A MEDRANK ALGORITMUS KORLÁTAI[1]

BEDNAY DEZSŐ[a] – FLEINER BALÁZS[b] – TASNÁDI ATTILA[c]
[a,c]*Budapesti Corvinus Egyetem* – [b]*Budapesti Gazdasági Egyetem*

A szavazási eljárások távolságminimalizálási problémák megoldásaiként is levezethetők. A MedRank algoritmus meglehetősen erős feltételek mellett minimalizálja a Spearman-féle colstok távolságot. Rávilágítunk ennek az eredménynek a korlátaira, és megvizsgáljuk a MedRank algoritmus egy finomítását is. Emellett megmutatjuk, hogy az analóg probléma nem merül fel a Spearman-féle rangkorreláció minimalizálásakor, ami a Borda-szavazást eredményezi.

Kulcsszavak: szavazási eljárások, Spearman-féle rangkorreláció, Spearman-féle colstok. *JEL kód:* D71

1 Bevezetés

Ebben a tanulmányban az „operációkutatási megközelítést” követjük, amely arra törekszik, hogy a szavazási eljárásokat, amelyeket társadalmi választási szabályoknak (TVSZ) is neveznek, megfelelően meghatározott távolságminimalizálási problémák megoldásaként válassza ki. Valójában a Kemény-Young eljárás (Kemény, 1959) a Kendall τ távolságot minimalizáló szavazási eljárásként definiált, míg a Borda-szavazás a Spearman-féle rangkorrelációs távolságot minimalizáló eljárásként, a MedRank algoritmus pedig a Spearman-féle colstok távolságot[2] minimalizáló eljárásként adódnak (Dwork et al., 2002). Az utóbbi két eredmény korlátait szintén Dwork et al. (2002) tartalmazza. Nevezetesen bebizonyították, hogy a Borda-szavazást bármely olyan preferenciaprofilra megkapjuk, amelyre a Borda-szavazás lineáris rendezést ad, míg a MedRank algoritmust akkor kapjuk, ha az alternatívák medián helyezései permutációt alkotnak. Megmutatjuk, hogy míg az első esetben a korlát hatása elhanyagolható, mivel a szavazók számának növekedésével a holtversenyek valószínűsége tetszőlegesen kicsivé tehető, addig azonos medián helyezések egyhez tetszőlegesen közeli valószínűséggel fordulnak elő.

A szavazási eljárások levezetésének mint a kívánt tulajdonságokkal rendelkező optimalizálási problémák megoldásai, gazdag a szakirodalma. Az irodalom jelentős része társadalmi választási függvényekkel (TVF) foglalkozik, amely csak egy győztest választ ki, nem pedig a teljes társadalmi rangsort. Farkas és Nitzan (1979) megkapta a Borda-szavazást az egyhangúság elvétől való távolság minimalizálásával. Nitzan (1981) más mérőszámokat figyelembe

[1] Beérkezett 2024. szeptember 8. DOI: https://doi.org/10.15170/SZIGMA.55.1236. E-mail: dezso.bednay@uni-corvinus.hu.

[2] A Spearman footrule-nak nincs a statisztikai irodalomban magyar nyelvű fordítása, mi Spearman-féle colstok távolságnak fordítottuk.

véve, más eljárások mellett a többségi szavazást vezette le. A kapott szavazási eljárások azonban a választott távolságfüggvények és elvek függvényei. Lehrer és Nitzan (1985), valamint Campbell és Nitzan (1986) megmutatták, hogy alapvetően bármely szavazási eljárás racionalizálható megfelelő távolságfüggvénnyel. Az egyhangú győztes, a többségi győztes vagy a Condorcet-győztes profilok halmazától való távolság minimalizálásának megközelítését többek között Elkind et al. (2015), Andiga et al. (2014) és Mahajne et al. (2015), valamint Zwicker (2014) fejlesztette tovább. Bednay et al. (2017) a legközelebbi diktátortól való távolságmaximalizáláson alapuló ,,duális" megközelítést alkalmazták, amely a TVF-ek és a diktatórikus szabályok közötti távolságot veszi alapul. A legközelebbi diktatórikus szabálytól legtávolabb eső szavazási szabályokat vizsgálva nagyon nem kívánatos szavazási eljárások adódnak. Ezt a megközelítést folytatva Bednay et al. (2019) bevezettek egy diktátormentességi indexet, amely segítségével több TVF-t hasonlítottak össze. Emellett Bednay et al. (2022) meghatározták a diktátormentességi index határértékét is.

Ebben a tanulmányban az operációkutatásban használt két, kiemelt jelentőségű távolságfüggvényt vizsgálunk, amelyek relevánsak a szavazáselmélet kontextusában. Ezek a távolságfüggvények jól ismertek a matematikában és a statisztikában. Diaconis és Graham (1977) permutációkra vonatkozó távolságként tekintett rájuk, a statisztikában pedig ordinális adatokra vonatkozó asszociációs kapcsolatok mértékeként használják őket (lásd Monjardet, 1997). A Spearman-féle rangkorreláció esetében a kapcsolódó optimális szavazási eljárás a Borda-szavazás, míg a Spearman-féle colstok tekintetében a MedRank algoritmust szokták optimumként említeni. Amint azonban az eredményeinkből kiderül, ez utóbbi csak a lehetséges rangsor-kombinációk kis halmazára igaz. A Bucklin-eljárást mint a MedRank algoritmus egy lehetséges finomítását szinten bevonjuk az elemzéseinkbe.

Történelmi perspektívából a Bucklinról elnevezett medián helyezéseken alapuló szavazási eljárást (a tanulmányunkban leírt formában) először 1909-től 1922-ig alkalmazták a Colorado állambeli Grand Junctionban. Az eljárást az Egyesült Államok összesen 55 városában alkalmazták. A FED igazgatóinak a megválasztásához is használták az eljárást. Az eljárást Condorcet már 1793-ban javasolta és a XVIII. században Genovában alkalmazták. A történelmi részleteket illetően lásd például Lagerspetz (2016, 65.oldal). A matematika és a statisztika tárgyak oktatása során tárgyalt elemi összefüggés, hogy az átlagos abszolút eltérést a median minimalizálja, így nem meglepő, sőt mondhatni triviális Dwork et al. (2002) első bekezdésben ismertetett MedRank algoritmusra vonatkozó eredménye. Érdemes megemlíteni magyar vonatkozású eredményként, hogy a kémia területén Héberger (2010) rangszámkülönbségek összegét (SRD) alkalmazva vezet be rangsoroló eljárást különböző módszerek összehasonlítására, illetve ötvözésére, amelyeket más statisztikai eljárásoknál előnyösebbnek talál. Az eljárása a Spearman-féle colstok távolsággal rokon. Sziklai és Héberger (2020) az SRD módszert a mandátumszámítási eljárások összehasonlítására és a választókerület-szabdalási problémára használják. Lin (2010) a rangszám összegző módszereket elsősorban biológiai és internetes alkalmazások által motiválva tárgyalja. Egy másik részben magyar vonatkozású eredményben Erdélyi et al. (2015) a Bucklin-eljárás tágabb értelemben vett manipulálhatóságának bonyolultságelméleti kérdéseivel foglalkozik.

A tanulmány felépítése a következő. A 2. szakasz bemutatja az alapvető jelöléseket és az alkalmazott metrikákat a TVSZ-ek halmazán. A 3. szakasz bemutatja a vizsgált TVSZ-eket. A 4. szakasz bemutatja a szimulációs eredményeinket. Az 5. szakasz rámutat a MedRank algoritmus korlátjára. A 6. szakasz a Borda-szavazással foglalkozik. Végezetül a 7. szakasz záró gondolatokat tartalmaz. További ábrák a függelékben találhatók.

2 Modellkeret

Legyen $A = \{a_1, \ldots, a_m\}$ az alternatívák halmaza, ahol $m \geq 2$ és $N = \{1, \ldots, n\}$ a szavazók halmaza. Jelöljük $\mathcal{P}$-vel az A feletti összes lineáris (vagy szigorú preferencia) rendezések halmazát, és $\mathcal{P}^n$-nel az összes preferenciaprofil halmazát. Ha $\succ \in \mathcal{P}^n$ és $i \in N$, akkor $\succ_i$ az i szavazó preferenciarendezése A felett. Ha a $\succ$ profilból elhagyjuk az i szavazó preferenciarendezését, akkor ezt $\succ_{-i}$-vel jelöljük. Ekkor $\succ = (\succ_i, \succ_{-i})$. Jelölje továbbá $rk[a, \succ]$ az a alternativa helyezését a $\succ \in \mathcal{P}$ rangsorban (azaz $rk[a, \succ] = 1$, ha a a legjobb alternatíva a $\succ$ rangsorban, $rk[a, \succ] = 2$, ha a a második legjobb, és így tovább).

2.1 definíció. Az $f : \mathcal{P}^n \to \mathcal{P}$ leképezést társadalmi választási szabálynak nevezzük (a továbbiakban TVSZ).

Egy TVSZ egy preferenciaprofilhoz választja ki a (társadalmi) lineáris rendezést. Mivel a TVSZ definíciónk nem engedi meg a lehetséges döntetleneket, míg a jól ismert szavazási eljárásokat meghatározó formulák csak gyenge sorrendeket határoznak meg, a döntetlenek feloldására rögzített anonim törési szabályokat alkalmazunk.[3] A $\tau : \mathcal{P}^n \to \mathcal{P}$ törési szabály preferenciaprofilokhoz A feletti lineáris rendezéseket határoz meg. Ha egy bizonyos pozícióban több közömbös alternatíva van egy TVSZ-t „majdnem” meghatározó formula alapján, akkor a közömbös alternatívák a törési szabály szerinti sorrendet követik. A vizsgált TVSZ-ek indexeinek meghatározásakor $a_1 \, \tau(\succ) \, a_2 \, \tau(\succ) \ldots \tau(\succ) \, a_m$ fix (más néven lexikografikus) törési szabályt alkalmazzuk minden egyes $\succ \in \mathcal{P}^n$ profilra, amely anonim és nem függ a tényleges $\succ \in \mathcal{P}^n$ preferenciaprofiltól. A továbbiakban röviden $a_1 \, \tau \, a_2 \, \tau \ldots \tau \, a_m$-et fogunk írni az alternatívák e törési szabály szerinti sorrendjére.

Legyen $\mathcal{F} = \mathcal{P}^{\mathcal{P}^n}$ a TVSZ-ek halmaza és $\mathcal{F}^{an} \subset \mathcal{F}$ az anonim TVSZ-ek halmaza. Az $\mathcal{F}$ diktatórikus eljárásokból álló részhalmazát $\mathcal{D} = \{D_1, \ldots, D_n\}$ jelöli, ahol D_i az a diktatórikus szabály, amelyben i szavazó a diktátor. $\mathcal{F}$-en egy metrikát és egy kvázimetrikát fogunk vizsgálni, amelyeket bármely $F, G \in \mathcal{F}$-re a

[3] Az anonim törési szabály által kiválasztott lineáris sorrend invariáns a szavazók preferenciáinak sorrendjére.

$$\rho_1(F,G) = \frac{1}{(m!)^n} \sum_{\succ \in \mathcal{P}^n} \sum_{i=1}^{m} |rk[a_i, F(\succ)] - rk[a_i, G(\succ)]| \tag{1}$$

és

$$\rho_2(F,G) = \frac{1}{(m!)^n} \sum_{\succ \in \mathcal{P}^n} \sum_{i=1}^{m} (rk[a_i, F(\succ)] - rk[a_i, G(\succ)])^2 \tag{2}$$

képletek segítségével definiálunk. Megjegyzendő, hogy az ilyen típusú távolságokat általában profilokra határozzák meg, és így a külső összeg hiányzik a fenti egyenletekből, amikor a távolságokat preferenciákra vagy permutációkra határozzák meg. Mivel minket a TVSZ-ek összehasonlítása érdekel, a TVSZ-ek közötti átlagos távolságot értelmeztük. Egy adott profil két szavazójára (azaz két lineáris rendezés távolságát figyelembe véve) az (1) és a (2) belső összegei által meghatározott távolságok a szakirodalomban Spearman-féle colstok és Spearman-féle rangkorreláció néven ismertek. Természetesen más súlyozási sémák is lehetségesek, itt azonban minden profilnak azonos súlyt adunk.

Segítségünkre lesz, hogy Diaconis és Graham (1977) a Spearman-féle colstok távolságra és a Spearman-féle rangkorrelációra két lineáris rendezés közötti maximális távolságot meghatározta, amelyek értékei rendre $\lfloor m^2/2 \rfloor$ és $(m^3 - m)/3$. Ezért a 2.2 definícióban a normalizálás érdekében legyen

$$C_1 = \lfloor m^2/2 \rfloor \quad \text{és} \quad C_2 = (m^3 - m)/3.$$

Megjegyzendő, hogy a fenti maximális értékek két ellentétes sorrend mellett vétetnek fel.

A normalizált indexünk értelmezésében az összes diktatórikus szabályt azonos súllyal vesszük figyelembe, ami azt jelenti, hogy a tőlük mért távolságok összegét vesszük az indexek kiszámításakor.

2.2 definíció. A kiegyensúlyozottsági index a

$$BI_K(F) = \frac{1}{n} \frac{\sum_{i \in N} \rho_K(F, D_i)}{C_K},$$

ahol $K \in \{1, 2\}$.

3 Néhány társadalmi választási szabály

Három szavazási eljárás szükséges az elemezésünkhöz. Ezen TVSZ-ek bemutatásához az *1. táblázatban* megadott 5 alternatívát és 7 szavazót tartalmazó profilt választjuk. Mivel a TVSZ-t részben meghatározó képlet alapján lehetnek döntetlen alternatívák, az $a \tau b \tau c \tau d \tau e$ rögzített törési szabállyal oldjuk fel a döntetleneket.

Helyezés	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$
1	e	d	a	e	d	b	b
2	a	a	d	c	a	e	e
3	d	c	b	d	e	d	d
4	c	e	c	b	c	c	a
5	b	b	e	a	b	a	c

1. táblázat. Egy profil 5 alternatívával és 7 szavazóval

1. A Borda-szavazás, röviden BC-vel jelölve, az alternatívákat a helyezéseik összege alapján rendezi. Az 1. táblázatban szereplő profil esetében az a, b, c, d és e alternatívák helyezéseinek összege 21, 24, 26, 16 és 18. Ezért a BC által meghatározott társadalmi sorrend $d \succ e \succ a \succ b \succ c$. A BC_τ TVSZ a Borda-szavazás, ha minden $(\succ_i)_{i=1}^n \in \mathcal{P}^n$ profilra és minden a és b különböző alternatívapárra

$$a\ BC_\tau\left((\succ_i)_{i=1}^n\right)\ b \quad\Leftrightarrow\quad \sum_{i=1}^n rk[a,\succ_i] < \sum_{i=1}^n rk[b,\succ_i] \quad \text{vagy}$$
$$\left(\sum_{i=1}^n rk[a,\succ_i] = \sum_{i=1}^n rk[b,\succ_i] \quad \text{és} \quad a\tau b\right).$$

2. A Bucklin-eljárás minden a alternatíva esetében meghatározza azt a legjobb h_a helyezést, amelyet még figyelembe kell venni, hogy az első h_a helyen rangsorolt alternatívát figyelembe véve több mint $n/2$-szer szerepeljen egy adott profilban. Az 1. táblázatot megvizsgálva azt látjuk, hogy egyetlen alternatíva sem kap többséget (azaz 4 szavazatot), ha csak a legmagasabban rangsorolt alternatívák számát vesszük figyelembe. Ha most a második alternatívákat is figyelembe vesszük, akkor azt látjuk, hogy a és e is négyszer szerepel, ahol a τ törési szabály az a-t részesíti előnyben. Így $h_a = h_e = 2$. Ha a harmadik alternatívákat is figyelembe vesszük, akkor d hatszor jelenik meg az első három sorban, és $h_d = 3$. Ha a negyedik helyen szereplő alternatívákat is figyelembe vesszük, akkor c hat, b pedig négy alkalommal jelenik meg. Ezért $h_b = h_c = 4$. A h_b és h_c értékek egyenlősége esetén a nagyobb számú előfordulással rendelkező alternatíva élvez elsőbbséget, és ha még ezek is azonosak, akkor a τ törési szabályt kell alkalmazni. Ekkor a Bucklin-eljárás az $a \succ e \succ d \succ c \succ b$ rangsort adja. Minden egyes a alternatívára a h_a helyezést az a helyezéseinek mediánja határozza meg. A Bucklin-eljárást BR-rel jelöljük. Formálisan a BR_τ a Bucklin-eljárás, ha minden $\succ \in \mathcal{P}^n$ profilra és minden egymástól eltérő a és b alternatívapárra

$$aBR_\tau(\succ)b \;\Leftrightarrow$$
$$(h_a < h_b) \text{ vagy}$$
$$(h_a = h_b \text{ és } \#\{i \in N \mid rk[a,\succ_i] \le h_a\} > \#\{i \in N \mid rk[b,\succ_i] \le h_b\}) \text{ vagy}$$
$$(h_a = h_b \text{ és } \#\{i \in N \mid rk[a,\succ_i] \le h_a\} = \#\{i \in N \mid rk[b,\succ_i] \le h_b\} \text{ és } a\tau b).$$

3. A Bucklin-eljárásnak több változata van, de van egy rokon változata is, amelyet az informatikai szakirodalomban MedRank algoritmusnak neveznek.

Mint az alternatívák medián helyezésein alapuló aggregációs szabály rövidítéseként, a továbbiakban MR-ként hivatkozunk rá. Az eltérés abban rejlik, hogy csak az alternatívák medián helyezései számítanak, azaz csak a fentebb definiált $h_{a_1},\ldots,h_{a_n}$ értékek (lásd például Dwork et al., 2002). A mi TVSZ keretünkben két különböző a és b alternatíva esetén, amelyek esetében $h_a = h_b$, a törési szabályt egyből kell alkalmazni, anélkül, hogy figyelembe vennénk, hogy ez a két alternatíva hányszor szerepel az összes szavazó első h_a helyein. Ezért a MedRank algoritmus kevésbé determinált, mint a Bucklin-eljárás, és valójában gyakrabban vannak holtversenyes alternatívák. Az 1. táblázatban szereplő preferenciaprofilra az elején említett törési szabállyal a MedRank algoritmus az $a \succ e \succ d \succ b \succ c$ rangsort adja, amely eltér a Bucklin-eljárás eredményétől. Dwork et al. (2002) kimutatta, hogy legalábbis azon profilok esetében, amelyekben az alternatívák összes medián rangja páronként különböző (azaz permutációt alkotnak), a MedRank algoritmus által kapott rangsor megegyezik a Spearman-féle colstok optimális rangsorral.

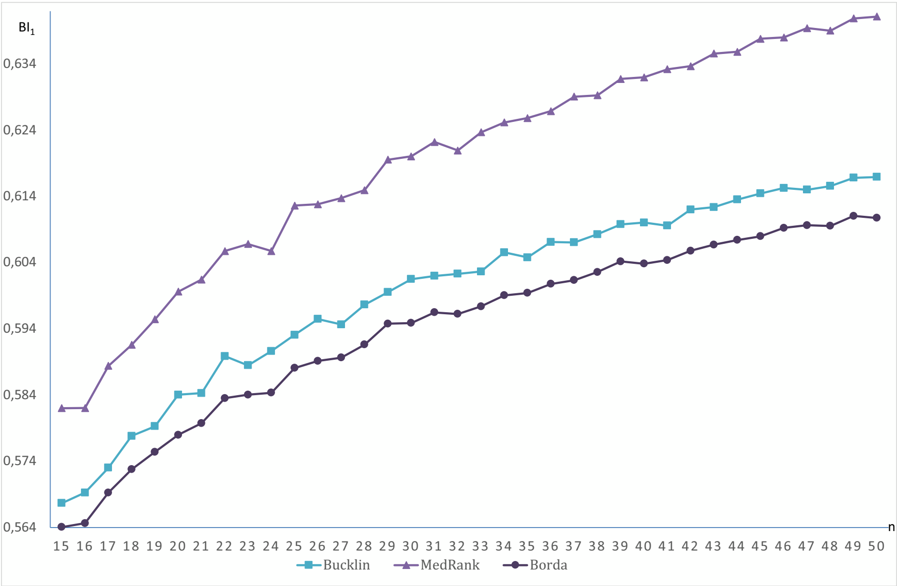
4 Szimulációk

Három alternatíva és legfeljebb 7 szavazó esetén a Spearman-féle colstok távolsághoz tartozó kiegyensúlyozottsági indexek pontos értékeit teljes leszámlálással határoztuk meg. Ezek az eredmények a 2. táblázatban találhatók. Több alternatíva és több szavazó esetén az indexek becslését 3, 4, 5 és 6 alternatíva és 3-tól 100 szavazóig végeztük el. Szimulációinkban 2500 véletlenszerű preferenciaprofilt generáltunk, ahol minden profil egyenlő valószínűséggel került kiválasztásra. A 3, 4 és 6 alternatívára vonatkozó grafikonok a függelékben találhatók. A lényegre összpontosítva, ebben a szakaszban csak az 5 alternatívás eset grafikonjait mutatjuk be. Minőségi szempontból 3, 4 és 6 alternatíva esetén ugyanarra a következtetésre jutunk. A táblázatban szereplő pontos értékek közel állnak a szimulációk útján kapott értékekhez három alternatíva esetén (lásd a függeléket is). Megfigyelhetjük, hogy legalább 4 szavazó esetén a MedRank algoritmus a másik két szavazási szabálynál rosszabbul teljesít, mivel a kisebb BI_1 értékek jobbak.

Az 1. ábrán tapasztalható lassú konvergenciasebesség és a három grafikon meredeksége miatt kis számú szavazó esetén csak a 15 és 50 szavazó közötti tartományban ábrázoljuk a grafikonokat. Az 1. ábra kiemeli, hogy a MedRank algoritmus a másik két szavazási eljáráshoz képest gyengébben teljesít. Még meglepőbb azonban, hogy a Borda-szavazás felülmúlja a Bucklin-eljárást, amely a MedRank algoritmus egyfajta finomítása.

TVSZ $\setminus n$	3	4	5	6	7
MedRank	0.370370	0.449074	0.464506	0.496914	0.513489
Borda	0.384259	0.439815	0.460648	0.482017	0.495842
Bucklin	0.370370	0.449074	0.453704	0.494663	0.493377

2. táblázat. Pontos értékek BI_1-re és $m = 3$-ra



1. ábra. Kiegyensúlyozottság BI_1 és $m = 5$ esetén

Ez ellentétben áll azzal az eredménnyel, hogy a MedRank algoritmus minimalizálja a Spearman-féle colstok távolságot azoknál a profiloknál, amelyeknél az alternatívák medián helyezései a helyezések permutációját alkotják (lásd Dwork et al., 2002). Könnyen belátható, hogy a profilok jelentős hányada nem felel meg ennek a követelménynek. Például az első helyezés mint egy alternatíva medián helyezése akkor és csak akkor jelenik meg egy profilban, ha valamely alternatíva esetén az első helyezés szigorú többségben van. Ennek a valószínűsége m alternatíva esetén

$$m \sum_{k=\lceil (n+1)/2 \rceil}^{n} \binom{n}{k} \left(\frac{1}{m}\right)^{k} \left(\frac{m-1}{m}\right)^{n-k}.$$

Például $m = 3$ és $n = 10$ esetén a fenti valószínűség megközelítőleg 0,229691. Sok profil nem felel meg ennek a kritériumnak, és lehetnek még más hiányzó helyezések is, továbbá nemcsak egy rögzített alternatíva esetén fordulhat ez elő. A számpélda is mutatja, hogy kevés szavazó esetén is nagy valószínűséggel – kizárólag az alternatívák medián helyezéseit véve – nem kapunk lineáris rendezést.

5 A MedRank algoritmus elemzése

Amint azt a 4. szakaszban a BI_1 indexek grafikonjaiból láttuk, a MedRank algoritmus a legtöbb profil esetében nem a Spearman-féle colstok optimális szabály. Ebben a szakaszban megmutatjuk ennek okát: a MedRank algoritmus esetében a döntetlenek valószínűsége egyhez tart, ha a szavazók számával a végtelenbe tartunk, sőt minden alternatíva rangszáma ugyanaz lesz páratlan sok alternatíva esetén, és ,,az egyik medián" rangszám lesz páros sok alternatíva esetén.

A MedRank*, a továbbiakban röviden MR^* algoritmus minden egyes a alternatívához hozzárendeli annak a 3. szakaszban értelmezett h_a 'medián helyezését', amely páros sok alternatíva esetén két különböző 'középső helyezés' közül a nagyobbikat, illetve – másképpen fogalmazva – a rosszabbikat jelenti. Tehát az MR^* a $\mathcal{P}^n$-ből az $\{1,\dots,m\}^A$-ra képez, és nem TVSZ, mivel gyenge preferenciákat eredményezhet. Ráadásul helyezéseket rendel az alternatívákhoz, és nem csak azok gyenge sorrendjét.[4] Másképpen fogalmazva, az MR^* algoritmus az MR algoritmus holtversenyek feloldása előtti helyezéseit rendeli az egyes alternatívákhoz. Mindazonáltal a $BI_1(MR^*)$ továbbra is ugyanúgy definiálható, mint a TVSZ-ek esetében. Továbbá, mivel bármely $j=1,\dots,m$ és bármely $\succ\in\mathcal{P}^n$ esetén a

$$\min_{r=1,\dots,m}\sum_{i\in N}|r-rk[a_j,\succ_i]|, \tag{3}$$

megoldása az $rk[a_j,\succ_1],\dots,rk[a_j,\succ_n]$ helyezések mediánja (az abszolút különbségek összegének minimumára vonatkozó standard tulajdonság alapján). Megjegyezzük, hogy bármely j és bármely $\succ$ esetén független minimalizálási problémáink vannak, mivel adott j-re és $\succ$ profilra a (3) minimumhelyére $rk[a_j,MR^*(\succ)]$,[5] és ezért

$$\begin{aligned}\sum_{\succ\in\mathcal{P}^n}\sum_{j=1}^{m}\min_{r=1,\dots,m}\sum_{i\in N}|r-rk[a_j,\succ_i]| &= \sum_{\succ\in\mathcal{P}^n}\sum_{j=1}^{m}\sum_{i\in N}|rk[a_j,MR^*(\succ)]-rk[a_j,\succ_i]|\\ &\le \sum_{\succ\in\mathcal{P}^n}\sum_{j=1}^{m}\sum_{i\in N}|rk[a_j,F(\succ)]-rk[a_j,\succ_i]|,\end{aligned}$$

amiből az összegzés sorrendjének felcserélésével, és a konstansokkal való beszorzások után

$$BI_1(MR^*)\le BI_1(F)$$

bármely anonim F TVSZ esetén.

A következő lépésünk, hogy meghatározzuk az $MR^*(\succ)$ aszimptotikus eloszlását, ahogy n tart a végtelenbe. Válasszunk ki egy tetszőleges a alternatívát, és vegyük észre, hogy minden $\succ_l$-re , ahol $l=1,\dots,n$, az a alternatíva r/m valószínűséggel kerül az első r helyre. Ha adott n esetén $X_n^{(r)}$ az a-t az első r helyen rangsoroló szavazók száma, akkor $X_n^{(r)}$ binomiális eloszlású n és r/m paraméterekkel. Ezért annak a valószínűsége, hogy a az MR^* szerint az első r helyen szerepel

$$P(X_n^{(r)}\ge\lfloor n/2\rfloor+1)=\sum_{i=\lfloor n/2\rfloor+1}^{n}\binom{n}{i}\left(\frac{r}{m}\right)^{i}\left(1-\frac{r}{m}\right)^{n-i} \tag{4}$$

[4] Ez megengedőbb, mint a helyezések szokásos kiosztása, hiszen például az olimpiai játékokon lehet két aranyérmes, de nem lehet két ezüstérmes egy aranyérmes nélkül.

[5] Az rk kiterjesztése a_j helyezését határozza meg $MR^*(\succ)$ szerint.

és annak a valószínűsége, hogy a az MR^* szerint r-edik helyen szerepel

$$P(MR^* = r) = P(X_n^{(r)} \geq \lfloor n/2 \rfloor + 1) - P(X_n^{(r-1)} \geq \lfloor n/2 \rfloor + 1).$$

Emlékeztetünk Arratia és Gordon (1989, 1. tétel) binomiális eloszlásra vonatkozó eredményére, amelyet a (4) valószínűség felső korlátjának megadására fogunk használni. Legyen Y_n binomiális eloszlású n és p paraméterekkel. Ekkor bármely $n = 1, 2, \ldots$ és bármely $p < \alpha < 1$ esetén

$$P(Y_n \geq \alpha n) \leq e^{-n\left(\alpha \log \frac{\alpha}{p} + (1-\alpha) \log \frac{1-\alpha}{1-p}\right)}. \tag{5}$$

A mi esetünkben p az $1/m$, $2/m$,...,$(m-1)/m$ értékeket veszi fel, és szimmetriaokokból kifolyólag csak a (4)-ben szereplő valószínűségek felső korlátját kell megadnunk az $1/m$, $2/m$,...,$\lfloor m/2 \rfloor/m$ értékekre, mivel a (4)-ben szereplő valószínűségekre a komplementer esemény figyelembevételével megadhatjuk a szükséges alsó korlátokat az $(\lfloor m/2 \rfloor + 1)/m$,...,$(m-1)/m$ értékekre. A $p = 1/2$ esetre, ha m páros, közvetlenül fogjuk meghatározni (4)-et. Ezért a $p < \alpha = 1/2$ esetből indulunk ki. Ezeket az értékeket az (5)-be behelyettesítve

$$P(Y_n \geq \frac{1}{2}n) \leq e^{-\frac{1}{2}n\left(\log \frac{1}{2p} + \log \frac{1}{2(1-p)}\right)} = e^{-\frac{1}{2}n \log \frac{1}{4p(1-p)}},$$

ami viszont $p(1-p) < 1/4$ révén azt jelenti, hogy $P(Y_n \geq \frac{1}{2}n)$ nullához tart, ha n tart a végtelenbe. Ezért azt is megkapjuk, hogy páros m-re $r = 1, 2, \ldots, \lfloor (m-1)/2 \rfloor$ esetén és páratlan m-re $r = 1, 2, \ldots, \lfloor m/2 \rfloor$ esetén a (4)-ben szereplő valószínűségek nullához tartanak, ha n tart a végtelenbe. Ha m páros, azaz $m = 2k$, akkor

$$\sum_{i=\lfloor n/2 \rfloor + 1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{n-i} = \frac{1}{2^{n+1}} 2^n = \frac{1}{2}.$$

A határértékekre vonatkozó eredményeinkből és a már említett szimmetriai okokból (a komplementer eseményekre) arra következtetünk, hogy amikor n a végtelenbe tart, a (4) valószínűségek rendre 1/2-hez és 1-hez tartanak $r = m/2$ és $r = m/2 + 1$ esetén,[6] míg a medián(ok)-nál kisebb r-re és a medián(ok)-nál nagyobb r-re nullához tartanak, ha n tart a végtelenbe. Ezek alapján igazoltuk a következő tételt.

5.1 állítás. *Adott m mellett annak a valószínűsége, hogy az alternatívák mediánja $\lfloor (m+1)/2 \rfloor$ vagy $\lceil (m+1)/2 \rceil$, 1-hez tart, ha n tart a végtelenbe.*[7]

Az 5.1 állítás eredménye úgy is megfogalmazható, hogy 'végtelen sok szavazó' esetén, ha az alternatívák száma páratlan, akkor az összes alternatíva rangszáma $(m+1)/2$, míg ha az alternatívák száma páros, akkor csak a két középső rangszám fordul elő.

[6] Ezek az esetek csak akkor állnak fenn, ha m páros.

[7] Ha m páratlan, akkor a medián értéke egyértelmű, és a két megadott szám azonos.

Az 1. ábrából (és a függelékben a 3, 4 és 6 alternatívára vonatkozó ábrákból) látható, hogy a lexikografikus törési szabály mellett a Borda-szavazás BI_1 indexe alacsonyabb, mint a másik két vizsgált TVSZ-é.

E szakasz fő mondanivalója, hogy a legtöbb profil esetén a MedRank algoritmus nem minimalizálja a Spearman-féle colstok távolságot, mivel egy valószínűséggel vannak egyenlő medián helyezések.

6 Spearman-féle rangkorreláció – Borda-szavazás

Ebben a szakaszban megmutatjuk, hogy a Borda-szavazás és a Spearman-féle rangkorrelációt minimalizáló TVSZ közötti kapcsolatot a már említett korlátozó feltétel feloldása nem rontja el, mint azt a MedRank algoritmus és a Spearman-féle colstok távolságot minimalizáló TVSZ esetén láttuk az előző szakaszban. A következő tétel szerint a Borda-szavazás csak a profilok egyre kisebb hányadánál térhet el a Spearman-féle rangkorrelációt minimalizáló szavazási eljárástól.

6.1 állítás. *Annak valószínűsége, hogy egy további szavazó megváltoztatja a Borda-szavazással kapott rangsort, nullához tart, ha n tart a végtelenbe.*

Bizonyítás. Az egyszerűség kedvéért legyen $i = n$ a plusz szavazó, amely a BC anonimitása miatt az általánosság megszorítása nélkül feltehető. Megmutatjuk, hogy azon profilok aránya, amelyeken a $BC(\succ_{-n})$ nem egyenlő a $BC(\succ)$-vel, n végtelenbe tartásával a nullához tart. Ennek megmutatásához vegyünk két tetszőleges egymástól különböző a_j és a_k alternatívát. Tekintsük az a_j és a_k alternatívák helyezéseinek kumulált különbsége által meghatározott véletlen bolyongást. Konkrétan, tekintsük az

$$X_n = \sum_{i=1}^{n} (rk[a_j, \succ_i] - rk[a_k, \succ_i])$$

valószínűségi változót, ahol $n = 1, 2, \ldots$. Megjegyezzük, hogy $-m < X_n < m$ szükséges feltétele annak, hogy $a_j BC(\succ_{-n}) a_k$ és $a_k BC(\succ) a_j$ előállhasson. A Markov-folyamatnak tekinthető véletlen bolyongás állapottere $S = \mathbf{Z}$, a kiinduló állapot $X_0 = 0$, az átmenet-valószínűségek

$$p_{s,s+d} = P(X_{i+1} = s + d \mid X_i = s) = (m - d)/(m(m - 1)),$$

ha $d = 1, \ldots, m - 1$ és

$$p_{s,s-d} = P(X_{i+1} = s - d \mid X_i = s) = (m - d)/(m(m - 1)),$$

ha $d = 1, \ldots, m - 1$ minden $i = 0, 1, \ldots, n - 1$-re. Jelölje $p_{i,j}^{(t)}$ az i állapotból a j állapotba való t-lépéses átmenet-valószínűséget. A definiált véletlen bolyongás:

1. visszatérő, ami azt jelenti, hogy egy valószínűséggel valamikor visszatér a kiinduló állapotba, ami igazolható Billingsley (1995) 8.2.(ii) tételével, mivel $\sum_{t=1}^{\infty} p_{i,j}^{(t)} > \frac{1}{m} \sum_{t=1}^{\infty} \widetilde{p}_{i,j}^{(t)} = \infty$, ahol $\widetilde{p}_{i,j}^{(t)}$ a szimmetrikus Bernoulli-féle egyszerű bolyongás megfelelő átmenet-valószínűségeit jelöli,

2. irreducibilis, azaz minden állapot véges számú átmenettel elérhető minden más állapotból és

3. aperiodikus, mivel $P(X_i = 0) > 0$ minden $i \geq 2$-re.

Billingsley (1995) 8.8. tételéből tudjuk, hogy $\lim_{n\to\infty} p_{i,j} = 0$ minden i és j esetén. Ezért tetszőlegesen kicsi $\varepsilon > 0$ és elég nagy n esetén

$$\varepsilon/(m(m-1)) > P(-m < X_n < m)\,.$$

Most még az összes különböző alternatívapárt figyelembe véve is annak a valószínűsége, hogy egy n-edik szavazó hozzáadása megváltoztatja a BC eredményét, kisebb, mint ε, továbbá nullához tart, ha n tart a végtelenbe. □

Megfogalmazzuk a 6.1 tétel egy egyszerű következményét.

6.1 következmény. *Annak a valószínűsége, hogy BC-nek vannak holtversenyes alternatívái, vagy – másképpen fogalmazva – azonos rangsorösszegű alternatívái, nullához tart, ha n tart a végtelenbe.*

Dwork et al. (2002) kimutatták, hogy bármely $\succ \in \mathcal{P}^n$ esetén, amelyre a Borda-szavazás lineáris rendezést határoz meg, BC minimalizálja a Spearman-féle rangkorrelációt. Eredményüket és a 6.1 következményt figyelembe véve tudjuk, hogy bármely anonim F TVSZ esetén

$$\lim_{n\to\infty} BI_2(BC_\tau) \leq \lim_{n\to\infty} BI_2(F).$$

7 Záró gondolatok

Három alternatíva és legfeljebb hét szavazó esetén a kiegyensúlyozottsági indexet teljes leszámolással határoztuk meg, míg nagyobb számú szavazó és 3-6 alternatíva esetén 2500 profilból vettünk véletlenszerű mintát. A fő megállapításaink, hogy a profilok többsége esetén a MedRank algoritmus nem minimalizálja a Spearman-féle colstok távolságot, míg a Borda-szavazás „majdnem minden” profil esetében minimalizálja a Spearman-féle rangkorrelációt.

Végül felvázoljuk a lehetséges további kutatási irányokat. Elemzésünkben az alternatívák száma rögzített, míg a szavazók száma változó. Nyilvánvalóan ez az érdekesebb eset a szavazáselméleti kontextusban. Az informatikai alkalmazásokban azonban a másik eset is ugyanolyan érdekes lehet, amikor a szavazók számát rögzítjük, és az alternatívák száma a változónk. Mivel

m alternatíva esetén $m!$ különböző rangsor létezik, ez az elemzés sokkal nehezebbnek ígérkezik. Már a Kemény-Young eljárás meghatározása is NP-nehéz feladat (Bartoldi et al., 1989). Ezért ettől a kutatási iránytól kevesebb eredményt várunk.

Irodalom

1. Andjiga, N. G., A. Y. Mekuko & I. Moyouwou, (2014). Metric rationalization of social welfare functions, *Mathematical Social Sciences*, 72, 14–23.
2. Arratia, R. & L. Gordon, (1989). Tutorial on large deviations for the binomial distribution, *Bulletin of Mathematical Biology*, 51, 125–131.
3. Bartholdi, J., C. A. Tovey & M. A. Trick, (1989). Voting schemes for which it can be difficult to tell who won the election, *Social Choice and Welfare*, 6, 157–165.
4. Bednay, D., Fleiner, B. & Tasnádi, A. (2022). The limit of the non-dictatorship index, *Pure Mathematics and Applications*, 30, 28–33.
5. Bednay, D., Moskalenko, A. & Tasnádi, A. (2017). Does avoiding bad voting rules lead to the good ones? *Operations Research Letters*, 45, 448–451.
6. Bednay, D., Moskalenko, A. & Tasnádi, A. (2019). Dictatorship versus manipulability, *Mathematical Social Sciences*, 101, 72–76.
7. Billingsley, P. (1995). *Probability and measure.* Wiley, New York.
8. Dwork, C., Kumar, R., Naor, M. & Sivakumar, D. (2002). Rank aggregation revisited, http://www.cse.msu.edu/~cse960/Papers/games/rank.pdf, accessed: January 6, 2022.
9. Diaconis, P. & R. L. Graham (1977). Spearman's Footrule as a Measure of Disarray, *Journal of the Royal Statistical Society: Series B (Methodological)* 39, 262–268.
10. Elkind, E., P. Faliszewski & A. Slinko (2015). Distance rationalization of voting rules, *Social Choice and Welfare* 45, 345–377.
11. Erdélyi G., M. R. Fellows, J. Rothe & L. Schend (2015). Control complexity in Bucklin and fallback voting: A theoretical analysis, *Journal of Computer and System Sciences* 81, 632–660.
12. Farkas, D. & S. Nitzan (1979). The Borda rule and Pareto stability: A comment, *Econometrica* 47, 1305–1306.
13. Héberger K. (2010). Sum of ranking differences compares methods or models fairly, *Trends in Analytical Chemistry* 29, 101–109.
14. Lagerspetz, E. (2016). *Social Choice and Democratic Values*, Springer International Publishing, Switzerland.
15. Lehrer, E. & S. Nitzan (1985). Some general results on the metric rationalization for social decision rules, *Journal of Economic Theory* 37, 191–201.
16. Lin, S. (2010). Rank aggregation methods, *WIREs Computational Statistics* 2, 555–570.
17. Knuth, D. E. (1973). *Sorting and searching, The art of computer programming, Volume 3*, Reading MA: Addison-Wesley.
18. Mahajne, M., S. Nitzan & O. Volij (2015). Level r consensus and stable social choice, *Social Choice and Welfare* 45, 805–817.

19. Monjardet, B. (1997). Concordance between two linear orders: The Spearman and Kendall coefficients revisited, *Journal of Classification* 14, 269–295.

20. Nitzan, S. (1981). Some measures of closeness to unanimity and their implications, *Theory and Decision* 13, 129–138.

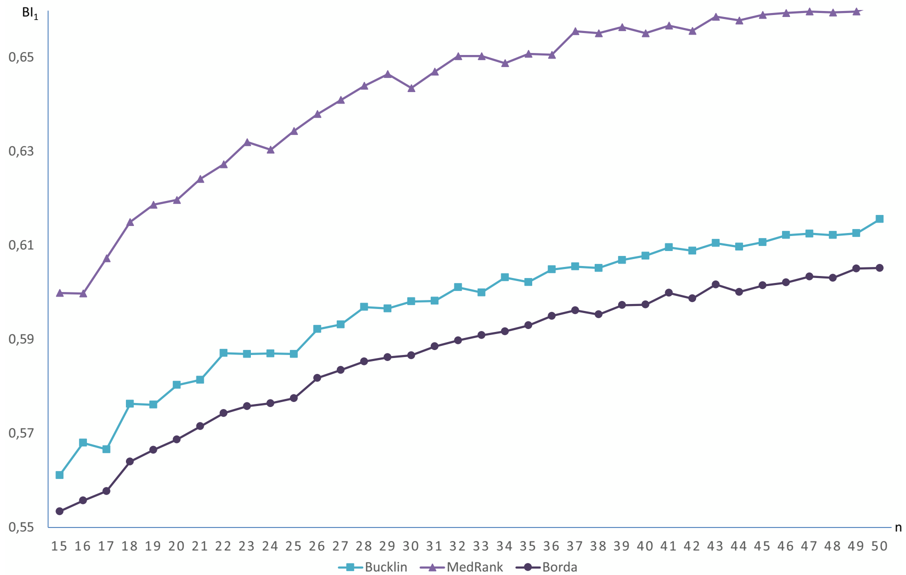
21. Nitzan, S. (1985). The vulnerability of point-voting schemes to preference variation and strategic manipulation, *Public Choice*, 47, 349–370.

22. Sziklai B. R. & Héberger K. (2020). Apportionment and districting by sum of ranking differences, *PLOS ONE*, 15, 1–20.

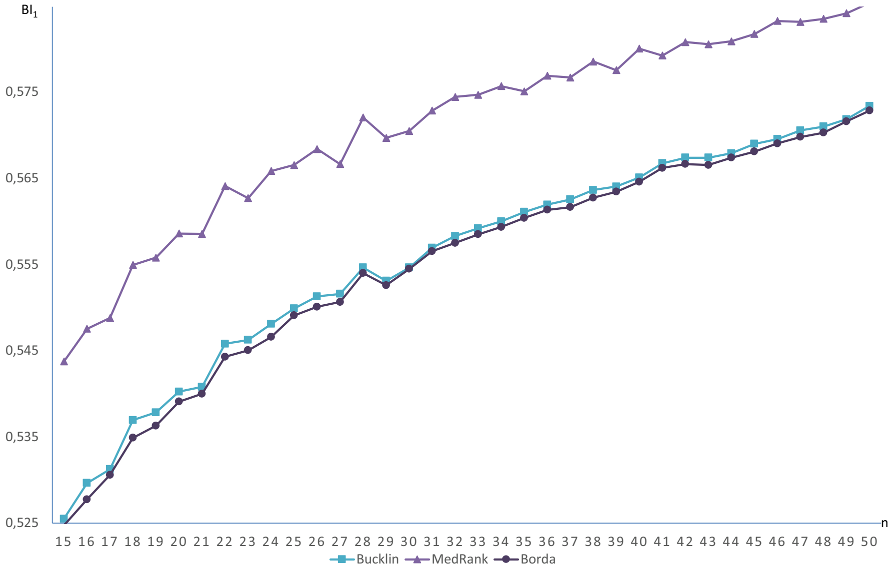
23. Zwicker, W. S. (2014). Universal and symmetric scoring rules for binary relations, *mimeo*.

Függelék

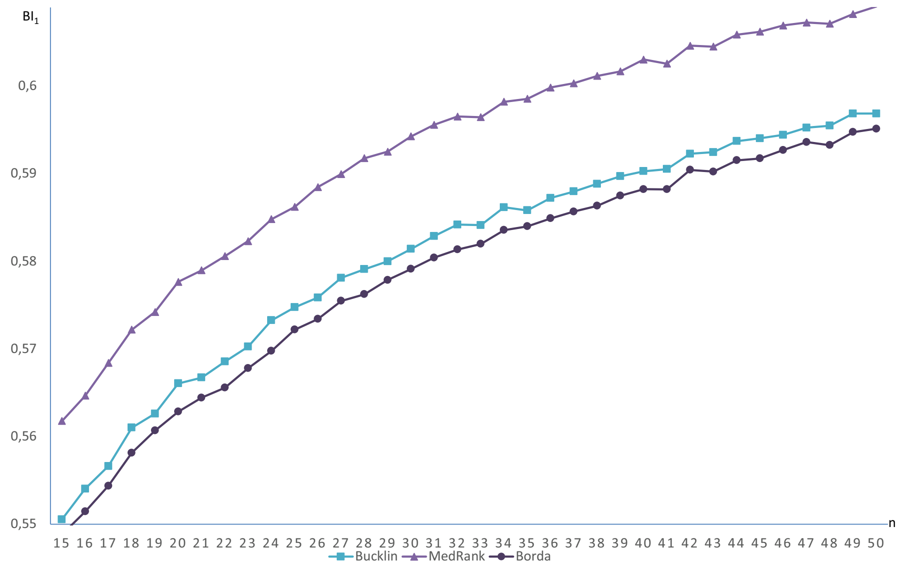
A *2–4. ábra* $m = 3$, $m = 4$ és $m = 6$ esetén ugyanazokat az adatokat tartalmazza, mint a 4. szakasz 1. ábrája $m = 5$ esetén. Mind a négy ábrán megfigyelhetjük, hogy a Borda-szavazás teljesít a legjobban, míg meglepő módon a MedRank algoritmus a legrosszabbul.



2. ábra. Kiegyensúlyozottság BI_1 és $m = 3$



3. ábra. Kiegyensúlyozottság BI_1 és $m = 4$



4. ábra. Kiegyensúlyozottság BI_1 és $m = 6$

LIMITATIONS OF THE MEDRANK ALGORITHM

Voting rules can be derived as distance minimization problems. Under quite restrictive conditions the MedRank algorithm minimizes the so-called Spearman footrule. We highlight the limitation of this result and also investigate the possibility of appropriate refinements of the MedRank algorithm. In addition, we show that the analogous problem does not arise when minimizing Spearman rank correlation, which results in the Borda count.