

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

PÁHÁN ANITA DALMA

Sportesemények kimenetelének előrejelzése

Szakdolgozat
Matematika Bsc

Témavezető:
JÜTTNER ALPÁR
Operációkutatási Tanszék



Budapest, 2025

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Jüttner Alpárnak, a rendszeres konzultációkat, az értékes tanácsokat, valamint a téma ajánlását, mely iránti lelkesedése rám is hamar átragadt. Továbbá szeretném megköszönni családomnak és barátaimnak is a sok támogatást.

Tartalomjegyzék

1	Bevezetés	4
2	Rangsorolási módszerek	5
2.1	Probléma bemutatása	5
2.2	Lineáris $p(x,y)$	8
2.3	Bradley–Terry-modell	9
2.3.1	Erősen összefüggő eset	14
2.3.2	A pontszámok kiszámítása	16
2.3.3	Gyorsabb konvergencia	19
2.3.4	Általános eset	20
2.4	Élő-pontozás	22
2.4.1	Konvergencia két játékosra	23
2.5	Több dimenziós pontszámok	27
3	Modellek kiértékelése	31
3.0.1	Valószínűségek kiértékelése	31
4	Alkalmazás	34

1 Bevezetés

Már az ókorban kötöttek fogadásokat gladiátorharcok és kocsihajtó versenyek győzte-seire. Napjainkban, amikor már számítógépünkön az elmúlt évek összes versenyének eredményét megtaláljuk, és az adatokat gyorsan fel is tudjuk dolgozni, felmerül a kérdés, hogy miként lehetne a mérkőzések eredményeit prediktálni ezek segítségével.

Az egyik megközelítés, hogy sorrendet alakítunk ki a versenyzők között úgy, hogy az a sportoló kerül előrébb, akit esélyesebbnek találunk a győzelemre. Ezután minden mérkőzésen arra tesszük a voksunk, aki hamarabb szerepel sorban. Ennél valamivel kifinomultabb módszer, ha olyan pontszámokat adunk a versenyzőknek, amiből valamilyen módon meghatározható az erőviszonyuk, azaz hogy milyen valószínűséggel nyer az egyik a másik ellen.

A továbbiakban megismertetünk néhány modellt, amikkel páronkénti összehasonlítások alapján meghatározhatjuk a játékosok erősségét. Ezután valós dzsúdómeccsek adatain leteszteljük a különböző predikciók teljesítményét. A modellek megválasztásához fontos tudni, hogy a dzsúdóban minden meccsen két sportoló küzd meg egymással, és nincs döntetlen. Az adatbázisunk közel 400000 meccset és 48000 játékost tartalmaz az elmúlt 20 évből, ezért elsősorban olyan módszereket vizsgálunk, amik kezelni tudják, hogy a versenyzőpárok többsége soha nem küzdött meg egymással, illetve nagy versenyzőszám esetén is viszonylag gyorsan meghatározzák a pontszámokat.

2 Rangsorolási módszerek

2.1 Probléma bemutatása

A célunk pontszámot adni a versenyzőknek a meccsek eredménye, illetve időpontja alapján. A pontszámoktól a következő három tulajdonságot várjuk el:

- A játékosok pontszáma alapján következtetni lehet a mérkőzés győztesére.
- A pontszámok gyorsan számolhatók, nagy versenyzőszám esetén is.
- A játékos pontszáma egy elvesztett meccs után sosem nő és soha nem csökken egy nyert meccs után. Ez azt jelenti, hogy soha senkinek nem éri meg veszteni egy mérkőzésen azért, mert ezzel nőne a pontszáma.

Az i . versenyző pontszámát jelöljük x_i -vel. Keresünk egy $p(x_i, x_j)$ függvényt, ami azt adja meg, hogy ha egy x_i és egy x_j pontszámmal rendelkező versenyző játszik egymással, akkor mekkora valószínűséggel győz az első. A p függvénynek teljesítenie kell néhány feltételt. Egyrészt $p(x, x) = 0.5$, hiszen két azonos pontszámú versenyző esetén azt várjuk, hogy 50% eséllyel nyer mindkettő. Másrészt $p(x, y) = 1 - p(y, x)$, hiszen a két versenyző közül az egyik biztosan nyer.

Megkövetelhetnénk még, hogy $0 \leq p(x, y) \leq 1$, hiszen $p(x, y)$ egy valószínűséget ad meg, azonban néha eltekintünk ettől, és így relaxáljuk a problémát.

Sportokban általában sokat számít a versenyző adott formája, ezért ha össze szeretnénk hasonlítani két embert, akik kétszer játszottak ezelőtt és ebből egy-egy meccset nyertek, az első gondolatunk, hogy megnézzük, ki nyert legutoljára. Mivel a versenyek adatbázisa sok évet ölel fel, így fontos valamilyen módon számításba vennünk az időpontokat. Minden k meccshez rendeljünk egy s_t súlyt, amit definiáljunk a következőképpen: $s_t = 0,5^{\frac{z-t}{365}}$, ahol $z - t$ a mérkőzéstől mostanáig eltelt idő napokban. Ezzel éppen azt érjük el, hogy a mérkőzés súlya évente (365 naponta) a felére csökken.

A versenyzők számát jelöljük n -nel. Első megközelítésként hozzunk létre egy $M \in \mathbb{R}^{n \times n}$ mátrixot az ismert mérkőzések alapján. Jelöljük M i . sorában és j . oszlopában lévő mezőt $M_{i,j}$ -nek. $M_{i,j}$ legyen az i játékos győzelmének relatív gyakorisága j ellen a

fent bemutatott súlyozással. Tehát ha $\mathcal{H}(i, j)$ azon meccsek halmaza, ahol i legyőzte j -t, akkor

$$M_{i,j} = \frac{\sum_{k \in \mathcal{H}(i,j)} s_k}{\sum_{k \in \mathcal{H}(i,j) \cup \mathcal{H}(j,i)} s_k}$$

Olyan a játékosok pontszámait tartalmazó $X = (x_1, x_2, \dots, x_n)$ vektort keresünk, amihez ha kiszámoljuk az $\widetilde{M}_{i,j} = p(x_i, x_j)$ mátrixot, az minél közelebb lesz az eredeti M -hez. Mivel azt szeretnénk, hogy a két mátrix azonos sorában és oszlopában lévő elemeinek a különbsége legyen a legkisebb, ezért $M - \widetilde{M}$ elemeinek a négyzetösszegeként határozzuk meg két mátrix közelségét. Ez éppen $M - \widetilde{M}$ Frobenius-, avagy $L_{2,2}$ normájának a négyzete.

Az általunk vizsgált dzsúdó meccsekből létrehozott M mátrix ritka, a játékos-párok többsége sosem játszott egymással. Ez azért jelent gondot, mert az ilyen (i, j) játékosokra $M_{i,j}$ és $M_{j,i}$ helyeken is 0 áll, azonban mikor az X vektort kiválasztjuk, nem célunk, hogy $\widetilde{M}$ ezekre a nullákra is jól illeszkedjen. Ezért hozunk létre egy W mátrixot, ahol az adatok "megbízhatósága" szerepel. Dönthetünk, hogy W -t hogyan definiáljuk: szeretnénk-e súlyozni a meccsek számával, azaz $W_{i,j} = \sum_{k \in \mathcal{H}(i,j) \cup \mathcal{H}(j,i)} s_k$, vagy a súlyokra csak mint indikátorra tekintünk és $W_{i,j} = \mathbb{I}(\mathcal{H}(i, j) \cup \mathcal{H}(j, i) \neq \emptyset)$. A két mátrix távolsága legyen a következő: $\|M - \widetilde{M}\| = \sum_{i=1}^n \sum_{j=1}^n W_{i,j} \cdot (M_{i,j} - \widetilde{M}_{i,j})^2$.

Konvex célfüggvények optimalizálására gyors módszereket ismerünk, ezért először keressünk olyan $p(x_i, x_j)$ függvényt, amire adott M esetén a két mátrix távolsága konvex X függvényében.

Definíció 2.1. Legyen $f : \mathbb{R}^d \rightarrow \mathbb{R}$ kétszer deriválható egy nyílt tartományon (például $\mathbb{R}^n$ fölött). Ekkor f pontosan akkor konvex, ha a Hesse-mátrixa pozitív szemidefinit.

$f(X) = \sum_{i=1}^n \sum_{j=1}^n W_{i,j} \cdot (M_{i,j} - p(x_i, x_j))^2$ esetén számoljuk ki f parciális deriváltjait. Az első deriváltak:

$$\begin{aligned} \frac{\partial}{\partial x_i} f(X) &= \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n W_{i,j} \cdot (M_{i,j} - p(x_i, x_j))^2 + W_{j,i} \cdot (M_{j,i} - p(x_j, x_i))^2 \right) = \\ \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n W_{i,j} p(x_i, x_j)^2 - W_{i,j} \cdot 2M_{i,j} p(x_i, x_j) + W_{j,i} p(x_j, x_i)^2 - W_{j,i} \cdot 2M_{j,i} p(x_j, x_i) \right) &= \end{aligned}$$

$$\begin{aligned}
& \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n W_{i,j} (p(x_i, x_j)^2 - 2M_{i,j}p(x_i, x_j) + (1 - p(x_i, x_j))^2 - 2(1 - M_{i,j})(1 - p(x_i, x_j))) \right) = \\
& \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n W_{i,j} (2p(x_i, x_j)^2 - 2p(x_i, x_j) + 1 - 2M_{i,j}p(x_i, x_j) - 2 + 2M_{i,j} + 2p(x_i, x_j) - 2M_{i,j}p(x_i, x_j)) \right) = \\
& = \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n W_{i,j} (2p(x_i, x_j)^2 - 4M_{i,j}p(x_i, x_j)) \right) = \\
& = \sum_{j=1}^n W_{i,j} \left(4p(x_i, x_j) \cdot \frac{\partial}{\partial x_i} p(x_i, x_j) - 4M_{i,j} \frac{\partial}{\partial x_i} p(x_i, x_j) \right) = \\
& = \sum_{j=1}^n 4W_{i,j} \left((p(x_i, x_j) - M_{i,j}) \cdot \frac{\partial}{\partial x_i} p(x_i, x_j) \right) \tag{1}
\end{aligned}$$

A második deriváltak:

$$\begin{aligned}
& \frac{\partial}{\partial x_l} \frac{\partial}{\partial x_i} f(X) = \frac{\partial}{\partial x_l} \left(\sum_{j=1}^n 4W_{i,j} (p(x_i, x_j) - M_{i,j}) \left(\frac{\partial}{\partial x_i} p(x_i, x_j) \right) \right) = \\
& = 4W_{i,l} \frac{\partial}{\partial x_l} p(x_i, x_l) \left(\frac{\partial}{\partial x_i} p(x_i, x_l) \right) + 4W_{i,l} (p(x_i, x_l) - M_{i,l}) \left(\frac{\partial}{\partial x_l} \frac{\partial}{\partial x_i} p(x_i, x_l) \right)
\end{aligned}$$

2.2 Lineáris $p(x,y)$

A legegyszerűbb eset, ha a valószínűségeket adó $p(x, y)$ függvény lineáris. Vizsgáljuk meg, hogy ilyenkor hogy néz ki a modellünk.

A lineáris p függvény $p(x, y) = ax + by + c$ alakú. Mivel teljesülnek rá a fent megadott feltételek: $p(x, x) = 0,5$, azaz $(a + b)x + c = 0,5$, minden x -re. Emiatt $a = -b$ és $c = 0,5$. Ez esetben a másik feltétel: $p(x, y) = 1 - p(y, x)$ automatikusan teljesül.

Észrevehetjük, hogy az a paraméter csak skálázza a pontszámokat, tehát ha az i versenyző j elleni győzelmének a valószínűsége q , akkor a pontszámaik különbsége $\frac{q-0,5}{a}$. Legyen $a = \frac{1}{2}$, hogy éppen akkor legyen a valószínűség 1, amikor a különbség 1. Tehát megkaptuk a függvényt: $p(x, y) = \frac{x-y}{2} + 0,5$.

Jelöljük H -val az $f_p(X)$ Hesse-mátrixát.

$$\begin{aligned} H_{i,j} &= 4W_{i,l} \frac{\partial}{\partial x_l} p(x_i, x_l) \left(\frac{\partial}{\partial x_i} p(x_i, x_l) \right) + 4W_{i,l} (p(x_i, x_l) - M_{i,l}) \left(\frac{\partial}{\partial x_l} \frac{\partial}{\partial x_i} p(x_i, x_l) \right) = \\ &= 4W_{i,l} \left(-\frac{1}{2} + 0,5 \right) \left(\frac{1}{2} \right) + 4W_{i,l} \left(\frac{x_i - x_j}{2} + 0,5 - M_{i,l} \right) (0) = 0 \end{aligned}$$

Mivel $H_{i,j}$ minden eleme 0, a Hesse-mátrix pozitív szemidefinit, azaz az $f_p(X)$ függvény konvex.

Mivel $f_p(X)$ konvex és kétszer deriválható, így a maximuma éppen ott van, ahol minden parciális deriváltja 0. Tehát az (1) egyenlet alapján a következő lineáris egyenletrendszer megoldását keressük:

$$\forall i : \frac{\partial}{\partial x_i} f(X) = \sum_{j=1}^n 4W_{i,j} \left(\left(\frac{x_i - x_j}{2} + 0,5 - M_{i,j} \right) \cdot \frac{1}{2} \right) = 0$$

Azaz minden i játékosra:

$$x_i \sum_{j=1}^n W_{i,j} - \sum_{j=1}^n W_{i,j} x_j = \sum_{j=1}^n W_{i,j} (2M_{i,j} - 1) \quad (2)$$

Ezt az egyenletrendszert felírhatjuk mátrix formában is. Legyen az A mátrixnak annyi sora és oszlopa, ahány játékos van. A i . sorában az i . helyen $\sum_{j=1}^n W_{i,j}$ áll, bármelyik másik j . helyen pedig éppen $-W_{i,j}$. Legyen a b vektor i . koordinátája $\sum_{j=1}^n W_{i,j} (2M_{i,j} - 1)$.

Oldjuk meg az $AX = b$ egyenletrendszert, és a kapott X vektor éppen az optimális pontszámokat tartalmazza lineáris $p(x, y)$ esetén.

2.3 Bradley–Terry-modell

Egy másik lehetőség, hogy úgy adjuk meg a p függvényt, hogy a két játékos egymás elleni győzelmének a valószínűsége éppen a pontszámaik aránya, azaz $p(x_i, x_j) = \frac{x_i}{x_i + x_j}$. Gyakran az itt definiált x_i és x_j pontszámok helyett a logaritmusunkat adjuk meg: π_i és π_j -t. Ezekre a valószínűséget a $p_e(\pi_i, \pi_j) = \frac{e^{\pi_i}}{e^{\pi_i} + e^{\pi_j}}$ függvény számolja ki. Az új p is teljesíti a korábbi megkötéseinket, azaz $p(x, x) = \frac{x}{x+x} = 0.5$ és $p(x, y) + p(y, x) = \frac{x}{x+y} + \frac{y}{x+y} = 1$.

Ezt a p függvényt használja az Élő-pontozás és a most következő Bradley–Terry-modell is.

A problémát a statisztika irányából először Zermelo közelítette meg 1929-ben. Az motiválta, hogy új pontozási rendszert találjon ki a sakkozók értékelésére. Akkoriban ugyanis a sakkversenyeken a játékosok körmérkőzést játszottak, mindenki mindenkivel pontosan két meccset. Ezeket úgy pontozták, hogy aki megnyerte mindkét mérkőzést, vagy döntetlent játszott, az 1 pontot kapott, aki elvesztette mindkettőt, az nullát. Ez a pontrendszer arra épül, hogy mindenki ugyanannyiszor játszik (ugyanannyi az elérhető összpontszám), ezért nem tudja kezelni, ha egy meccs elmarad, vagy félbeszakad. Egy másik hibája, hogy a versenyzők egymáshoz mért erősségét nem számolja pontosan, hiszen ha egy versenyen egy nagymester és sok amatőr jelentkezik, akkor a nagymester minden meccsét megnyerve is csak kétszer akkora pontszámot kap, mint a kezdők.

Zermelo [8] egy olyan pontszámítást javasolt, ahol mindenkihez egy x_i nemnegatív valós pontszámot tartozik, amire az i játékos j elleni győzelmének a valószínűsége $p(x_i, x_j) = \frac{x_i}{x_i + x_j}$. Az a feladatunk, hogy a meccsek ismeretében meghatározzuk az $x_1, x_2, \dots, x_n$ pontszámokat.

Ehhez vegyük a korábban említett M mátrix egy M^* változatát, ahol $M_{i,j}^* = \{\text{azon meccsek száma, ahol } i \text{ legyőzte } j\}$.

Tegyük fel, hogy a meccsek függetlenek, és minden (i, j) versenyzőpárra az első játékos $\frac{x_i}{x_i + x_j}$ valószínűséggel nyeri meg bármely mérkőzésüket (valamilyen ismeretlen x_i, x_j értékekre). Az pontszámok meghatározásához maximum likelihood becslést fogunk használni.

Definíció 2.2. Diszkrét esetben a likelihood függvény $L_{\vartheta}(X)$ a modell paraméterének egy függvénye, ami azt adja meg, hogy az adott X mintának mennyi a valószínűsége a ϑ paraméter esetén.

A loglikelihood függvény a likelihood logaritmusa: $\ell_{\vartheta}(X) = \log(L_{\vartheta}(X))$.

Ha egy (i, j) versenyzőpárra i t_{ij} , míg j t_{ji} meccset nyert a másik ellen (azaz összesen $T_{ij} = t_{ij} + t_{ji}$ -szer játszottak), akkor adott x_i, x_j -re ennek a valószínűsége:

$$\mathbb{P}(\text{a versenyzők } t_{ij} \text{ és } t_{ji} \text{ meccset nyertek}) = \frac{x_i^{t_{ij}} \cdot x_j^{t_{ji}}}{(x_i + x_j)^{T_{ij}}} \quad (3)$$

Azaz M^* mátrixnak, mint mintának a likelihood függvénye:

$$L_{x_1, x_2, \dots, x_n}(M^*) = \frac{x_1^{t_1} \cdot \dots \cdot x_n^{t_n}}{\prod_{i \neq j} (x_i + x_j)^{T_{ij}}} \quad (4)$$

ahol t_i az i versenyző összes megnyert meccsének a száma a többiek ellen.

Vegyük észre, hogy ha minden pontszámot ugyanazzal a skalárral szorzunk, akkor a valószínűségek nem változnak (hiszen a pontszámok aránya megmarad), ezért az általánosság megszorítása nélkül feltehetjük, hogy $\sum_{i=1}^n x_i \leq 1$. Tehát ekkor a $L_{x_1, x_2, \dots, x_n}(M^*)$ függvényt a korlátos és zárt $\mathcal{B} = \{(x_1, x_2, \dots, x_n) | \forall i : x_i \geq 0, \sum_{i=1}^n x_i \leq 1\}$ halmazon értelmezzük.

Az L függvény valószínűséget határoz meg, azaz 0 és 1 közti értékeket vesz fel, így létezik w , ami felülről határolja. w -ról tudjuk, hogy $1 \geq w \geq 0$. Ha minden versenyző ugyanakkora pontszámot kap, akkor a likelihood függvény pozitív: $L = \left(\frac{1}{2}\right)^{\sum_{i=1}^n t_i} > 0$, ezért $w > 0$.

Vegyük észre, hogy a függvény értéke nem meghatározott olyankor, amikor legalább két versenyző pontszáma nulla, hiszen ekkor

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \lim_{x \rightarrow 0} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) = \lim_{x \rightarrow 0} \left(\frac{\lambda_1 x}{\lambda_1 x + \lambda_2 x} \right) = \frac{0}{0+0} = \lim_{x \rightarrow 0} \left(\frac{\lambda_3 x}{\lambda_3 x + \lambda_4 x} \right) = \frac{\lambda_3}{\lambda_3 + \lambda_4}$$

Ezért ahelyett, hogy a függvény maximumát keressük, vegyünk konvergens $(\bar{p})_n$ sorozatokat $\mathcal{B}' = \{(x_1, x_2, \dots, x_n) | \forall i : x_i > 0, \sum_{i=1}^n x_i \leq 1\}$ -ből, és a $\lim_{n \rightarrow \infty} L(\bar{p}_n)$ határértékekből keressünk minél legnagyobbat, hogy ezzel becsüljük w -t. A $(\bar{p})_n$ sorozat határértéke legyen $\bar{r} = (r_1, r_2, \dots, r_n)$.

Bontsuk a feladatot négy esetre aszerint, hogy hol van $\bar{r}$.

1. eset: $\bar{r}$ a $\mathcal{B}$ belsejében van.

$\bar{r} \in \text{Int}(\mathcal{B})$, tehát létezik olyan $\varepsilon > 0$, hogy $\forall i : r_i > \varepsilon$. Ekkor a $\mathcal{B}_2 = \{(x_1, x_2, \dots, x_n) | \forall i : x_i \geq \varepsilon, \sum_{i=1}^n x_i \leq 1\}$ egy zárt, korlátos tartomány, ahol az L függvény folytonos, azaz a Weierstrass-szélsőértéktétel szerint $\inf_{\mathcal{B}_2} L = \max_{\mathcal{B}_2} L$, azaz L felveszi a maximumát. Ha $\bar{r}$ -ben van a maximumhely, akkor a loglikelihood függvény minden parciális deriváltja 0.

$$\ell_{r_1, r_2, \dots, r_n}(M^*) = \sum_{i=1}^n t_i \cdot \log(r_i) - \left(\sum_{i \neq j} T_{ij} \cdot \log(r_i + r_j) \right) \quad (5)$$

$$\frac{\partial}{\partial r_i} \ell_{r_1, r_2, \dots, r_n}(M^*) = \frac{t_i}{r_i} - \left(\sum_{j=1, i \neq j}^n \frac{T_{ij}}{r_i + r_j} \right) = 0 \quad (6)$$

Ebből következik, hogy $t_i = \sum_{j=1, i \neq j}^n T_{ij} \frac{r_i}{r_i + r_j}$, minden i -re, így kaptunk egy n egyenletből álló egyenletrendszert. Vegyük észre, hogy $T_{ij} \frac{r_i}{r_i + r_j}$ éppen annak a várható értéke, hogy hány meccset nyer i a j ellen a T_{ij} -ből. Tehát ha $r_1, \dots, r_n$ kielégítik az egyenletrendszert, akkor minden versenyző éppen annyi meccset nyert, ami az összes eddigi mérkőzésén a győzelmei számának a várható értéke.

2. eset: $\bar{r}$ a $\mathcal{B}$ határán van, de minden r_i pozitív.

Ekkor $\sum_{i=1}^n r_i = 1$. Emítettük, hogy ha minden pontszámot ugyanazzal a pozitív skalárral szorzunk, akkor a likelihood függvény értéke nem változik. Legyen $1 > \lambda > 0$, ekkor $(\lambda r_1, \lambda r_2, \dots, \lambda r_n)$ már $\mathcal{B}$ belső pontja és $L_{r_1, r_2, \dots, r_n}(M^*) = L_{\lambda r_1, \lambda r_2, \dots, \lambda r_n}(M^*)$, tehát visszavezettük a feladatot az 1. esetre.

3. eset: $\bar{r}$ -nek pontosan egy koordinátája 0.

Az általánosság megsértése nélkül feltehetjük, hogy $r_1 = 0$ és $\forall i > 1 : r_i \geq \varepsilon > 0$, tehát a $\mathcal{B}_3 = \{(x_1, x_2, \dots, x_n) | x_1 \geq 0, \forall i > 1 : x_i \geq \varepsilon, \sum_{i=1}^n x_i \leq 1\}$ zárt, korlátos tartományon L folytonos, így felveszi a maximumát.

Állítás 2.1. *Az $\bar{r}^*$ maximumhelyen pontosan akkor kap egy versenyző $r_1^* = 0$ pontszámot, ha senki ellen nem nyert mérkőzést.*

Bizonyítás. Ha $r_1^* = 0$, $\forall i > 1 : r_i^* > 0$, akkor a likelihood függvényben az r_1^* -ot tartalmazó tényezők $\frac{(r_1^*)^{t_{1i}}}{(r_1^* + r_i^*)^{t_{1i}}}$ és $\frac{(r_i^*)^{t_{i1}}}{(r_i^* + r_1^*)^{t_{i1}}}$ alakúak. Tehát ha indirekten feltesszük, hogy nyert meccset az 1. játékos, akkor $L_{r_1^*, r_2^*, \dots, r_n^*}(M^*) = 0$, hiszen megjelenik egy

$\frac{0^{t_{1i}}}{(r_i^*)^{t_{1i}}} = 0$ tényező a szorzatban, ahol $t_{1i} > 0$. Mivel láttuk, hogy a likelihood függvény felvesz 0-nál nagyobb értéket, ezért $\bar{r}^*$ nem maximumhely. Ezzel ellentmondásra jutottunk. $\nexists$

Nézzük a másik irány bizonyítását. Tegyük fel indirekten, hogy az 1. versenyző nem nyert egyetlen mérkőzést sem, de $r_1^* > 0$. Ekkor a likelihood függvényben az r_1^* csak $\frac{(r_i^*)^{t_{i1}}}{(r_i^* + r_1^*)^{t_{i1}}}$ alakú tényezőkben szerepel. Ha csökkentjük r_1^* -et (hogy még pozitív maradjon), akkor mivel minden másik pontszám pozitív, minden ilyen tényező szigorúan nő, azaz a likelihood függvény is nő. Azaz ellentmondásra jutottunk, hiszen találtunk olyan helyet, ahol L nagyobb értéket vesz fel, mint $\bar{r}^*$ -nál. $\nexists$

□

4. eset: $\bar{r}$ -ben több versenyző is 0 pontszámot kap.

Itt már nem folytonos az L , így $L(\bar{p}_n)$ sorozatok határértékeinek keressük a maximumát, ahol $\bar{p}_n \in \mathcal{B}$ és $\bar{p}_n \rightarrow \bar{r} \in \mathcal{B}$. Mivel L csak a $[0, 1]$ intervallumból vesz fel értékeket, így minden konvergens $L(\bar{p}_n^{(i)})$ sorozat egy $[0, 1]$ -beli $a_{\bar{r}^{(i)}}$ értékhez konvergál. Vegyük azt a $\bar{p}_n^* \rightarrow \bar{r}^*$ sorozatot, amire $a_{\bar{r}^*} = \max\{a_{\bar{r}^{(i)}}\}$ ¹ és vizsgáljuk most az $\bar{r}^*$ pontban a versenyzők pontszámait.

Feltehetjük, hogy $\bar{r}^*$ -ban az első k versenyző pontszáma 0, a többieké pozitív. A 3. eset miatt $\forall i \leq k < j : t_{ij} = 0$, azaz akiknek a pontszáma 0, minden mérkőzésüket elvesztették azon sportolók ellen, akiknek pozitív a pontszámuk. Ha nyertek volna meccset, például az i a j ellen, akkor létezne olyan $\left(\frac{x_i}{x_i + x_j}\right)^{t_{ij}}$ tényező L -ben, ami 0-hoz tart, ha $x_i \rightarrow 0$, minden $\bar{r}^*$ -hoz konvergáló sorozatra. Ezért ekkor $L(\bar{r}^*) = 0$, azaz $\bar{r}^*$ nem maximumhely.

Vizsgáljuk meg a másik irányt: tegyük fel, hogy létezik olyan csoportja a versenyzőknek, akik egy meccset sem nyertek a többiek ellen. Legyenek az első csoportban az 1, 2, ..., k , a másodikban a $k + 1, k + 2, \dots, n$ versenyzők. Bontsuk az L függvényt három tényező szorzatára úgy, hogy az egyikbe kerülnek az első csoporton belüli, a másikba a második csoporton belüli, a harmadikba pedig a két csoport közti meccsekhez tartozó

¹ Kihasnálva hogy a $p_n^{(i)}$ sorozatok a $\mathcal{B} \in \mathbb{R}^n$ zárt és korlátos tartományban vannak és az $a_{\bar{r}^{(i)}}$ értékek is korlátosak, meggondolható, hogy $\sup\{a_{\bar{r}^{(i)}}\} = \max\{a_{\bar{r}^{(i)}}\}$.

$\frac{x_i^{t_{ij}} \cdot x_j^{t_{ji}}}{(x_i + x_j)^{T_{ij}}}$ alakú tényezők.

$$\begin{aligned} L_{x_1, x_2, \dots, x_n}(M^*) &= L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(12)}(M^*) = \\ &= L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*) \cdot \prod_{i \leq k < j} \frac{x_i^{t_{ij}} \cdot x_j^{t_{ji}}}{(x_i + x_j)^{T_{ij}}} = \\ &= L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*) \cdot \prod_{i \leq k < j} \frac{x_j^{t_{ji}}}{(x_i + x_j)^{t_{ji}}} \end{aligned}$$

Tegyük fel indirekten, hogy van olyan pozitív pontszámú játékos az első csoportban, aki kikapott valakitől a második csoportból. Feltehetjük, hogy ez a mérkőzés éppen az 1. és $k + 1$. versenyzők között történt. Ha a $\bar{r}_{k+1}^* = 0$, akkor bármilyen $\bar{r}^*$ -hoz tartó sorozatra $x_{k+1} \rightarrow 0$, azaz a $k + 1$. versenyző győzelmének a valószínűsége tart a nullához az $\bar{r}_1 > 0$ pontszámú versenyző elleni meccsen, azaz $L \rightarrow 0$. Tehát $\bar{r}_{k+1}^* > 0$.

Ha minden az első csoportba tartozó versenyző pontszámának vesszük a λ -szorosát (ahol $1 > \lambda > 0$), akkor nekik egymáshoz képest a relatív pontszámuk nem változik, azaz $L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) = L_{\lambda x_1, \lambda x_2, \dots, \lambda x_n}^{(1)}(M^*)$. A második csoportbeli meccsek valószínűségei is ugyanazok maradnak, azaz az egyetlen eltérés $L^{(12)}$ -ben van. Itt minden tényező számlálója változatlan, hiszen a számlálóban csak a második csoport versenyzőinek pontszámai szerepelnek. Az olyan nevezők, amikben az első csoport játékosának pozitív pontszáma van, csökkennek. Tehát $L^{(12)}$ -ben minden tényező változatlan vagy nő. Mivel esetünkben van olyan első csoportbeli versenyző (az 1-es), akinek pozitív pontszáma van, így $L^{(12)}$ szigorúan nő. Tehát ellentmondásra jutotunk, ha van $r_i > 0$ ($i \leq k$), hogy az i . versenyző játszott meccset egy második csoportbeli játékosal, akkor L növelhető, tehát $\bar{r}^*$ nem maximumhely.

Ha a játékosok két csoportra oszthatók úgy, hogy az első csoport tagjai egy mérkőzést sem nyertek a másik csoport ellen, akkor L csak akkor lehet a legnagyobb, ha az első csoportból mindenkinek a pontszáma, aki veszített a másik csoport egy tagja ellen, 0-hoz tart. Ilyenkor L a következő alakba írható: $L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*) \cdot \prod_{i \leq k < j} \frac{x_j^{t_{ji}}}{(x_i + x_j)^{t_{ji}}} = L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*) \cdot \prod_{i \leq k < j} \frac{x_j^{t_{ji}}}{x_j^{t_{ji}}} = L_{x_1, x_2, \dots, x_n}^{(1)}(M^*) \cdot L_{x_1, x_2, \dots, x_n}^{(2)}(M^*)$.

Ábrázoljuk a mérkőzéseket irányított gráfként, aminek a csúcsai a versenyzők, és pontosan akkor rajzolunk élt u pontból v -be, ha volt olyan meccs, ahol a v legyőzte az u -t.

Definíció 2.3. *Egy irányított gráf erősen összefüggő, ha bármelyik csúcsból bármelyikbe vezet irányított út.*

A fentiek miatt, ahol a gráfban egyirányú vágás van, azaz két csúcshalmaz, melyek között csak az egyik irányba futnak élek, ott a feladattól két kisebb részfeladatot csinálhatunk. Ilyen módon a csúcsokat addig bonthatjuk kisebb csoportokra, míg erősen összefüggő komponenseket nem kapunk. A likelihood függvény csak akkor lehet maximális, ha bármelyik két komponensre, melyek között megy irányított út, vagy az elsőben a pontszámok nullához, vagy a másodikban a pontszámok végtelenhez tartanak, azaz az elsőben minden r_i elhanyagolhatóan kicsi a másodikban lévő r_j -hez képest. Mivel szeretnénk a versenyzők erősségeit összehasonlítani, ezért a gyakorlatban olyan esetekben, amikor a mérkőzések gráfja nem erősen összefüggő, a maximum likelihood becslés nem alkalmas a pontszámok meghatározására, hiszen egy komponensen kívül minden más versenyző pontszámára 0-t vagy ∞ -t ad. Ilyenkor más becslést fogunk alkalmazni.

2.3.1 Erősen összefüggő eset

Ha a mérkőzések gráfja erősen összefüggő, akkor az is következik, hogy nincs olyan versenyző, aki minden meccsét elveszítette vagy megnyerte volna.

Tegyük fel, hogy van olyan, akinek a pontszáma nem 0. Ahogy közelítünk $\bar{r}^*$ -hoz, nincs olyan versenyző, akinek a pontszáma 0-hoz tart. Ellenkező esetben, például, ha az 1. versenyző pontszámával 0-hoz tartunk, akkor L csak akkor nem tart 0-hoz, ha minden versenyzőnek, akit az 1. legyőzött, a pontszáma 0-hoz tart. Mivel a gráf erősen összefüggő, ezért ekkor minden játékos pontszáma 0-hoz tart.

Így ha a mérkőzések rendszere erősen összefüggő, akkor a fenti négy eset közül csak az első kettő következhet be. Ilyenkor létezik a maximum, és a korábban felírt egyenletek teljesülnek:

$$\forall i : t_i = \sum_{j=1, i \neq j}^n T_{ij} \frac{r_i}{r_i + r_j} \quad (7)$$

Már beláttuk, hogy az első két eset ekvivalens, hiszen ha minden értéket ugyanazzal a skalárral szorozzuk, akkor L nem változik. Lássuk be, hogy a maximum egyértelmű azzal a feltétellel, hogy $\sum_{i=1}^n r_i = 1$. Indirekt tegyük fel, hogy létezik két maximális megoldás $\bar{r}$ és $\bar{r}'$. Mivel a két megoldás nem azonos, így $\exists k$, hogy a koordináták sorrendjét meg tudjuk úgy változtatni, hogy $\forall i \leq k : r_i < r'_i$ és $\forall j > k : r_j \geq r'_j$. A feltételünk miatt, hogy a pontszámok összege 1, azt is tudjuk, hogy $0 < k < n$.

Ekkor minden $i \leq k < j$ -re $\frac{r'_i}{r'_j} > \frac{r_i}{r_j}$, azaz $\frac{r'_i}{r'_i + r'_j} > \frac{r_i}{r_i + r_j}$.

Vegyük a következő összeget és rendezzük át

$$\begin{aligned} \sum_{i=1}^k \sum_{j=k+1}^n T_{ij} \frac{r_i}{r_i + r_j} &= \sum_{i=1}^k \sum_{j \leq n, j \neq i} T_{ij} \frac{r_i}{r_i + r_j} - \sum_{i=1}^k \sum_{j \leq k, j \neq i} T_{ij} \frac{r_i}{r_i + r_j} = \\ &= \sum_{i=1}^k t_i - \sum_{i < j \leq k} \left(T_{ij} \frac{r_i}{r_i + r_j} + T_{ji} \frac{r_j}{r_i + r_j} \right) = \sum_{i=1}^k t_i - \sum_{i < j \leq k} T_{ij} \end{aligned}$$

A fenti összeg megegyezik az átrendezés után első k versenyző által összesen megnyert meccsek számának és az első k versenyző egymás ellen játszott mérkőzéseinek számának a különbségével.

Mivel $\forall i \leq k < j : \frac{r'_i}{r'_i + r'_j} > \frac{r_i}{r_i + r_j}$, így ezt behelyettesítve:

$$\sum_{i=1}^k t_i - \sum_{i < j \leq k} T_{ij} = \sum_{i=1}^k \sum_{j=k+1}^n T_{ij} \frac{r_i}{r_i + r_j} < \sum_{i=1}^k \sum_{j=k+1}^n T_{ij} \frac{r'_i}{r'_i + r'_j} = \sum_{i=1}^k t_i - \sum_{i < j \leq k} T_{ij}$$

Amivel ellentmondásra jutottunk, hiszen a megnyert, illetve lejátszott meccsek száma konstans, nem függ a pontszámoktól. $\nexists$

Tehát a $\sum_{i=1}^n r_i = 1$ megkötés mellett a megoldás egyértelmű.

Térjünk vissza az általános feladatra. Az erősen összefüggő komponensek között felállíthatunk egy részleges rendezést. Ebben $\mathcal{C}_1$ és $\mathcal{C}_m$ komponensekre $\mathcal{C}_1 < \mathcal{C}_m$, ha van a komponenseknek egy olyan $\mathcal{C}_1, \mathcal{C}_2, \dots$ és $\mathcal{C}_m$ sorozata, hogy $\forall 1 \leq i < m : \mathcal{C}_i$ és $\mathcal{C}_{i+1}$ komponensekre van olyan versenyző $\mathcal{C}_{i+1}$ -ben, aki legyőzött valakit $\mathcal{C}_i$ -ből. Mivel a komponenseket egy ponttá összehúzáva aciklikus gráfot kapunk, így nincs két

komponens, amire $\mathcal{C}_1 < \mathcal{C}_m$ és $\mathcal{C}_1 > \mathcal{C}_m$ is igaz.

Két komponens pontosan akkor nem összehasonlítható, ha a gráfban nem vezet irányított út az egyikből a másikba.

Számoljuk ki a komponensekre külön-külön az optimális r_i értékeket. Ha egy komponens minden versenyzőjének a pontszámát ugyanazzal a $\lambda > 0$ számmal szorozzuk, akkor a komponensre az L függvény értéke változatlan. Ha minden versenyzőre keressük a legjobb $\bar{r}$ pontszámokat, akkor a 3. és 4. eset miatt akkor kapjuk a w közelítését, ha a komponenseken belül a pontszámok aránya megmarad, míg minden olyan komponenspárra, ahol $\mathcal{C}_i < \mathcal{C}_j$, ott az i és j komponensbeli pontszámok aránya a 0-hoz tart.

2.3.2 A pontszámok kiszámítása

Ford 1957-es cikkében [4] bizonyította, hogy erősen összefüggő esetben a maximum egyértelmű, és iterációval tudunk konvergálni hozzá. Ő a maximum létezését és egyértelműségét kicsit eltérő irányból közelítette meg. A skálázhatóság miatt az eredeti feladattal ekvivalens, ha olyan pontszámokat keresünk, melyekre $\{0 > x_i, \sum x_i = 1\}$. A maximum létezéséhez vegyük a tartomány lezártját: $\{0 \geq x_i, \sum x_i = 1\}$ -t. Ford belátta, hogy ha a likelihood függvényt ennek a kisebb felületnek a határpontjain 0-nak definiálja, akkor L folytonos a tartományon.

Legyen $\bar{x}'$ egy pont a határon, azaz $\exists j : x'_j = 0$. Mivel $\sum x_i = 1$, ezért $\exists k : x'_k > 0$. Feltettük, hogy erősen összefüggő a meccsek rendszere, azaz van a versenyzőknek olyan $v_0, v_1, \dots, v_m$ sorozata, amire $v_0 = j$, $v_m = k$ és $\forall 0 \leq i < m : t_{v_i, v_{i+1}} > 0$, azaz a sorozatban bármely két egymást követő emberből az első legalább egyszer nyert a második ellen. A sorban az első olyan versenyző, akinek nem 0 a pontszáma, legyen v_l . Az L függvénynek van $\left(\frac{x'_{v_{l-1}}}{x'_{v_{l-1}} + x'_{v_l}} \right)^{t_{v_{l-1}, v_l}}$ alakú tényezője (és a kitevő pozitív), azaz bármilyen $(\bar{x})_n$ sorozattal tartunk $\bar{x}'$ -höz, a tényező 0-hoz tart. Mivel minden tényező legfeljebb 1, így ilyenkor L is 0-hoz tart. Tehát ha a határpontokon 0-nak definiáljuk, L folytonos a tartomány lezártján.

Egy korlátos, zárt tartományon folytonos függvénynek van maximuma a Weierstrass-tétel szerint, azaz a maximum létezését így is be tudtuk látni.

Ahogy korábban leírtuk, azt az X pontszám-vektort keressük, ami teljesíti a

következő feltételeket:

$$\forall i : t_i = \sum_{j=1, i \neq j}^n T_{ij} \frac{x_i}{x_i + x_j}$$

A megoldás megtalálására nincs gyors, explicit megoldás, ezért iterációkkal konvergálunk a maximumhelyhez. A kezdőpont legyen egy $\bar{x}^{(0)}$, amire $\forall i : x_i^{(0)} > 0$ és egy iterációs lépés:

$$x_i^{(k+1)} = \frac{t_i}{\sum_{j=1}^n \frac{T_{ij}}{x_i^{(k)} + x_j^{(k)}}}$$

$$x_j^{(k+1)} = x_j^{(k)}, \forall j \neq i$$

ahol $i = 1, 2, \dots, n$ ciklizálva.

Jegyezzük meg, hogy ha belső pontból indultunk (azaz $\forall i : x_i > 0$), akkor végig a tartomány belsejében maradunk iteráció közben. Egy iteráció után a $\sum_i x_i = 1$ feltétel már nem lesz igaz. Megtehetnénk, hogy iterációnként normáljuk az összes pontszámot, azonban az a tapasztalat, hogy mivel nekünk a pontszámok arányai a fontosak, ezért nem lesz szükség arra, hogy folyton normáljunk, ezt elég csak körönként megtennünk, miután minden koordinátát egyszer frissítettünk.

Lássuk be az iteráció konvergenciáját. Vizsgáljunk egyszerre egy lépést (ahol épp az i versenyző pontszámán javítunk). Bebizonyítjuk, hogy $\bar{x}^{(k-1)}$ és $\bar{x}^{(k)}$ között $L_{\bar{x}}$ monoton nő. Először vezessük be a következő két függvényt:

$$\delta_i(\bar{x}) = \frac{\partial l_{\bar{x}}(M^*)}{\partial x_i} = \frac{t_i}{x_i} - \left(\sum_{j=1}^n \frac{T_{ij}}{x_i + x_j} \right) \quad (8)$$

$$D_i(\bar{x}) = \delta_i(\bar{x}) \cdot x_i = t_i - \left(\sum_{j=1}^n \frac{T_{ij} \cdot x_i}{x_i + x_j} \right) \quad (9)$$

$D_i = \delta_i \cdot x_i$ monoton $x_i \in \mathbb{R}^+$ -re, ha minden mászt rögzítünk, hiszen ahogy x_i nő, úgy minden $\frac{T_{ij} \cdot x_i}{x_i + x_j}$ tag nő. Tehát, ha tudnánk, hogy $\delta_i(\bar{x}^{(k-1)})$ és $\delta_i(\bar{x}^{(k)})$ előjele megegyezik, akkor az is igaz, hogy $\bar{x}^{(k-1)}$ és $\bar{x}^{(k)}$ közötti szakasz minden pontjára ugyanaz δ_i előjele.

Lássuk be, hogy minden $k \geq 1$ -re $\delta_i(\bar{x}^{(k)})$ és $\delta_i(\bar{x}^{(k-1)})$ előjele megegyezik. Használjuk a fenti két egyenletet:

$$\begin{aligned}
\delta_i(\bar{x}^{(k)}) &= \frac{t_i}{x_i^{(k)}} - \left(\sum_{j=1}^n \frac{T_{ij}}{x_i^{(k)} + x_j^{(k)}} \right) = \frac{t_i}{\frac{\sum_{j=1}^n \frac{t_i T_{ij}}{x_i^{(k-1)} + x_j^{(k-1)}}}{\sum_{j=1}^n \frac{t_i T_{ij}}{x_i^{(k-1)} + x_j^{(k-1)}}}} - \left(\sum_{j=1}^n \frac{T_{ij}}{\frac{\sum_{l=1}^n \frac{t_i T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}}{\sum_{l=1}^n \frac{t_i T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}} + x_j^{(k)}} \right) = \\
&= \left(\sum_{j=1}^n \frac{T_{ij}}{x_i^{(k-1)} + x_j^{(k-1)}} \right) - \left(\sum_{j=1}^n \frac{T_{ij} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}}{t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}} \right) = \\
&= \sum_{j=1}^n T_{ij} \left(\frac{1}{x_i^{(k-1)} + x_j^{(k-1)}} - \frac{\sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}}{t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}} \right) = \\
&= \sum_{j=1}^n T_{ij} \frac{t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} - \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} (x_i^{(k-1)} + x_j^{(k-1)})}{(x_i^{(k-1)} + x_j^{(k-1)}) \left(t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} \right)} = \\
&= \sum_{j=1}^n T_{ij} \frac{t_i - x_i^{(k-1)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}}}{(x_i^{(k-1)} + x_j^{(k-1)}) \left(t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} \right)} = \\
&= \sum_{j=1}^n T_{ij} \frac{\delta_i(\bar{x}^{(k-1)}) \cdot x_i^{(k-1)}}{(x_i^{(k-1)} + x_j^{(k-1)}) \left(t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} \right)} = \\
&= \delta_i(\bar{x}^{(k-1)}) \cdot \sum_{j=1}^n T_{ij} \frac{x_i^{(k-1)}}{(x_i^{(k-1)} + x_j^{(k-1)}) \left(t_i + x_j^{(k)} \cdot \sum_{l=1}^n \frac{T_{il}}{x_i^{(k-1)} + x_l^{(k-1)}} \right)}
\end{aligned}$$

Mivel a második tényező mindig pozitív (ha $\bar{x}$ a tartomány belső pontja), ezért $\delta_i(\bar{x}^{(k-1)})$ és $\delta_i(\bar{x}^{(k)})$ előjele megegyezik, tehát $\delta_i(\bar{x})$ előjele a két pont közötti egész szakaszon ugyanaz marad.

A frissítés szabálya szerint $\sum_{j=1}^n \frac{T_{ij}}{x_i^{(k-1)} + x_j^{(k-1)}} = \frac{t_i}{x_i^{(k)}}$, azaz

$$\delta_i(\bar{x}^{(k-1)}) = \frac{t_i}{x_i^{(k-1)}} - \left(\sum_{j=1}^n \frac{T_{ij}}{x_i^{(k-1)} + x_j^{(k-1)}} \right) = \frac{t_i}{x_i^{(k-1)}} - \frac{t_i}{x_i^{(k)}} = \frac{t_i(x_i^{(k)} - x_i^{(k-1)})}{x_i^{(k)} \cdot x_i^{(k-1)}}$$

$\delta_i(w)$ definíció szerint $\frac{\partial l_{\bar{x}}(M^*)}{\partial x_i}$. Mivel δ_i előjele nem változik $\bar{x}^{(k-1)}$ és $\bar{x}^{(k)}$ között, és $\delta_i(\bar{x}^{(k-1)})$ éppen akkor pozitív vagy negatív, ha $\bar{x}_i^{(k)} - \bar{x}_i^{(k-1)}$ is az, ezért $\bar{x}^{(k-1)}$ és $\bar{x}^{(k)}$ között $L_{\bar{x}}$ monoton nő (hiszen a deriváltja végig nemnegatív). Sőt, amennyiben $\delta_i(\bar{x}^{(k-1)}) \neq 0$, az L függvény szigorúan monoton nő. Ha $\delta_i(\bar{x}^{(k-1)}) = 0$, akkor $\bar{x}_i^{(k)} = \bar{x}_i^{(k-1)}$, azaz az iteráció közben nem mozdultunk.

Most térjünk át az egész algoritmus vizsgálatára. Minden körben végigmegyünk $\bar{x}$ minden koordinátáján, közben L monoton nő. Minden kör végén normalizáljuk a pontszámokat, hogy az összegük 1 legyen, eközben a függvényérték változatlan. A normált $\bar{x}$ -ekből kapjuk az $\bar{x}^K$ sorozatot. Mivel L korlátos és monoton nő a körök során, ezért $L_{\bar{x}^K}$ konvergens. Legyen $\bar{x}^K$ egy torlódási pontja $\bar{x}^*$. Ha L konvergál, akkor $L_{\bar{x}}$ körönkénti növekedése a 0-hoz tart. Az, hogy $L_{\bar{x}}$ mennyit nő egy kör alatt folytonos $\bar{x}$ -ben, azaz $\bar{x}^*$ -ban L körönkénti növekedése 0.

Ha egy egész körön át nem javítunk, akkor $\forall i : \delta_i(\bar{x}) = 0$, ami átrendezve épp azt jelenti, hogy $\bar{x}$ teljesíti az (5) egyenleteket, tehát megtaláltuk az L függvény maximumát. Mivel az $\bar{x}^K$ sorozat bármely $\bar{x}^*$ torlódási pontja maximumhely (amiről beláttuk, hogy egyértelmű), így pontosan egy torlódási pont van, azaz a sorozat konvergál: $\bar{x}^K \rightarrow \bar{x}^*$.

2.3.3 Gyorsabb konvergencia

Beláttuk, hogy az előbb mutatott iterációval konvergálunk az optimális pontszámokhoz, a konvergencia lassú. 1957 óta többféle iterációs lépést is találtak, amik mind bizonyítottan konvergálnak ugyanehhez a maximumhoz. M. E. J. Newman [6] is bemutatótt egy másik iterációs eljárást, amivel bár nem bizonyítható, hogy minden esetben gyorsabb lenne a konvergencia, de általában gyorsabban megtaláljuk a keresett pontszámokat.

Ennek az iterációnak a frissítési szabályához is azt használjuk fel, hogy a loglikelihood deriváltja 0 az optimális X -re ((6) egyenlet), de a tagokat máshogy csoportosítjuk:

$$\frac{t_i}{r_i} - \left(\sum_{j=1, i \neq j}^n \frac{T_{ij}}{r_i + r_j} \right) = \sum_{j=1, i \neq j}^n \left(\frac{t_{ij}}{r_i} \cdot \left(\frac{r_j}{r_i + r_j} + \frac{r_i}{r_i + r_j} \right) \right) - \left(\sum_{j=1, i \neq j}^n \frac{T_{ij}}{r_i + r_j} \right) =$$

$$= \sum_{j=1, i \neq j}^n \left(\frac{t_{ij}}{r_i} \cdot \left(\frac{r_j}{r_i + r_j} \right) \right) - \left(\sum_{j=1, i \neq j}^n \frac{t_{ji}}{r_i + r_j} \right) = 0$$

Így az iteráció:

$$r_i^{(k+1)} = \frac{\sum_{j=1, i \neq j}^n \frac{t_{ij} r_j^{(k)}}{r_i^{(k)} + r_j^{(k)}}}{\sum_{j=1, i \neq j}^n \frac{t_{ji}}{r_i^{(k)} + r_j^{(k)}}}$$

A módszernek egy előnye, hogy már az első körben kiderül, ha valaki egy meccset sem nyert, vagy minden meccsét megnyerte. Ha i nem nyert meccset, akkor a számlálóban 0 áll, tehát az első iteráció után 0 lesz a pontszáma, míg ha egy meccset sem veszített, akkor a nevezőbe kerül 0 és a pontszáma ∞ lesz rögtön az első körben.

2.3.4 Általános eset

A Bradley–Terry-modellhez eddig felsorolt pontszámítási módszerek mind csak olyan esetben működtek, amikor a meccsek rendszere erősen összefüggő gráfot alkotott. A továbbiakban bemutatunk egy módszert, M. E. J. Newman [6] cikke alapján, amivel kezelni tudjuk a nem erősen összefüggő eseteket is.

Keressünk pozitív pontszámokat és az eddig használt $\sum_{i=1}^n x_i = 1$ feltétel helyett a következőkben vegyük a $\prod_{i=1}^n x_i = 1$ megkötést. Korábban beláttuk, hogy ha $\sum_{i=1}^n x'_i = 1$ teljesül, akkor pontosan egy pontszám-vektor létezik, ami maximalizálja a likelihood függvényt. Ekkor, ha az x'_i pontszámokat mind elosztjuk $\sqrt[n]{\prod_{i=1}^n x_i}$ -vel, akkor a szorzatuk éppen 1. Egyértelmű megfeleltetés van a $\sum_{i=1}^n x'_i = 1$ és $\prod_{i=1}^n x_i = 1$ megoldások között, ezért mindkét megkötéssel is egyértelmű a megoldás. Azért tértünk át erre a normalizációra, mert ebben az esetben könnyen látszik, hogy az átlagos versenyző éppen az 1 pontszámot kapja, azaz annak a valószínűsége, hogy az x_i pontszámú versenyző nyerjen az átlagos játékos ellen éppen $\frac{x_i}{x_i + 1}$.

A továbbiakban az $\bar{x}$ pontszám-vektor helyett keressük a pontszámok logaritmusát. Ezt jelöljük $\bar{\pi}$ -vel.

Ehhez a maximum likelihood becslés helyett használjunk maximum a posteriori (MAP) becslést. Ilyenkor van egy előzetes tudásunk a keresett paraméter ($\bar{\pi}^*$) eloszlásáról, ez az *a priori* eloszlás. Vesszük a $\mathbb{P}(\bar{\pi}|T)$ feltételes valószínűséget, ahol T a

minta, azaz esetünkben a meccsek rendszere. Ez a posteriori eloszlás, és mi olyan $\bar{\pi}$ paramétert keresünk, ami ezt maximalizálja.

Válasszuk meg az π_i -k feltételezett eloszlását. Mivel nincs előzetes tudásunk az $\bar{\pi}$ -ról, ezért legyen minden i -re $\frac{x_i}{x_i+1}$ -nek – azaz az átlagos játékos elleni győzelem valószínűségének – eloszlása egyenletes a $[0; 1]$ intervallumon.

Ekkor a sűrűségfüggvény transzformációval megkapjuk, hogy x_i sűrűségfüggvénye $f_{x_i}(x) = \frac{1}{(1+x)^2}$ és s_i sűrűségfüggvénye pedig $f_{\pi_i}(\pi) = \frac{e^\pi}{(1+e^\pi)^2} = \frac{1}{(1+e^\pi)(1+e^{-\pi})} = \frac{x}{(1+x)^2}$.

A feltételes valószínűség:

$$\mathbb{P}(\bar{\pi}|T) = \prod_{i,j} \left(\frac{e^{\pi_i}}{e^{\pi_i} + e^{\pi_j}} \right)^{t_{ij}} \cdot \prod_i \frac{1}{(1 + e^{\pi_i})(1 + e^{-\pi_i})} \cdot C$$

ahol C konstans független $\bar{\pi}$ -től és T -től. Olyan $\bar{\pi}$ erősségeket keresünk, amik $\mathbb{P}(\bar{\pi}|T)$ -t maximalizálják.

Vegyük észre, hogy $\prod_{i,j} \left(\frac{e^{\pi_i}}{e^{\pi_i} + e^{\pi_j}} \right)^{t_{ij}} \cdot \prod_i \frac{1}{(1+e^{\pi_i})(1+e^{-\pi_i})} = \prod_{i,j} \left(\frac{x_i}{x_i+x_j} \right)^{t_{ij}} \cdot \prod_i \frac{x}{(1+x)^2}$ éppen annak a valószínűsége, hogy $\forall i$ játékos az eddigi meccsein kívül egy mérkőzést megnyert és egyet elveszített az átlagos játékos ellen. Szemléletesen tehát a mérkőzések gráfját úgy tettük összefüggővé, hogy létrehoztunk egy új csúcsot, amiből minden másikba rajzoltunk oda-vissza egy-egy irányított élt. Ezután a fent bemutatott iterációs lépésekkel meg tudjuk találni az optimális pontszámokat.

Tehát azt kaptuk, hogy pontszámok maximum a posteriori becslése ekvivalens azzal, mintha maximum likelihood becslést alkalmaztunk volna kibővített mérkőzések rendszerére.

2.4 Élő-pontozás

Ahogy a Bradley–Terry-modellt, úgy az Élő-pontozást is a sakkversenyek pontozása ihlette. A pontrendszer kitalálója Élő Árpád fizikus, sakkozó és az Egyesült Államok Sakkszövetségének alapító tagja.

A pontozás menete a következő: kezdőként minden versenyző pontszáma 1500, majd a pontszámot minden meccs után frissítjük attól függően, hogy nyert vagy veszített az illető és hogy milyen erős volt hozzá képest az ellenfele. Két versenyző z_i és z_j pontszámaiból az első győzelmének a valószínűségét a $\mathbb{P}(\text{az } i \text{ legyőzi } j\text{-t}) = \frac{1}{1+10^{\frac{z_j-z_i}{400}}}$ képlettel számoljuk.

Vegyük észre, hogy a győzelem valószínűségének ilyen számolása (a pontszámok skalárszorosától eltekintve) ekvivalens a korábban említett Bradley–Terry-modellével, hiszen átrendezve $\mathbb{P}(\text{az } i \text{ legyőzi } j\text{-t}) = \frac{1}{1+10^{\frac{z_j-z_i}{400}}} = \frac{10^{\frac{z_i}{400}}}{10^{\frac{z_i}{400}}+10^{\frac{z_j}{400}}} = \frac{e^{c \cdot z_i}}{e^{c \cdot z_i}+e^{c \cdot z_j}}$, ahol $c = \frac{\log(10)}{400}$.

Az eltérés az előző modellel a pontszámok kiszámításában rejlik. Iteratív eljárás helyett most online algoritmust használunk. Ennek nagy előnye, hogy ha lejártsanak egy új mérkőzést, akkor nem kell előről kezdenünk a pontszámok számítását, hanem meccsenként egy rövid képletbe behelyettesítve megkapjuk az új pontszámokat.

A győzelem valószínűsége nem függ a pontszámoktól, csak azok különbségétől, ezért bevezethetjük a $b(z_i - z_j) = \frac{1}{1+10^{\frac{z_j-z_i}{400}}} = \frac{1}{1+e^{-cx}}$ jelölést. Ahogy ezt korábban beláttuk $b(z_i - z_j) + b(z_j - z_i) = 1$, illetve $b(0) = 0.5$.

Tegyük fel, hogy a k . mérkőzés az i és j játékosok között zajlott, akiknek a pontszáma eddig z_i^k és z_j^k volt. A pontszámok frissítésére a szabály a következő:

$$z_i^{k+1} = z_i^k + 2K(\mathbb{I}(i \text{ nyerte a } k. \text{ meccset}) - b(z_i - z_j))$$

ahol K egy előre megadott konstans. Ezzel a frissítéssel a két pontszám összege nem változik, hiszen $z_i^{k+1} + z_j^{k+1} = z_i^k + z_j^k + 2K(\mathbb{I}(i \text{ nyerte a } k. \text{ meccset}) + \mathbb{I}(j \text{ nyerte a } k. \text{ meccset}) - b(z_i - z_j) - b(z_j - z_i)) = z_i^k + z_j^k + 2K(1 - 1) = z_i^k + z_j^k$.

2.4.1 Konvergencia két játékosra

Tegyük fel, hogy az általunk vizsgált i és j játékosok ereje nem változik az idők során, azaz minden mérkőzésüket i ugyanazzal a p_{ij} valószínűséggel nyeri meg. Krifa, Spinelli és Junca bemutattak egy bizonyítást, hogy ha a két játékosra a kettejük közti mérkőzések száma tart a végtelenhez, akkor a $b(z_i - z_j)$ -k (a prediktált valószínűségek) átlaga tart p_{ij} -hez.

Mivel mindig csak a játékosok pontszámainak a különbségével számolunk, ezért vezessük be a $d_{ij}^k = z_i^k - z_j^k$ változót. Erre a frissítési szabály a $d_{ij}^{k+1} = d_{ij}^k + 2K(\mathbb{I}(i \text{ nyerte a } k. \text{ meccset}) - b(d_{ij}^k))$. Legyen d_{ij}^0 adott valós szám és jelöljük $\mathcal{D}^k$ -nel a k . meccs előtt a lehetséges d_{ij}^k -k halmazát. Minden mérkőzésnél, az eredménytől függően $d_{ij}^{k+1} = d_{ij}^k + 2K(1 - b(d_{ij}^k))$ vagy $d_{ij}^k - 2Kb(d_{ij}^k)$. Vizsgáljuk meg az $\mathcal{M}^k = \max \mathcal{D}^k$ sorozatot. Ezt úgy kapjuk, hogy minden körben $d_{ij}^k + 2K(1 - b(d_{ij}^k))$ -val frissítünk (azaz minden meccset az i játékos nyert). Vezessük be a $g_1(x) = x + 2K(1 - b(x))$ függvényt, erre $\mathcal{M}^{k+1} = g_1(\mathcal{M}^k)$. A g_1 szigorúan monoton nő, ha a deriváltja pozitív, azaz ha $2K \cdot b'(x) < 1$. Esetünkben $b'(x) = \left(\frac{1}{1+e^{-cx}}\right)' = -\frac{1}{(1+e^{-cx})^2} \cdot e^{-cx} \cdot -c$, ahol $\frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} \leq \frac{1}{4}$ a számtani-mértani közepek közötti összefüggés miatt mindig igaz. Tehát $K < \frac{2}{c} = \frac{800}{\log(10)}$ választással a $b'(x)$ -re vonatkozó feltételt mindig teljesítjük.

Lássuk be, hogy az $\mathcal{M}^k$ monoton növekvő sorozat a végtelenbe tart. Ehhez az kell, hogy minden G számra létezzen olyan N indexe a sorozatnak, hogy $\forall k' > N$ -re $\mathcal{M}^{k'} > G$. Ha $\mathcal{M}^k \leq G$, akkor $\mathcal{M}^{k+1} - \mathcal{M}^k = 2K(1 - b(\mathcal{M}^k))$. Tudjuk, hogy $b(\mathcal{M}^k) = \frac{1}{1+10^{-\frac{\mathcal{M}^k}{400}}} \leq \frac{1}{1+10^{-\frac{G}{400}}}$, azaz $2K(1 - b(\mathcal{M}^k)) > 2K \left(1 - \frac{1}{1+10^{-\frac{G}{400}}}\right)$. Mivel minden iterációkor $\mathcal{M}^k$ értéke legalább egy adott $G_1 = 2K \left(1 - \frac{1}{1+10^{-\frac{G}{400}}}\right)$ konstanssal nő, ezért egy idő után átlépjük G -t. Azonban annak a valószínűsége, hogy $\mathcal{D}^k = \mathcal{M}^k$, a nullához tart, hiszen azt jelenti, hogy i minden mérkőzést megnyert.

Tétel 2.1. *Ha k elég nagy $\mathcal{M}^k \sim \frac{\log(k)}{c}$, ahol $c = \frac{\log(10)}{400}$.*

A bizonyításhoz először vezessük be a $\mathcal{O}(f)$ (nagy-ordó) és $o(f)$ (kis-ordó) fogalmakat.

Definíció 2.4. Legyen $f(n)$ és $g(n)$ két egész számokon értelmezett függvény. $g = \mathcal{O}(f)$, ha létezik s konstans és N_s index, hogy $\forall n > N_s : g(n) \leq s \cdot f(n)$.

Definíció 2.5. $f(n)$ és $g(n)$ továbbra is egészen értelmezett függvények, továbbá $g(n)$ csak véges sokszor veszi fel a 0 értéket. $f = o(g)$, ha $\frac{f(n)}{g(n)} \rightarrow 0$.

Másképp megfogalmazva bármilyen $s' > 0$ valós számra létezik $N_{s'}$ index, hogy $\forall n > N_{s'} : f(n) < s' \cdot g(n)$.

Próbáljuk $\mathcal{M}$ növekedését megbecsülni, ehhez vizsgáljuk a $g_1(x) - x$ függvényt. $g_1(x) - x = 2K(1 - \frac{1}{1+e^{-cx}}) = 2K(\frac{e^{-cx}}{1+e^{-cx}}) = 2Ke^{-cx}(\frac{1}{1-(-e^{-cx})}) = 2Ke^{-cx}(1 - e^{-cx} + e^{-2cx} - e^{-3cx} + \dots) = 2Ke^{-cx} + \mathcal{O}(e^{-c2x})$. Itt felhasználtuk, hogy $\frac{1}{1-y}$ Taylor-sora $|y| < 1$ esetén konvergens. Mivel $d_{ij}^n \rightarrow \infty$, így elég nagy n -re $d_{ij} > 0$, azaz $x > 0$. Ekkor $e^{-cx} < 1$, tehát használhatjuk a fenti becslést.

$x > G$ (G megfelelően nagy konstans) esetén $g_1(x) - x$ felülről becsülhető Ce^{-cx} -vel, ahol adott $C > 2K$. $g_1(\mathcal{M}^k) - \mathcal{M}^k$ éppen $\mathcal{M}^{k+1} - \mathcal{M}^k$, azaz azt adja meg, hogy $\mathcal{M}^k$ mennyit nőtt a k . lépésben. Tehát $G < k$ -ra ($k \in \mathbb{Z}$) $\mathcal{M}^{k+1} - \mathcal{M}^k \leq Ce^{-c\mathcal{M}^k}$.

Állítás 2.2. $\mathcal{M}^k = o(\sqrt{k})$

Bizonyítás 2.1. Ehhez lássuk be, hogy létezik $N_{s'} \in \mathbb{Z}^+$, hogy $\forall k > N_{s'} : \frac{\mathcal{M}^k}{\sqrt{k}} < s'$. Mivel az $f(x) = \sqrt{x}$ monoton növekvő és a deriváltja $f'(x) = \frac{1}{2\sqrt{x}}$ monoton csökken, ezért $\sqrt{k+1} - \sqrt{k}$ alulról becsülhető $f'(\sqrt{k+1})$ -gyel, tehát az $f(1), f(2), \dots$ sorozat k . és $k+1$. eleme között $\frac{1}{2\sqrt{k+1}}$ -vel nő. Mivel az $\mathcal{M}^k$ sorozat elég nagy k -ra már kevesebb, mint $Ce^{-c\mathcal{M}^k}$ -val nő, ezért bármilyen $s'' > 0$ -ra létezik olyan $N_{s''}$ index, hogy $\forall k > N_{s''}$ esetén $\mathcal{M}^k$ pozitív és $\frac{1}{Ce^{-c\mathcal{M}^k}} > s''$, azaz minden lépésben a $\sqrt{k}$ sorozat s'' -szor annyival nő, mint az $\mathcal{M}^k$. Mivel $\mathcal{M}^k \rightarrow \infty$, így van olyan N_{s^*} , hogy $\mathcal{M}^{N_{s^*}} > 2\mathcal{M}^{N_{s''}}$, azaz $\mathcal{M}^k$ megduplázódott az $N_{s''}$. és N_{s^*} . lépések között. Ezalatt $\sqrt{k}$ legalább $s'' \cdot \mathcal{M}^{N_{s''}}$ -sal nőtt, azaz $\sqrt{N_{s^*}} > s'' \cdot \mathcal{M}^{N_{s''}}$. Legyen A olyan konstans, amire $A\mathcal{M}^{N_{s''}} > \mathcal{M}^{N_{s^*}}$, ekkor $s'' = \frac{A}{s'}$ választással $\frac{\mathcal{M}^{N_{s^*}}}{\sqrt{N_{s^*}}} > \frac{A\mathcal{M}^{N_{s''}}}{s'' \cdot \mathcal{M}^{N_{s''}}} > \frac{A}{s''} = \frac{As'}{A} = s'$. Mivel a növekményekre továbbra is igaz, hogy $\sqrt{k}$ gyorsabban, $s'' > \frac{1}{s'}$ -szor annyival nő, mint $\mathcal{M}^k$, így a $\frac{\mathcal{M}^k}{\sqrt{k}}$ sosem emelkedik s' fölé.

Vezessük be az $r_k = \frac{\mathcal{M}^k}{\log k}$ sorozatot. Mivel $\mathcal{M}^k \rightarrow \infty$, ezért $r_k > 0$ elég nagy k esetén. A 2.1-es tétel bizonyításához elég lenne belátnunk, hogy $r_n \rightarrow \frac{1}{c}$.

Ehhez használjuk fel, hogy elég nagy k -ra $\mathcal{M}^k = o(\sqrt{k})$, valamint hogy $\mathcal{M}^{k+1} - \mathcal{M}^k \leq Ce^{-c\mathcal{M}^k}$. Rendezzük át az egyenleteket és helyettesítsük be r_k -t:

$$r_k = o\left(\frac{\sqrt{k}}{\log k}\right)$$

$$\begin{aligned} \frac{\log(k+1)}{\log k} r_{k+1} = \frac{\mathcal{M}^{k+1}}{\log k} &\leq \frac{1}{\log k} \left(\mathcal{M}^k + Ce^{-c\mathcal{M}^k} \right) \leq r_k + \frac{1}{\log k} \left(Ce^{-c\mathcal{M}^k} \right) \\ \frac{\log k}{\log(k+1)} r_k &\leq r_{k+1} \end{aligned}$$

Ezek segítségével határozzuk meg, hogy $r_{k+1} - r_k$ mikor pozitív és negatív:

$$\begin{aligned} r_{k+1} - r_k &\leq \left(1 - \frac{\log(k+1)}{\log k} \right) r_{k+1} + \frac{1}{\log k} \left(Ce^{-c\mathcal{M}^k} \right) = \\ &= \frac{1}{\log k} \left((\log k - \log(k+1)) r_{k+1} + Ce^{-c\mathcal{M}^k} \right) = \frac{1}{\log k} \left(-\log\left(\frac{k+1}{k}\right) r_{k+1} + Ce^{-c\mathcal{M}^k} \right) \end{aligned}$$

Ismert, hogy $\log\left(\frac{k+1}{k}\right)$ Taylor-sora $\frac{1}{k} - \frac{1}{2k^2} + \frac{1}{3k^3} - \dots$. Ezt behelyettesítve:

$$\begin{aligned} \frac{1}{\log k} \left(-\log\left(\frac{k+1}{k}\right) r_{k+1} + Ce^{-c\mathcal{M}^k} \right) &= \frac{1}{\log k} \left(-\frac{r_{k+1}}{k} + \mathcal{O}\left(\frac{r_{k+1}}{k^2}\right) + Ce^{-c\mathcal{M}^k} \right) \leq \\ &\leq \frac{1}{\log k} \left(-\frac{r_{k+1}}{k} + \mathcal{O}\left(\frac{\mathcal{M}^k + Ce^{-c\mathcal{M}^k}}{k^2 \log(k+1)}\right) + Ce^{-c\mathcal{M}^k} \right) \leq \\ &\leq \frac{1}{\log k} \left(-\frac{\log k}{\log(k+1)} \frac{r_k}{k} + \mathcal{O}\left(\frac{\sqrt{k}}{k^2}\right) + Ce^{-c\mathcal{M}^k} \right) \end{aligned}$$

Legyen $r_k \geq G_2$, ahol $G_2 \cdot c > 1$. Ekkor $\mathcal{M}^k \geq G_2 \cdot \log k$, azaz $e^{-c\mathcal{M}^k} \leq k^{-c G_2} < \frac{1}{k}$.

Ha $k > G$, akkor $\frac{\log k}{\log(k+1)}$ alulról becsülhető egy konstanssal: C_2 -vel.

Tehát elég nagy k és $r_k > G_2$ esetén $r_{k+1} - r_k < \frac{1}{\log k} \left(C \frac{1}{k} - C_2 \frac{r_k}{k} + \mathcal{O}\left(\frac{\sqrt{k}}{k^2}\right) \right) < 0$. Mivel az $r_{k+1} - r_k$ különbség ilyenkor $-\frac{1}{k \log k}$ -nel arányos, és $\sum_{l=k}^{\infty} \frac{1}{l \log l} = \infty$, így az (r_k) sorozat véges sok lépés után lecsökken G_2 alá. Mivel bármikor, amikor $r_k > \frac{1}{c}$, $r_{k+1} < r_k$ (a sorozat csökkenni kezd), mindig csak két szomszédos elem különbségével kerül G_2 fölé (bármely $G_2 > \frac{1}{c}$ -re). Az r_k sorozat felülről korlátos, azaz $|r_{k+1} - r_k| \leq \mathcal{O}\left(\frac{1}{k \log k}\right)$. Tehát $\limsup r_k \leq \frac{1}{c}$.

Hasonlóképpen, ha $r_k \leq G_2$, ahol $G_2 \cdot c < 1$, akkor $e^{-c\mathcal{M}^k} \geq k^{-c G_2} > \frac{1}{k}$. Tehát elég

nagy k és $r_k < G_2$ esetén $r_{k+1} - r_k > 0$, és az (r_k) sorozat véges sok lépés után eléri G_2 -t (bármely $G_2 < \frac{1}{c}$ -re). Így $\liminf r_k \geq \frac{1}{c}$, tehát bebizonyítottuk, hogy $r_k \rightarrow \frac{1}{c}$ és $\mathcal{M}^k \sim \frac{\log k}{c}$. $\square$

Végül lássuk be, hogy ha vesszük $b(d_{ij}^0), b(d_{ij}^1), \dots, b(d_{ij}^k)$ átlagát, az tart p_{ij} -hez.

Állítás 2.3.

$$\frac{1}{k} \sum_{l=0}^k b(d_{ij}^l) = p_{ij} + \mathcal{O}\left(\frac{\log(\log(k))}{k}\right)$$

Bizonyítás 2.2. Nézzük meg, hogy változott a pontszámok különbsége az első k mérkőzés alatt.

$$\begin{aligned} d_{ij}^k - d_{ij}^0 &= \sum_{l=1}^k (d_{ij}^l - d_{ij}^{l-1}) = \sum_{l=1}^k (2K (\mathbb{I}(i \text{ nyerte az } l. \text{ meccset}) - b(d_{ij}^{l-1}))) \\ d_{ij}^k - d_{ij}^0 &= 2K \left(\sum_{l=1}^k \mathbb{I}(i \text{ nyerte az } l. \text{ meccset}) - \sum_{l=1}^k b(d_{ij}^{l-1}) \right) \end{aligned} \quad (10)$$

A $\sum \mathbb{I}(i \text{ nyerte az } l. \text{ meccset})$ független, p_{ij} paraméterű Bernoulli-eloszlású változók összege. Legyen $X_l = \mathbb{I}(i \text{ nyerte az } l. \text{ meccset}) - p_{ij}$. Ekkor az $X_1, X_2, \dots$ változók várható értéke 0 és a szórása véges, azaz használhatjuk az iterált logaritmus tételt. Ebből azt kapjuk, hogy $\frac{1}{k} \sum_{l=1}^k \mathbb{I}(i \text{ nyerte az } l. \text{ meccset}) = p_{ij} + \mathcal{O}\left(\sqrt{\frac{\log(\log(k))}{k}}\right)$ majdnem mindenütt. Átrendezve az (10) egyenletet:

$$\frac{1}{k} \sum_{l=1}^k b(d_{ij}^{l-1}) = p_{ij} + \mathcal{O}\left(\sqrt{\frac{\log(\log(k))}{k}}\right) - \frac{d_{ij}^k - d_{ij}^0}{k \cdot 2K}$$

Mivel d_{ij}^0 konstans és tudjuk, hogy $d_{ij}^k < \max \mathcal{D}^k \sim \frac{\log(k)}{c}$, ezért a jobb oldalt a következőképpen becsülhetjük:

$$\frac{1}{k} \sum_{l=1}^k b(d_{ij}^{l-1}) = p_{ij} + \mathcal{O}\left(\sqrt{\frac{\log(\log(k))}{k}}\right) - \frac{\mathcal{O}(\log(k))}{k \cdot 2K} = p_{ij} + \mathcal{O}\left(\sqrt{\frac{\log(\log(k))}{k}}\right)$$

Megjegyzés. Vegyük észre, hogy a (d_{ij}^k) sorozat egy Markov-láncot alkot. Ha (d_{ij}^k) ergodikus Markov-lánc, akkor a fenti eredményből következik, hogy $\lim_{k \rightarrow \infty} b(d_{ij}^k)$ várható értéke p_{ij} .

2.5 Több dimenziós pontszámok

Eddig mindegyik modellünkben a versenyzők erősségeit egy számmal fejeztük ki. Felmerülhet a kérdés, hogy javíthatunk-e a modellek teljesítményén, ha mindenkinek több pontszámot is adunk, és ezeknek a függvényeként számoljuk ki a győzelem valószínűségét.

Egy kézenfekvő lehetőség az, hogy azt a fent bevezetett M mátrixot közelítsük egy alacsonyabb rangúval, ami a játékosok győzelmeinek relatív gyakoriságait tartalmazza. Feltételezzük, hogy a mérkőzések számának növekedésével az $M_{i,j}$ értékek tartanak a versenyzők győzelmének valószínűségéhez. Az M mátrixra igaz, hogy $M_{i,j} = 1 - M_{j,i}$, így átalakíthatjuk antiszimmetrikussá azzal, hogy minden értékéből kivonunk 0.5-öt. A kapott M' mátrixra tehát igaz, hogy $M' = -(M')^T$. Feltettük, hogy M főátlójában mindenhol 0.5 állt, ezért M' főátlójában minden érték 0.

Egy másik lehetőség, hogy az $M''_{i,j} = \frac{M_{i,j}}{M_{j,i}}$ mátrixot, vagy ennek a logaritmusát $M^l_{i,j} = \log(M_{i,j}) - \log(M_{j,i})$ -t szeretnénk közelíteni. Ez azért intuitív, mert ha a játékosok erősségei egy $\omega \in \mathbb{R}^n$ vektorral leírhatók lennének, akkor az M'' mátrix $M'' = \omega \times \frac{1}{\omega}$ alakban megadható lenne.

Vegyük észre, hogy M^l -ben $\log(M_{i,j}) - \log(M_{j,i}) = \log\left(\frac{M_{i,j}}{M_{j,i}}\right) = \log\left(\frac{t_{ij}}{t_{ji}}\right)$, tehát a nyert mérkőzések arányának a logaritmusai szerepelnek az M^l mátrixban. Ekkor is igaz, hogy M^l antiszimmetrikus, továbbá az eddigi arányok helyett itt a mátrix elemei lineárisan függenek a játékosok erősségeinek a logaritmusától. Ha két játékos közül minden mérkőzést ugyanaz nyerte, akkor M^l -ben $-\infty$ vagy ∞ szerepelne (hiszen $\frac{M_{i,j}}{M_{j,i}}$ nulla vagy $\frac{1}{0}$), ennek az elkerülése érdekében határozzunk meg egy $\varepsilon > 0$ küszöböt, hogy ha valamelyik i, j versenyzőpárra $M_{i,j} < \varepsilon$, akkor cseréljük ki az értékeket $M_{i,j} = \varepsilon$ -ra és $M_{j,i} = 1 - \varepsilon$ -ra. ε -t választhatjuk a legnagyobb $\frac{M_{i,j}}{M_{j,i}}$ arány konstansszorosának.

Egy mátrix alacsonyabb rangú közelítését szinguláris érték felbontás segítségével határozzuk meg.

Definíció 2.6. Az A mátrix szinguláris érték (SVD) felbontása USV^T , azaz A -t három mátrix szorzataként írjuk fel, melyekre U és V ortonormált bázist alkotnak, és S egy diagonális mátrix, aminek minden eleme nemnegatív. S főátlójában lévő elemeit nevezzük szinguláris értékeknek, míg U és V oszlopait jobb-, illetve baloldali

szinguláris vektoroknak.

Minden mátrixnak létezik SVD felbontása, és a szinguláris értékei egyértelműek.

Ha azt az $\widetilde{M}$ mátrixot keressük, aminek a rangja r és $\|M - \widetilde{M}\|_F$ a lehető legkisebb, akkor a megoldás éppen $\widetilde{M} = U_r S_r V_r^T$, ahol S_r az r legnagyobb szinguláris értékből alkotott diagonális mátrixot, valamint U_r és V_r ezekhez a legnagyobb értékekhez tartozó r jobboldali és baloldali szinguláris vektort jelentik. Tehát SVD felbontással meghatározhatjuk egy mátrix alacsony rangú közelítését. Minden i játékosra megjegyezve az $(U_{i,1}, \dots, U_{i,r}, V_{i,1}, \dots, V_{i,r})$ számokat, $S_1, \dots, S_r$ ismeretében ezekből visszakaphatjuk az $\widetilde{M}_{i,j}$ értékeket, hiszen $\widetilde{M}_{i,j} = \sum_{k=1}^r U_{i,k} S_k V_{j,k}$.

A mi gyakorlati példánkban ritka a mérkőzések mátrixa, amit közelíteni szeretnénk, ami probléma, hiszen szinguláris érték felbontás alkalmazásakor a hiányzó elemek helyén álló nullákra is illeszkedni próbálunk az alacsonyabb rangú közelítéssel. Keressünk olyan módszert, amivel a hiányzó értékektől el tudunk tekinteni, avagy pótolni tudjuk őket úgy, hogy a rang minél kisebb legyen. Vezessük be a $P_W(M)$ jelölést, ahol $P_W(M)$ egy M -mel azonos méretű mátrix, amire $P_W(M)_{i,j} = M_{i,j}$, ha $W_{i,j} > 0$, azaz i és j már játszottak egymás ellen, egyébként $P_W(M)_{i,j} = 0$. Ekkor az általános feladat megfogalmazása: minimalizáljuk a $\text{rang}(\widetilde{M})$, miközben $P_W(\widetilde{M}) = P_W(M^l)$. Ez egy NP -nehéz feladat, ezért valamiképpen relaxálnunk kell.

A legismertebb megközelítés, hogy $\widetilde{M}$ rangja helyett az $\widetilde{M}$ nukleáris normáját minimalizáljuk, ahol az A mátrix nukleáris normája a szinguláris értékeinek az összege. Ezt $\|A\|_*$ -gal jelölve a feladat a következő:

$$\begin{aligned} \min \|\widetilde{M}\|_* \\ \text{ahol } P_W(\widetilde{M}) = P_W(M) \end{aligned}$$

Nekünk az a célunk, hogy $\widetilde{M}$ egy alacsony rangú közelítése legyen a mátrixnak, ezért nem várjuk el, hogy M minden ismert értékére pontosan illeszkedjen az $\widetilde{M}$, csak azt hogy közel legyen a kettő. Ezért megadunk egy r rangot és olyan r rangú $\widetilde{M}^*$ mátrixot keresünk, hogy $\|P_W(\widetilde{M}^*) - P_W(M)\|_F^2$ minimális legyen.

A jól közelítő mátrix megtalálásához ALS (Alternating Least Squares) módszert használunk. Az $\widetilde{M}$ -ot UV^T alakban keressük, a célunk olyan U^* és V^* , amire $\|P_W(U^*(V^*)^T) - P_W(M)\|_F^2$ minimális, azaz ahol $\sum_{i,j:W_{i,j}>0} (U_i^* V_j^{*T} - M_{i,j}^l)^2$ a lehető legkisebb.

Vegyük észre, hogy a célfüggvény általánosan nem konvex, azonban ha U vagy V egyikét rögzítjük, akkor már konvex optimalizációs problémát kapunk. $U = A$ rögzítése esetén a feladatunk éppen az $AV \approx M^l$ többváltozós lineáris regresszióban a V becslése úgy, hogy M^l ismert értékeire a négyzetes eltérés a legkisebb legyen. Az ALS módszer tehát a következő:

1. Inicializáljuk $U^{(0)} \in \mathbb{R}^{n \times r}$ mátrixot, mint $P_W(M^l)$ SVD felbontásában az r legnagyobb szinguláris értékhez tartozó baloldali szinguláris vektorokat. i értéke legyen 1.

2. Rögzítjük $U^{(i-1)}$ -et és meghatározzuk $V^{(i)}$ -t a következőképpen:

$$V^{(i)} = \arg \min_V \|P_W(U^{(i-1)}V) - P_W(M^l)\|_F$$

3. Rögzítjük $V^{(i)}$ -t és meghatározzuk $U^{(i)}$ -t a következőképpen:

$$U^{(i)} = \arg \min_U \|P_W(UV^{(i)}) - P_W(M^l)\|_F$$

4. Legyen $U^{(i)} = \frac{U^{(i)}}{\|U^{(i)}\|_{\max}}$. Ha nem értük el az előre megadott *maxiter* iterációs számot, akkor $i = i + 1$ és visszatérünk a 2. ponthoz.

Az algoritmus után minden játékosra megjegyezzük a kapott U és V mátrixok megfelelő sorát, ezekből rekonstruálható UV minden eleme. Az így kapott alacsonyabb rangú mátrix struktúrája azonban nem lesz feltétlen azonos, mint az eredeti M -é. Már említettük, hogy $r = 1$ esetén, amikor M^l -t két diád szorzatával próbáljuk közelíteni, akkor a mátrixot UV^T alakban várjuk, ahol $U \sim \omega$ és $V \sim \frac{1}{\omega}$. Ekkor valóban teljesül, hogy $UV_{i,j}^T = U_i V_j = \frac{1}{V_i} \frac{1}{U_j} = \frac{1}{UV_{j,i}^T}$. Úgy tudjuk elérni, hogy az algoritmusban végül $U = \frac{1}{V}$ -t kapjunk, minden iteráció végén az új $U^{(i)}$ -t és $V^{(i)}$ -t megváltoztatjuk a következőképpen: $U^{(i)} = \frac{U^{(i)} + \frac{1}{V^{(i)}}}{2}$ és $V^{(i)} = \frac{2}{U^{(i)} + \frac{1}{V^{(i)}}}$.

M^l közelítésekor nem tudunk ilyen egyszerű megkötést arra, hogy V milyen függvénye legyen U -nak, hogy a szorzatuk megfelelő struktúrájú legyen. Azonban az itt is igaz, hogy ha a kapott $\widetilde{M}$ mátrix helyett vesszük $\widetilde{\widetilde{M}} = \frac{\widetilde{M} - \widetilde{M}^T}{2}$ -t, akkor egy antiszimmetrikus mátrixot kapunk, amire $\|P_W(\widetilde{M}) - P_W(M^l)\|_F^2 \geq \|P_W(\widetilde{\widetilde{M}}) - P_W(M^l)\|_F^2$. Ezt lássuk be: minden olyan i, j párra, ahol $\widetilde{\widetilde{M}}_{i,j} = -\widetilde{\widetilde{M}}_{j,i}$ a mátrix változatlan, azaz $\widetilde{\widetilde{M}}_{i,j} = \widetilde{M}_{i,j} = -\widetilde{M}_{j,i} = -\widetilde{\widetilde{M}}_{j,i}$. Ha valahol $\widetilde{M}_{i,j} = M_{i,j}^l + a$ és $\widetilde{M}_{j,i} = M_{j,i}^l + b$ $a \neq b$ valós számokra, akkor a Frobenius-normában ezek az elemek $(\widetilde{M}_{i,j} - M_{i,j}^l)^2 + (\widetilde{M}_{j,i} - M_{j,i}^l)^2 =$

$a^2 + b^2$ -ként jelennek meg. Ha $\widetilde{M}$ -et $\widetilde{\widetilde{M}}$ -re cserélnénk, akkor ugyanezek a tagok:

$$\begin{aligned} (\widetilde{\widetilde{M}}_{i,j} - M_{i,j}^l)^2 + (\widetilde{\widetilde{M}}_{j,i} - M_{j,i}^l)^2 &= \left(\frac{\widetilde{M}_{i,j} - (\widetilde{M}^T)_{i,j}}{2} - M_{i,j}^l\right)^2 + \left(\frac{\widetilde{M}_{j,i} - (\widetilde{M}^T)_{j,i}}{2} - M_{j,i}^l\right)^2 = \\ &= \left(\frac{\widetilde{M}_{i,j} - \widetilde{M}_{j,i}}{2} - M_{i,j}^l\right)^2 + \left(\frac{\widetilde{M}_{j,i} - \widetilde{M}_{i,j}}{2} - M_{j,i}^l\right)^2 = \left(\frac{\widetilde{M}_{i,j} + \widetilde{M}_{i,j} - a - b}{2} - M_{i,j}^l\right)^2 + \left(\frac{\widetilde{M}_{j,i} + \widetilde{M}_{i,j} - a - b}{2} - \right. \\ &\left. M_{j,i}^l\right)^2 = 2\left(\frac{a+b}{2}\right)^2, \text{ ahol felhasználtuk, hogy } \widetilde{M}_{i,j} + \widetilde{M}_{j,i} = a + b. \end{aligned}$$

Ismert azonosság, hogy $a^2 + b^2 \geq 2\left(\frac{a+b}{2}\right)^2$, tehát beláttuk, hogy $\|P_W(\widetilde{M}) - P_W(M^l)\|_F^2$ nem nőtt, amikor $\widetilde{M}$ -ot $\widetilde{\widetilde{M}}$ -ra cseréltük.

3 Modellek kiértékelése

Keressünk mérőszámokat arra, hogy melyik modellel kapott pontszámok fedik a legjobban a valóságot.

Ehhez vezessünk be néhány fogalmat:

Definíció 3.1. Legyen $\mathcal{X}$ és $\mathcal{Y}$ a lehetséges bemenetek és kimenetek halmaza. Az $f : \mathcal{X} \rightarrow \mathcal{Y}$ függvények modellek. Gyakran $\mathcal{X} = \mathbb{R}^d$ és $\mathcal{Y} = \mathbb{R}$.

Definíció 3.2. A $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ bemenet-kimenet párok véges sorozatát mintának nevezzük. A veszteségfüggvény $L : \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ megadja, hogy (y_i) kimenetek és $\hat{y}_i$ prediktált kimenetek esetén mennyire volt jó vagy rossz a modellünk, azaz mennyire találta el a célváltozót. Minél kevésbé találtuk el a valós kimeneteket, annál nagyobb a veszteségfüggvény értéke.

Definíció 3.3. A kockázat a veszteségfüggvény várható értéke, azaz $R = \mathbb{E}(L(f(X), Y))$. Az a célunk, hogy a kockázatot minimalizáljuk.

Határozzuk meg, hogy milyen veszteségfüggvényt használjunk a modellünk jóságának mérésére. Az egyik módszer, hogy vesszünk meccseket és leszámoljuk, hogy hányat találunk el, ha mindig arra a sportolóra tippelünk, akinek nagyobb a pontszáma, azaz aki az erősebb.

Mivel a fent taglalt pontszámítási szabályok esetén tudunk valószínűségeket is számolni, ezért keressünk olyan függvényt, amivel a prediktált valószínűségeket értékelhetjük.

3.0.1 Valószínűségek kiértékelése

Tegyük fel, hogy minden (i, j) versenyzőpárra az első játékos állandó $q_{i,j}$ valószínűséggel nyeri meg a mérkőzéseket. Abban az esetben, ha a meccseket, amiken kiértékelünk, egy rövidebb időintervallumból választjuk, akkor ez az állítás hihető.

Tegyük fel, hogy az m mérkőzés között, amiken kiértékelünk, minden mérkőzés azonos valószínűséggel fordul elő.

Legyenek az éppen vizsgált m mérkőzés játékosai $(i_1, j_1), (i_2, j_2), \dots, (i_m, j_m)$ úgy, hogy $\forall k : x_{i_k} > x_{j_k}$, azaz mindig az az első, akinek a győzelmét valószínűbbnek

tartjuk. Emellett $y_i = \mathbb{I}(x_i \text{ nyerte a mérkőzést})$ és legyen $\forall k : p_i = p(x_{i_k}, x_{j_k})$ az általunk jóslott valószínűség, hogy a k . meccset x_{i_k} nyeri.

Olyan $L((i_1, j_1, y_1), (i_2, j_2, y_2), \dots, (i_m, j_m, y_m), (p_1, p_2, \dots, p_m))$ veszteségfüggvényt keresünk, aminek a várható értéke éppen akkor a legkisebb, ha $\forall k : p_k = q_{i_k, j_k}$, tehát ha éppen eltaláltuk a győzelmek valószínűségét.

Legyen $L((i_1, j_1, y_1), (i_2, j_2, y_2), \dots, (i_m, j_m, y_m), (p_1, p_2, \dots, p_m)) = \sum_{k=1}^m L((i_k, j_k), y_k, p_k)$ és keressük L -et $L((i_k, j_k), y_k, p_k) = g_2(y_k) \cdot p_k^2 + g_1(y_k) \cdot p_k + g_0(y_k)$ alakban, azaz a veszteségfüggvény függ attól, hogy jól tippeltünk-e, illetve az általunk előrejelzett valószínűségek legfeljebb másodfokú függvénye. Mivel a várható érték a meccseken összeadódik, így elég az $m = 1$ esetet vizsgálnunk és olyan L -t keresnünk, amire a kockázat akkor minimális, ha $p_{i,j} = q_{i,j}$.

A veszteségfüggvény feltételes várható értéke ekkor az (i, j) párra:

$$\begin{aligned} \mathbb{E}(L((i_k, j_k), y_k, p_k) | i_k = i, j_k = j) &= L((i, j), 1, p_{i,j}) \cdot q_{i,j} + L((i, j), 0, p_{i,j}) \cdot q_{j,i} = \\ &= (g_2(1) \cdot p_{i,j}^2 + g_1(1) \cdot p_{i,j} + g_0(1)) \cdot q_{i,j} + (g_2(0) \cdot p_{i,j}^2 + g_1(0) \cdot p_{i,j} + g_0(0)) \cdot q_{j,i} \end{aligned}$$

Mivel a mérkőzés során i és j közül pontosan az egyikük nyer, így $q_{i,j} = 1 - q_{j,i}$.

Ezt helyettesítsük be:

$$\begin{aligned} &(g_2(1) \cdot p_{i,j}^2 + g_1(1) \cdot p_{i,j} + g_0(1)) \cdot q_{i,j} + (g_2(0) \cdot p_{i,j}^2 + g_1(0) \cdot p_{i,j} + g_0(0)) \cdot (1 - q_{i,j}) = \\ &= ((g_2(1) - g_2(0)) \cdot q_{i,j} + g_2(0)) \cdot p_{i,j}^2 + ((g_1(1) - g_1(0)) \cdot q_{i,j} + g_1(0)) \cdot p_{i,j} + (g_0(1) - g_0(0)) \cdot q_{i,j} + g_0(0) \end{aligned}$$

A minimum megtalálásához vegyük az első és második deriváltakat. Az első derivált legyen 0 a második pedig nemnegatív:

$$\begin{aligned} &\mathbb{E}(L((i_k, j_k), y_k, p_k) | i_k = i, j_k = j)' = \\ &= 2((g_2(1) - g_2(0)) \cdot q_{i,j} + g_2(0)) \cdot p_{i,j} + (g_1(1) - g_1(0)) \cdot q_{i,j} + g_1(0) = 0 \\ &\mathbb{E}(L((i_k, j_k), y_k, p_k) | i_k = i, j_k = j)'' = 2(g_2(1) - g_2(0)) \cdot q_{i,j} + 2g_2(0) \end{aligned}$$

Akkor veszi fel a függvény a minimumot, ha $q_{i,j} = p_{i,j}$, $\forall q_{i,j} : 0 \leq q_{i,j} \leq 1$, azaz

$$\begin{aligned} 0 &= 2((g_2(1) - g_2(0)) \cdot q_{i,j} + g_2(0)) \cdot q_{i,j} + (g_1(1) - g_1(0)) \cdot q_{i,j} + g_1(0) = \\ &= 2(g_2(1) - g_2(0)) \cdot q_{i,j}^2 + (g_1(1) - g_1(0) + 2g_2(0)) \cdot q_{i,j} + g_1(0) \end{aligned}$$

$q_{i,j}$ polinomja éppen akkor konstans 0, ha minden együttható 0, tehát

- $g_1(0) = 0$
- $g_1(1) - g_1(0) + 2g_2(0) = 0 \implies 2g_2(0) = -g_1(1)$

$$\bullet \quad g_2(1) - g_2(0) = 0 \implies 2g_2(1) = 2g_2(0) = -g_1(1)$$

Legyen mivel minden megfelelő polinomban tér csak el, a minimumhelyük ugyanott van, így megválaszthatjuk $g_2(0)$ -át 1-nek. Ekkor a kapott veszteségfüggvény

$$L((i_k, j_k), y_k, p_k) = \begin{cases} p_k^2 & , \text{ ha } y_k = 1, \text{ azaz jól tippeltünk} \\ (1 - p_k)^2 & , \text{ ha } y_k = 0, \text{ azaz rosszul tippeltünk} \end{cases}$$

A valószínűségek ismeretében már nem csak meccsek, hanem versenyek eredményeit is képesek vagyunk előrejelezni. Például egy egyenes kieséses versenyen, a sorsolás ismeretében minden mérkőzésre meg tudjuk határozni, hogy melyik versenyzőnek mennyi esélye van eljutni odáig és megnyerni azt, azaz felírható, hogy kit mekkora valószínűséggel várunk a dobogóra vagy az első helyre. Tehát azt is használhatjuk a modellünk helyességének egyik mérőszámának, hogy eltaláltuk-e a verseny dobogósait, illetve milyen valószínűséggel tippeltük a valódi nyerteseket a dobogóra.

4 Alkalmazás

Vizsgáljuk meg, hogy milyen eredményt adnak a fenti modelleink valós adatokon. Ehhez a már korábban említett dzsúdómérkőzéseket tartalmazó adathalmazt fogjuk használni. Mivel az adatok nagy időintervallumból származnak (van, amelyik több mint 20 éves), exponenciális súlyozást alkalmazunk úgy, hogy az egy éve történt mérkőzések súlya fele akkora, mint egy mai meccs.

A dzsúdóversenyeket ritkán szervezik körmérkőzéses formában, ehelyett rendszerint egyenes kieséses lebonyolításúak, teljes- vagy negyeddöntőktől kezdődő vigaszággal. Emiatt a mérkőzések gráfja nem erősen összefüggő, tehát a Bradley–Terry-modell esetén a MAP-beclést tudjuk rá alkalmazni.

A módszerek összehasonlításánál a modelleket a 2024. évi nyári olimpia előtti egy év mérkőzésein teszteltük. Azért választottuk ezt az intervallumot, mert az olimpiára készülő versenyzők ritkán váltanak súlycsoportot az utolsó évben, illetve ekkor már nem történt szabálymódosítás sem, így ezek nem befolyásolták a méréseinket. Vettük az összes 2023. július 26. előtti mérkőzést, és ezekre futtattuk le a három pontszámító algoritmust. Először nézzük meg a futásidőket. Az adatok beolvasása és előkészítése mindegyik esetben 5-10 másodpercet vett igénybe. Az Élő-rangsorolás, ha a meccseket már sorbarendeztük időpont szerint, kevesebb, mint 1 másodpercet vesz igénybe. A lineáris modellben és a több dimenziós közelítésnél is egy nagy – közel 50000 sorból és oszlopból álló –, feltehetően ritka mátrixszal dolgoztunk. Ahhoz, hogy ekkora mátrixot kezelni tudjunk, a pont ilyen helyzetekre kitalált SciPy.sparse Python-csomagot használtunk.

A lineáris modellben egy közel 50000 sorból és változóból álló egyenletrendszert kell megoldanunk minél gyorsabban. Erre léteznek iteratív módszerek, mi a *gmres* algoritmust hívtuk meg. Így a lineáris pontozás futásideje (a beolvasás után) 15 másodperc volt. A több dimenziós modell esetén az ALS algoritmust implementáltuk, hogy megkapjuk a versenyzők erőviszonyának logaritmusát tartalmazó mátrix 5 rangú közelítését. Próbálkoztunk nagyobb, például 10, illetve 50 rangú közelítéssel is, azonban egyre több iterációra volt szükség a konvergenciához, és a teljesítmény nem javult lényegesen. Az algoritmus során egy iteráció közel egy percet vett igénybe, azaz az általunk lefuttatott 90 iterációval másfél óra alatt végeztünk.

A Bradley–Terry-modell szerinti pontszámok számolásánál az volt a megállási

feltételünk, hogy a pontszámok abszolút változása a legújabb iterációban összesen kevesebb, mint 0,1. Azt tapasztaltuk, hogy ahhoz, hogy a pontszámok változása 1-nél kevesebb legyen, nagyjából 7000, míg az algoritmus leállásáig 9000 iterációra volt szükség. Így a pontszámok kiszámításához összesen 30 percre volt szükség.

Ezután vettük az olimpia előtti utolsó évből és az olimpiáról a mérkőzéseket, és a versenyeket körülbelül 4000 mérkőzésenként csoportosítva kiértékeljük a három modellt. A csoportokat az időrendben egymást követő versenyekből alakítottuk ki, a lent látható táblázatban a csoporthoz tartozó utolsó verseny időpontja szerepel. Minden csoportra megszámoltuk, hogy hány olyan meccs volt, ahol valamelyik versenyzőt nem ismertük, és a többi mérkőzésen vizsgáltuk a predikciónk eredményét. Az alábbi táblázatban látható, hogy a különböző modellekkel milyen arányban találtuk el helyesen a mérkőzések eredményét, illetve hogy mennyi a 3. szekcióban leírt veszteségfüggvény értéke. Az utolsó oszlopba írtuk viszonyítási alapként, hogy mennyi lenne a veszteségfüggvény, ha a modellünk minden játékost ugyanolyan erősnek prediktál, azaz minden mérkőzésnél 0,5 valószínűséggel nyernek mindketten.

			Találati arány				Veszteségfüggvény				
Időpont	Összes meccs	Hiányos meccsek	Lineáris modell	Élő-pontok	Bradley-Terry	Több dimenzió	Lineáris	Élő-pontok	Bradley-Terry	Több dimenzió	50%
23.aug	4142	948	0.688	0.626	0.695	0.586	279.295	174.931	175.526	292.578	798.50
07.okt	4091	305	0.715	0.645	0.725	0.613	347.961	134.124	227.969	136.972	946.50
11.nov	4246	969	0.679	0.644	0.693	0.601	282.117	153.736	170.184	312.287	819.25
02.feb	4146	1077	0.712	0.639	0.711	0.622	264.730	106.425	172.366	324.599	767.25
01.márc	4106	1256	0.671	0.617	0.679	0.589	255.195	117.338	164.122	368.250	712.50
23.márc	4228	1223	0.695	0.652	0.695	0.569	254.203	129.901	156.674	358.168	751.25
13.ápr	4221	1899	0.677	0.623	0.680	0.602	187.122	118.113	119.486	521.022	580.50
03.máj	4388	1424	0.684	0.614	0.694	0.592	245.631	134.140	150.472	408.581	741.00
25.máj	4459	1974	0.684	0.623	0.703	0.592	209.312	112.943	126.558	535.073	621.25
27.júl	4083	1336	0.692	0.653	0.697	0.565	212.917	124.254	126.695	385.579	686.75
09.aug	4072	1248	0.678	0.625	0.678	0.546	235.985	150.080	143.343	354.713	706.00
16.aug	3661	1949	0.709	0.606	0.693	0.539	113.014	156.697	39.620	512.185	428.00
Összes	49843	15608	0.691	0.632	0.696	0.587	2887.482	1612.684	1773.014	4510.007	8558.75

Ahogy azt a táblázatból látjuk, a mérkőzések kimenetelét a Bradley–Terry-modell és a lineáris modell lényegesen jobban eltalálja, mint az Élő-pontozás vagy a többdimenziós közelítés. Meglepő, hogy a lineáris modell egyszerűsége ellenére alig maradt el a széles körben használt Bradley–Terry-modell teljesítményétől. Ami a valószínűségeket illeti, ott az Élő-modell a legjobb, a Bradley–Terry-modell a második, végül a lineáris és a többdimenziós modellek adják a legkevésbé pontos becsléseket.

Irodalomjegyzék

- [1] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. “A Singular Value Thresholding Algorithm for Matrix Completion”. In: *SIAM Journal on Optimization* 20.4 (2010), pp. 1956–1982.
- [2] Arpad E. Elo. *The Rating of Chessplayers, Past and Present*. New York: Arco Publishing, 1978, pp. 1–35. ISBN: 0668047216.
- [3] Adrien Krifa, Florian Spinelli, and Stéphane Junca. *On the convergence of the Elo rating system for a Bernoulli model and round-robin tournaments*. Research Report hal-03286065. Université Côte d’Azur, 2021.
- [4] Jr. L.R. Ford. “Solution of a Ranking Problem from Binary Comparisons”. In: *Mathematische Zeitschrift* 64.8 (1957), pp. 28–33.
- [5] Tal Levy, Alireza Vahid, and Raja Giryes. “Ranking Recovery from Limited Comparisons using Low-Rank Matrix Completion”. In: *arXiv preprint arXiv:2102.07054* (2021).
- [6] M. E. J. Newman. “Efficient Computation of Rankings from Pairwise Comparisons”. In: *Journal of Machine Learning Research* 24.238 (2023), pp. 1–25.
- [7] Reza Zadeh et al. *CME 323: Distributed Algorithms and Optimization, Spring 2015, Lecture 14*. <http://stanford.edu/~rezab/dao>. Stanford University, scribed lecture notes. May 2015.
- [8] Ernst Zermelo. “Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung”. In: *Mathematische Zeitschrift* 29 (1929), pp. 436–460.

Alulírott Jahán Anita Dalma nyilatkozom, hogy szakdolgozatom elkészítése során az alább felsorolt feladatok elvégzésére a megadott MI alapú eszközöket alkalmaztam:

Feladat	Felhasznált eszköz	Felhasználás helye	Megjegyzés
Irodalomkeresés	GPT-4o	2.5 fejezet	
Adatfeldolgozás	Gemini 2.5 Flash	2.2 fejezet	saját C++ kód átalakítása Python kóddá
Nyelvhelyesség ellenőrzése	Writefull	Teljes dolgozat	

A felsoroltakon túl más MI alapú eszközt nem használtam.