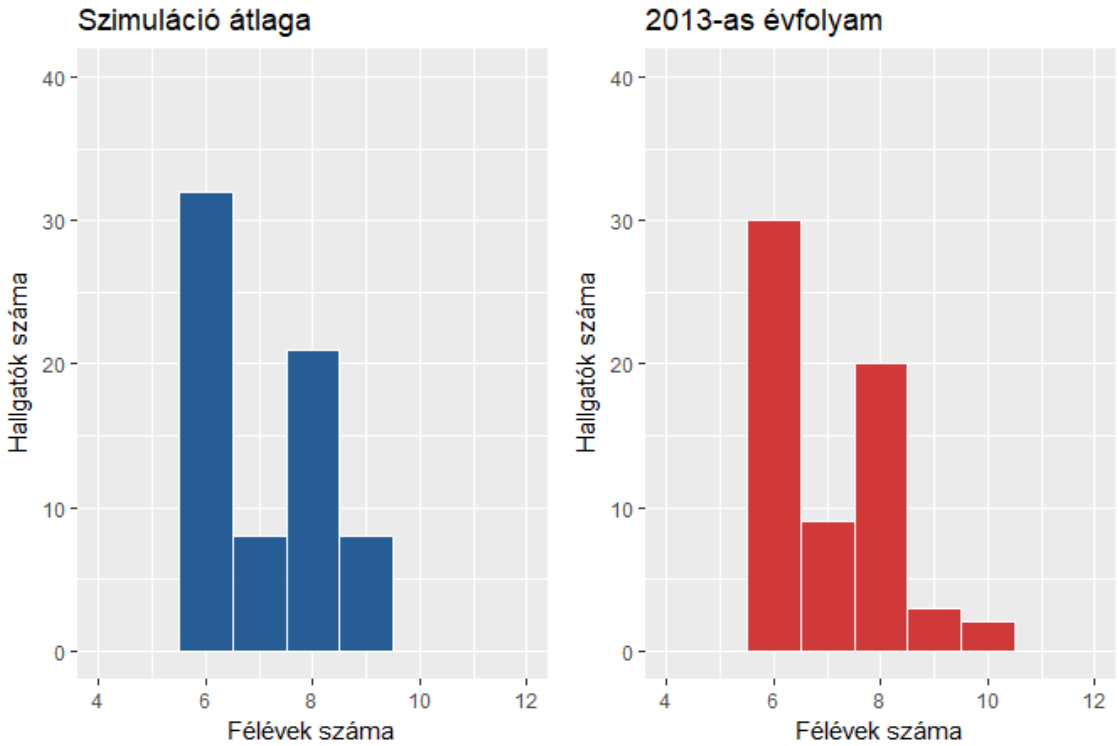
A fenti kiléptetések alapján tehát azok csúsznak ki a $6 + 2$ államilag támogatott félévből, akik előrehaladásuk során nagyrészt a *D osztály* hallgatói voltak, vagy valamilyen oknál fogva a hatodik félév végére lerontották tanulmányaikat, és ez okozta a csúszásukat. Eszerint a kilépés definíciója nem Markov, mivel a hat szemeszter mindegyikét figyelembe veszi. Azonban ez nem jelent problémát, hiszen minden információnk meg van a döntéshez.

A végzési kritériumrendszer kialakítása során arra öszpontosítottunk, hogy monoton kritériumokat határozzunk meg, tehát a felsőbb osztályokban a sikeresség valószínűsége nagyobb legyen, mint az alsóbb csoportok esetében.

Ahogy az eddigi módszereknél véletlenítettünk, úgy itt is vizsgálhatnánk ennek lehetőségét. Információ hiányában azonban a végzések hátteréről még az eddigiekhez képest is megalapozatlanabb lenne bármilyen véletlen struktúra kialakítása. De a biztosító működése során új adatokhoz fog jutni és lesz lehetősége a módszer, és ebből adódóan a díjak finomhangolására is.

3.3.4. Ellenőrzések

Az előző alfejezet során definiáltuk a pozitív kiléptetések menetét félévekre bontva. Most ellenőrizzük a modell hitelességét, és hasonlítsuk össze a 100 évfolyamra generált adataink félévekre vizsgált végzési eloszlását a tapasztalati mintánk eloszlásához viszonyítva.

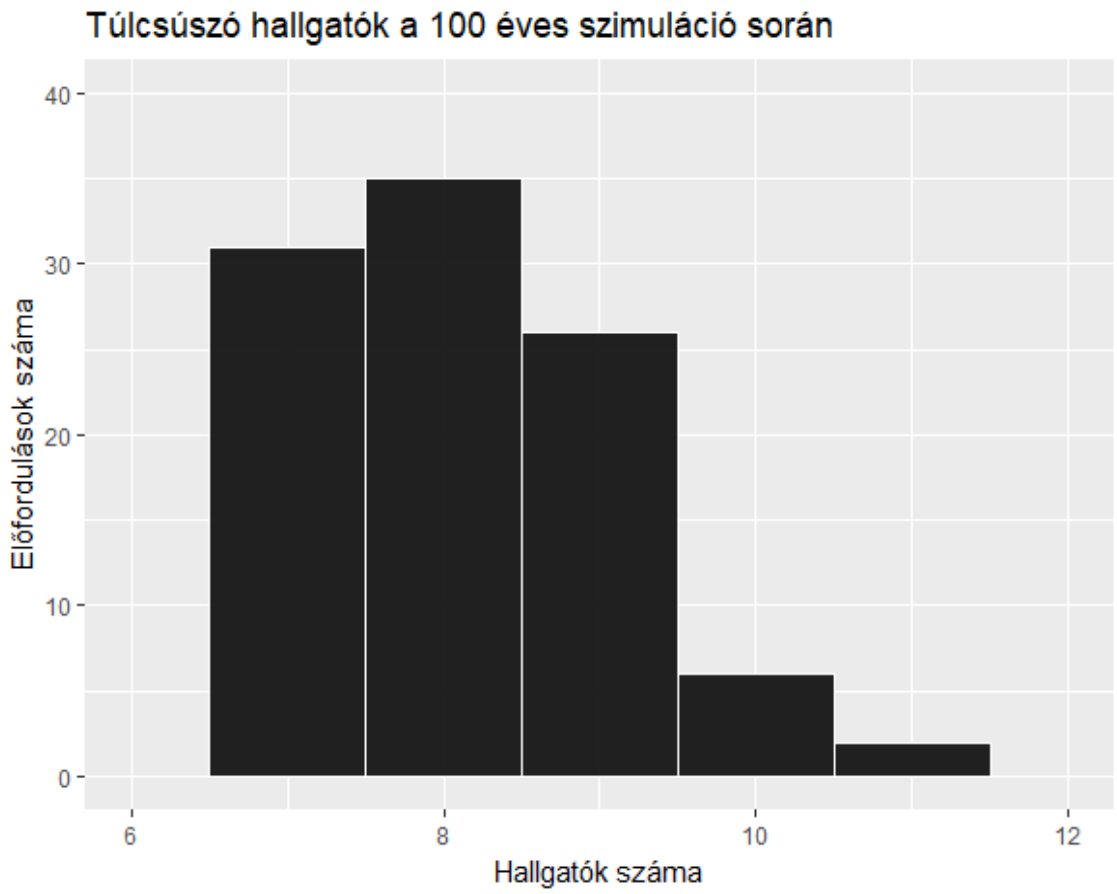


3. ábra. Szimuláció és tapasztalati eloszlás összehasonlítása.

Ahogy azt korábban lefektettük, a szimuláció során nem teszünk különbséget a $6 + 2$ államilag támogatott féléven túli csúszásoknál az egy, és a két szemeszterrel megcsúszó hallgatók között. Ennek elsődleges oka, hogy a biztosítás állja mindkét félévet a diák számára, így számolnunk kell azzal a lehetőséggel, hogy a diák feléli a károsszeg egészét.

A (3) ábrán a bal oldali hisztogramm mutatja a szimulált 100 évfolyam átlagolt végzésszámát félévekre bontva, míg a jobb oldali hisztogrammon a tapasztalati mintául szolgáló 2013-as évfolyam adatait figyelhetjük meg. A két ábrán látható, hogy csak kis mértékben tér el a mintától a generált adathalmaz.

Vizsgáljuk meg a szimuláció során a túlcsúszásokat. A (4) ábrán látható, hogy az évfolyamonként felvett 100 diákból leggyakrabban 8-an kerültek át önköltséges képzésbe. Ez az érték a $[7, 11]$ intervallumon mozog, viszont a tapasztalati mintában a végzett hallgatókhoz viszonyítva kaptunk 8% körüli értéket. Így mondhatjuk, hogy a szimulációnk kissé pesszimista. Ez nem jelent problémát, hiszen nem tér el jelentős mértékben a tapasztaltaktól, és a biztonsági költségpótlék szempontjából is pozitív hatással lehet a *StudIn Biztosító* jövőjére.



4. ábra. A szimuláció során a túlcsúszó hallgatók számának előfordulásai.

3.3.5. Díjak meghatározása

A szimuláció ellenőrzésével megbizonyosodtunk arról, hogy egy kellően óvatos előrejelzést konstruáltunk meg. A díjak kiszabásához azonban szükségünk van a generált adatok alapján a a tapasztalati kockázat kiszámítására.

A bekövetkezett káreseményeknél ismerjük a hallgató korábbi életútját, így számolhatunk a szimulált 100 évfolyam esetén osztályokra bontott kockázatot. Itt lényegében meghatározzuk, hogy azok közül, akik az első félévben az A osztályban voltak, hány esetben csúsztak túl a $6 + 2$ féléven. Majd elvégezzük ezeket a számításokat minden félévre, és azon belül is mind a négy osztályra. A kapott számokat leosztjuk a 100 év során felvett hallgatók összlétszámával, ami 10 000 fő volt a szimulációnkban. Ennek eredményeként mátrixos alakban megkapjuk azokat a tapasztalati valószínűségeket, melyek jellemzik a kockázatot félévre és osztályra lebontva. Ezzel tudunk a későbbiekben díjakat számolni, melyet a hallgató előző állapotbeli helyzete szerint szabunk ki. Így megtartjuk a Markov-

folyamat sajátosságát, és csak az előző féléves teljesítményt vesszük figyelembe a díjkalkuláció során.

A kalkuláció algoritmusa a következő lépésekből áll:

1. létrehozunk egy *array* objektumot, melynek dimenziói $:= (100, 4, 8)$,

2. iterálunk a harmadik dimenzión 1-től 8-ig, ez jelöli a féléves bontásokat,

3. iterálunk a másik két dimenzión (évfolyam száma, állapot), ezzel mátrixokat definiálunk minden félévre,

4. értéket adunk a mátrix elemeinek,

5. a kapott *array* objektumot a második és harmadik dimenzióra felösszegezzük, így kapunk egy 4 sorból és 8 oszlopból álló mátrixot,

6. a mátrix elemeit leosztjuk a szakon a 100 év alatt felvett hallgatók számával, ami pontosan 10 000 a szimulációnkban.

Az algoritmus során a **4.** pontban összeadjuk, hogy a 100 éves szimulációban osztályonként hány olyan hallgató volt, aki túlcsúszott az állami féléveken.

Az algoritmus során kapott kockázat mátrix:

$$M = \begin{pmatrix} 0.0062 & 0.0104 & 0.0057 & 0.0025 & 0 & 0 & 0 & 0 \\ 0.0145 & 0.0091 & 0.0127 & 0.0117 & 0.0078 & 0 & 0 & 0 \\ 0.0284 & 0.0288 & 0.0326 & 0.0346 & 0.0379 & 0.0330 & 0 & 0 \\ 0.0322 & 0.0330 & 0.0303 & 0.0325 & 0.0356 & 0.0483 & 0.0813 & 0.0813 \end{pmatrix}$$

A kapott M mátrix soraiban az egyes osztályokra (A,B,C,D fentről lefelé haladva) vonatkozó tapasztalati kockázatot láthatjuk. Az oszlopokban balról jobbra növekedő sorrendben találhatóak a félévek. Első ránézésre látszik, hogy előfordul olyan félév, ahol az alsóbb kategória kisebb kockázattal jár. Ez bár furcsának tűnik, mégis előfordulhat. Ugyanis a D osztály tagjai már nem kerülhetnek rosszabb osztályba, hiszen vagy továbbra is a D osztályban maradnak, vagy feljebb lépnek a C-be. Ha felkerülnek, akkor valamivel nagyobb eséllyel lépnek ki idő előtt.

Példa: *A 100 év során összesen 62 olyan diák volt, aki az első félévben az A osztályba lett kategorizálva, és végül több, mint nyolc félévig járta a szakot, azaz túlcsúszott. Ezt*

pontosan $M(1,1)$ írja le, tehát az első félévre az A osztályba sorolt hallgatók esetén 0.0062 *a tapasztalati kockázat.*

Alkalmazzuk a várható érték elvet, melyet a (2.1.2) alfejezetben definiáltunk. Legyen $\lambda = 0.1$ paraméterünk, $X_{i,j} : \Omega \to \Re_0^+$ a kockázat az egyes osztályokhoz félévente, ahol

$$P(X_{i,j} = 1000000) = \begin{cases} M(i,j), & \text{ha } i = 1 \dots 4,\ j = 1 \dots 8 \\ 0 & \text{egyébként.} \end{cases}$$

Ebből a következő egyenletrendszert kapjuk:

$$\Pi_{X_{i,j}}(X_{i,j}) = (1 + 0.1) \cdot M(i,j) \cdot 1000000$$

. Így kapjuk meg az alábbi táblázat elemeit, melyek a osztályokra kalkulált díjakat mutatják félévenként.

	1. félév	2. félév	3. félév	4. félév	5. félév	6. félév	7. félév	8. félév
A	6820	11440	6270	2750	0	0	0	0
B	15950	10010	13970	12870	8580	0	0	0
C	31240	31680	35860	38060	41690	36300	0	0
D	35420	36300	33330	35750	39160	53130	89430	89430

2. táblázat. Kalkulált díjak a várható érték elve alapján $\lambda = 0.1$ paraméterrel

A kapott díjak az adott félévre értendő összeget jelölik minden kategóriához. Láthatjuk, hogy az A,B osztályok esetében ez jóval alacsonyabb, mint a alsó két kategóriában.

Továbbá láthatjuk azt is, hogy előfordulnak olyan esetek, mikor egy rosszabb kategóriának jobb díjat kalkulált az algoritmus. Ennek oka, hogy többen kerültek végül önköltséges kategóriába a jobb csoportból.

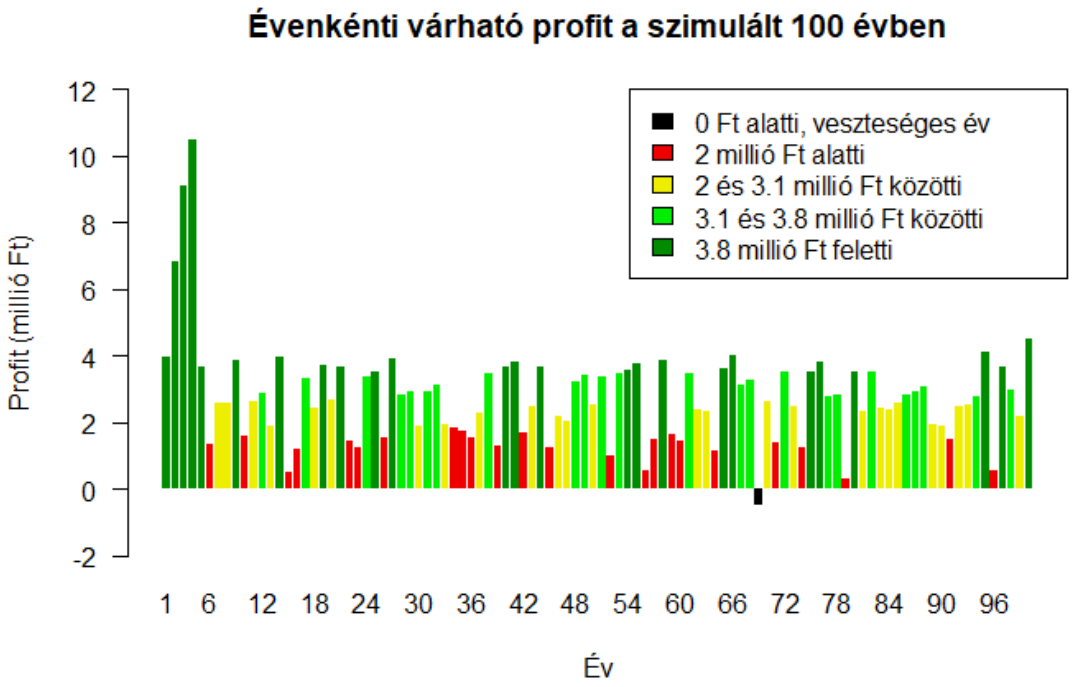
3.3.6. Cashflow

A (2) táblázatban kiszabott díjak alapján vissza tudunk helyettesíteni a szimulációba. Ennek eredményeként elemezhetjük, hogy a befizetett díjak és kifizetett károsszegek függvényében milyen pénzmozgások fordulhatnak elő a jövőben. Lényegében a modell ered-

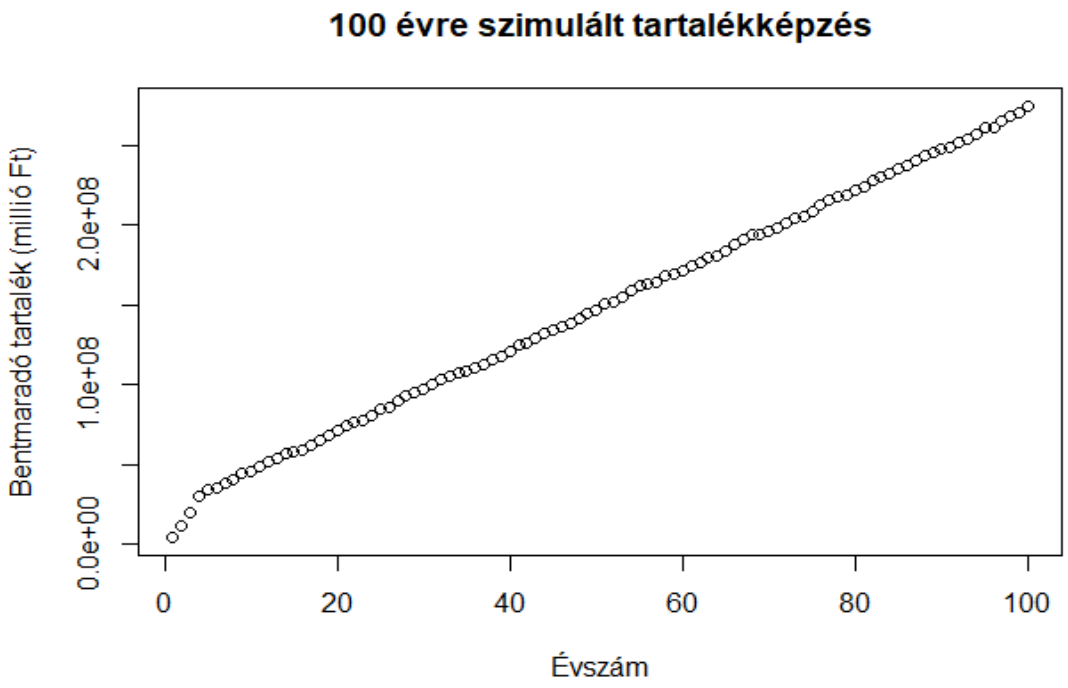
ményéül szolgál a *Cashflow* ábrázolása. A vezetőség ugyanis ebből figyelheti meg a kilengéseket, trendeket, az első néhány félév után felhalmozódott biztosítási költségtartalékot, az első káresemények hatását az éves bevételre.

Az (5) ábrán évekre összegezve mutatjuk a biztosító pénzének áramlását. Ennek kiszámítását úgy végezzük, megvizsgáljuk egy adott évre, hogy milyen befizetések és kifizetések történtek az összes aktív évfolyamon, majd az adott évben befizetett díjakból kivonjuk a károsszeget, ami kiadja az éves bontásban vett profitot. Láthatjuk, hogy az első négy év során több, mint 10 millió Forint halmozódott fel. Csupán egy év esetén negatív a mérleg, és ekkor sem éri el az 1 millió Ft-os értéket. Amennyiben elosztjuk az összegzett profitot a összes bevétellel, akkor megkapjuk hogy összességében $0.234832 \approx 23.5\%$-os növekedésre számíthatunk a 100 év során.

Most pedig nézzünk teljes képet az alakuló pénzügyekről, tehát a ténylegesen a *StudIn Biztosító*hoz kerülő pénzösszegekről. A (6) ábrán az előrejelzés szerint az első négy év megugrása után beáll egy trend, és egy lineáris függvénnyel leírható a tartalékképzés. Látszólag ez egy monoton növekvő folyamat, mely nem mutat durva kilengéseket. Ennek eredményeként a biztonsági költségtartalékunk összege folyamatos növekedést mutat az elkövetkező 100 évre.



5. ábra. Évenkénti profit millió Ft-ban a szimuláció 100 éve során.



6. ábra. A 100 évre vetített tartalékok alakulása millió forintban.

4. Konklúziók

A dolgozatom során igyekeztem bemutatni és modellezni a biztosítás egy olyan formáját, mely jelenleg nem létezik a piacon. A matematika Bsc szakos hallgatók számára nyújtott pénzügyi támogatás sok esetben segítené azon diákokat, akik gyengébb teljesítménnyel végzik tanulmányaikat. Jelenleg ugyanis a képzésen nagyjából 40%-os a végbizonyítvány nélkül távozók aránya. Reményeink szerint az esetek többségében ezek a hallgatók maradásra bírhatóak lennének egy a modellben felépített, vagy ehhez hasonló biztosítás ellenében.

Láthattuk, hogy kevés információval és kis elemszámú mintával is lehetőség nyílt bizonyos statisztikai ismérvek, trendek megismerésére. A modell készítése során több helyen is becslésekre, feltevésekre szorítkoztunk, melyek pontosítására a jövőben adottak különböző módszerek. Például kérdőívekkel, piackutatással érdemes lenne felmérni a diákok igényeit, és a biztosítás lehetőségét. Kérdés, hogy a hallgatók mennyire érzik ténylegesen szükségszerűnek egy ilyen típusú biztosítás meglétét? Mekkora összeget lennének hajlandóak félévente fizetni azért, hogy be legyenek biztosítva további támogatott félévek számukra? Emellett egyeztetni kell a biztosításfelügyelettel a lefektetett szabályokat.

Továbbá érdemes megfigyelnünk a felépített modell gyenge pontjait. Hol tévedhetünk, milyen paraméteren kéne finomítani, mely feltevéseinket kell pontosítani a biztosabb előrejelzés érdekében.

A szimuláció során alapvetően pozitív eredményt kaptunk, hiszen kellően óvatos becslést készítettünk a kockázatra, mégis pozitív mérleget jósoltunk a biztosítónak. A számításaink szerint sikerült teljesíteni a 20%-os profitot, ami a kezdetektől fogva elvárás volt. Ez reményeink szerint elég lehet arra is, hogy a gyakorlatban a modelltől jelentkező eltérésekre is fedezetet biztosítson.

A modell számunkra kedvező eredményt hozott, a kapott díjak megfizethetőnek tűnnek. Mégis azt érezzük, hogy egy ilyen összetett vállalkozás esetén elengedhetetlen a precízió. Ennek fényében évről évre szükséges az aktuális adatok elemzése, ezzel is biztosítva a modell megfelelő működését, és eredményes jövőt a *StudIn Biztosító*nak.

Hivatkozások

[1] Arató Miklós. *Nem-életbiztosítási matematika.* Egyetemi tankönyvkiadó, Budapest, 2001.

[2] Mogyoródi József. *Biztosítási matematika 1.* [*A biztosítási díjkalkuláció matematikai elveiről*]. Az Eötvös Loránd Tudományegyetem Természettudományi Kar Valószínségelmélet és Statisztika Tanszéke és az Állami Biztosító kiadványsorozata

[3] Olaf Mersmann. Package 'truncnorm'.
`https://cran.r-project.org/web/packages/truncnorm/truncnorm.pdf`

[4] Kiss Emil. BSc Matematika alapszak Tantervi háló - 2013.
`http://ewkiss.web.elte.hu/html/ujbsc/ujhonlap/BSc_Mat_halo_2013.html`

[5] Varga László gyakorlati jegyzete, Idősorok és többdimenziós statisztikai módszerek - 2018.
`http://vargal4.elte.hu/Elmelet_IdosStat18.pdf`

[6] Márkus László előadás fóliái, Idősorok és többdimenziós statisztikai módszerek.
`http://web.cs.elte.hu/probability/markus/ElemzoIdosor.html`

[7] Peter D. England, *Financial Simulation Models In General Insurance.* 2003
`https://www.actuariesindia.org/downloads/gcadata/5thGCA/Financial%20Simulation%20Models%20In%20General%20Insurance.pdf`

[8] ELTE TTK Tanulmányi Osztály, *Költségtérítés*
`http://to.ttk.elte.hu/koltsegterites`

[9] Magyar Nemzeti Bank, *A bonus-malus besorolási rendszer*
`https://www.mnb.hu/fogyasztovedelem/biztositasok/gepjarmu-biztositas/a-bonus-malus-besorolasi-rendszer`

[10] Becsky-Nagy Patrícia – Fazekas Balázs, *Speciális kockázatok és kockázatkezelés a kockázatitőke finanszírozásban*
`http://unipub.lib.uni-corvinus.hu/1965/1/vt2015n3p57.pdf`

Ábrák jegyzéke

1. A félévek számának eloszlása a 2013-ban felvett, és a szakot azóta elvégzett hallgatók statisztikái alapján. 17

2. Az első féléves tanulmányi átlagok, és bukások száma hallgatónként a *Vizsga data.frame* objektumból. Piros szaggatott vonallal az algoritmusban használt határvonalak láthatóak, a színek a kezdeti besorolás osztályait jelölik. 22

3. Szimuláció és tapasztalati eloszlás összehasonlítása. 32

4. A szimuláció során a túlcsúszó hallgatók számának előfordulásai. 33

5. Évenkénti profit millió Ft-ban a szimuláció 100 éve során. 37

6. A 100 évre vetített tartalékok alakulása millió forintban. 37