A Bradley-Terry modell elemzése

Szakdolgozat

Készítette: Bókkon Andrea
Témavezető: Csiszár Villő, adjunktus

Matematika B.Sc., Matematikai elemző szakirány

Valószínűségelméleti és Statisztika Tanszék

Eötvös Loránd Tudományegyetem
Természettudományi Kar
2010

Tartalomjegyzék

1. A szakdolgozat témája és felépítése

1.1. Bevezetés

Hogy miről is szól a szakdolgozatom? Mit takar a cím? Azt szeretném közelebbről bemutatni.

Sportrajongók és lelkes fogadók tudják, hogy a mérkőzések előtt mindig megjósolják, hogy ki az esélyesebb, egyik csapat, vagy versenyző mennyivel jobb a másiknál, mi az előzetesen várt eredmény, esetleg mennyi a gól, illetve pontkülönbség.

Bonyolítja a helyzetünket, ha egyik csapat/versenyző hazai pályán játszik. Lehet, hogy az ellenfelet kiáltják ki előzetesen esélyesnek, ám az esélytelenebb versenyző hazai pályán jobban teljesít. Ekkor a *hazai pálya előnyé*ről beszélünk. De van, amikor hátránnyá is tud válni az otthoni helyszín. Ekkor azt mondjuk, hogy a *hazai pálya hátrányá*ról van szó. De nem csak azt nézhetjük, hogy ki nyer, vagy veszít, hanem a döntetlenekre is kiterjesztjük a modellünket, akár a hazai pályán, akár idegenben.

Hogy mindezt egy való életből vett példára - matematikai algoritmusok felhasználásával, statisztikai modellek illesztésével, továbbá az R programcsomag felhasználásával - hogyan határozhatjuk meg, arról szól a szakdolgozatom a továbbiakban.

1.2. A szakdolgozat felépítése

Ahhoz, hogy a modellt a későbbiekben bevezethessük és megérthessük, előzetesen ismernünk kell pár statisztikai becslést, illetve algoritmust.

A szakdolgozatom első felében vázolom a *maximum-likelihood becslés*t, majd bemutatom az *EM- és az MM-algoritmus*t, melyek elengedhetetlenek a továbbiakban.

A szakdolgozatom főtémájában részletesen leírom a modellt, majd kitérek a modell általánosításaira, kiterjesztéseire is. Vizsgálom a *hazai pálya előnyé*t, és a *döntetlen* eseteit is; a modellt, és az algoritmusok kapcsolatát.

Majd valós sportesemények eredményeit elemzem az R, statisztikai program segítségével, a BradleyTerry modell nevű installált programcsomag felhasználásával, s ebből vonok le konklúziókat.

2. Felhasznált eszközök

2.1. Maximum-likelihood becslés (ML)

A momentumok módszerén kívül a pontbecslés másik módszere. A maximális valószínűség angolul: *maximum-likelihood*, tehát az $L = L(k; \lambda)$ likelihood függvény maximumát keressük.

Általánosítva: Ismerjük a sokaság eloszlását, de nem ismerjük az eloszlást jelző paramétert vagy paramétereket. A paraméter vagy paraméterek értékét olyan értékkel vagy értékekkel becsüljük, amely vagy amelyek esetén az adott minta bekövetkezése lenne a legnagyobb valószínűségű. A maximális valószínűséget az adott minta valószínűségét megadó likelihood-függvény maximumával vagy a logaritmusának a maximumával keressük meg.

2.1. Definíció. *. Legyen $X_1, ..., X_n$ minta F_ϑ eloszlásból, $\vartheta \in \theta$. Ekkor a ϑ maximum likelihood (ML) becslése $\hat{\vartheta}$, ha $L_n(X; \hat{\vartheta}) = \max\{L_n(X; \vartheta) : \vartheta \in \theta\}$.*

Ha ez nem egyértelmű, vagy nem létezik, de $L_n(X; \vartheta)$ "elég sima", akkor a $\frac{\partial}{\partial\vartheta}\ln L_n(X; \vartheta) = 0$ *likelihood-egyenlet* megoldására vagyunk kíváncsiak.

A maximum-likelihood becslés az egyik legelterjedtebb módszer a gyakorlatban. Bár, a becslés általában nem torzítatlan, bizonyos erős feltételek mellett jó aszimptotikus tulajdonságai vannak.

2.2. Tétel. *Bizonyos (erős) regularitási feltételek mellett elég nagy n -re a $\hat{\vartheta}_n$ ML becslés létezik, és konzisztens. Egyes esetekben:*

Aszimptotikusan normális eloszlású: $\sqrt{n} \cdot (\hat{\vartheta}_n - \vartheta) \to N(m(\vartheta), \sigma(\vartheta))$, $(n \to \infty)$

Aszimptotikusan torzítatlan: $m(\vartheta) = 0$

Aszimptotikusan optimális: $\sigma^2(\vartheta) = \frac{1}{I_1(\vartheta)}$.

Példa maximum-likelihood becslésre

Egy fonalgyárban a fonalak szakadását vizsgáljuk. A fonalak szakadása egymástól független. Kimutatható, hogy ebben az esetben egy adott időtartam alatt a fonalszakadások száma: X jó közelítésben *Poisson-eloszlású.*

Az ismeretlen λ paraméterre célszerű olyan értéket választani, amely esetén

$$X = k$$

esemény valószínűsége maximális. (Ha az adott időtartamban a fonalszakadások száma $X = k$ volt.)

Tehát, mint fent említettem, a $L = L(k, \lambda)$ likelihood-függvény maximumát keressük.

Vizsgáljuk, ha n mérési intervallumban nézzük a fonalszakadások számát!

$X_1 = k_1, X_2 = k_2, ..., X_n = k_n$,

Azt keressük, hogy milyen λ paraméterérték esetén maximális minta valószínűsége. Mivel a fonalszakadások függetlenek, ezért:

$$L = L(k_1, k_2, ..., k_n; \lambda) = P(X_1 = k_1, X_2 = k_2, ...X_n = k_n) =$$

$$= P(X_1 = k_1)P(X_2 = k_2)...P(X_n = k_n) = \prod_{i=1}^{n} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{k_i}}{k_i!}$$

.

Egyszerűbb, ha L helyett ln L maximumát keressük.

$$\ln L = -n\lambda + \sum_{i=1}^{n} (k_i \ln \lambda - \ln k_i!)$$

$$\frac{d \ln L}{d\lambda} = -n + \sum_{i=1}^{n} \frac{k_i}{\lambda} = 0$$

$$\lambda = \frac{\sum_{i=1}^{n} k_i}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{x}$$

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{1}{\lambda^2} \sum k_i < 0$$

Célszerű a λ paramétert $\bar{x}$-sal becsülni, mert $\lambda = \bar{x}$ esetén maximális annak a valószínűsége, hogy az $X_1 = k_1, X_2 = k_2, ..., X_n = k_n$ mintát kapjuk. A λ paraméter maximum-likelihood becslése tehát a mintaátlag.

Megjegyzés.

Egy T paraméter esetén a likelihood-függvény a következő:

I. *Diszkrét eset:*

$L(X_1, X_2, ..., X_n, T) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) =$
$= P(X_1 = x_1)P(X_2 = x_2)...P(X_n = x_n)$.

Szorzat helyett (néha) könnyebb összeget kezelni. Ekkor az

$$\ln L = \sum_{i=1}^{n} P(X_i = x_i)$$

függvényt tekintjük a log-likelihood függvénynek.

II. *Folytonos eset:* egy pont felvételének valószínűsége az n-dimenziós térben 0. Annak a valószínűségét kell maximalizálni, hogy a pont az $(x_1, x_2, ..., x_n)$ pont közvetlen környezetébe, illetve pontosabban az

$$x_1 \leq X_1 \leq x_1 + \Delta x_1, ..., x_n \leq X_n \leq x_n + \Delta x_n$$

n-dimenziós téglatestbe esik. Ennek valószínűsége:

$$f(x_1, T)f(x_2, T)...f(x_n, T)\Delta x_1 \Delta x_2 ... \Delta x_n.$$

Ez ott maximális, ahol $L(x_1, x_2, ..., x_n; T) = f(x_1, T)f(x_2, T)...f(x_n, T)$ függvény maximális.

Ezt, vagy ennek a logaritmusát, vagyis az ln $L = \sum_{i=1}^{n}$ ln $f(x_i, T)$ tekintjük likelihood-függvénynek.

Ennek megfelelően: $\frac{dL}{dT} = 0$ vagy ha a logaritmusát nézzük, akkor: $\frac{d \ln L}{dT} = 0$

A parciális deriváltak zérushelyeit keressük. Itt lehet a maximum. (Ha van.) Hogy van-e, vagy nincs, azt a Hesse-determinánssal tudjuk eldönteni.

2.2. Az EM-algoritmus

Hiányos megfigyelések esetén alkalmazzuk ezt az algoritmust.

Tegyük fel, hogy a teljes megfigyelésünk Z, és valamilyen β paramétervektor írja le az eloszlását, de Z -nek csak valamilyen X függvényét tudjuk megfigyelni. Ezt hiányos megfigyelésnek nevezzük.

A β maximum-likelihood becslését keressük, iteratív módszerrel, ahol az itereráció $\beta^{(0)}$ kezdőértékből indul, és minden iteráció a következő két lépésből áll:

1.lépés: **E-lépés** (expectation)

2. lépés: **M-lépés** (maximization)

Így hajtjuk végre az algoritmust:

1. E-lépés: Van egy X megfigyelésünk, ami hiányos. A feltételes várható értéket keressük, a hiányos megfigyelés mellett. (Majd ezt a $\beta^{(t)}$ paraméterérték mellett maximalizáljuk β-ban.)

$$Q(\beta, \beta^{(t)}) = E(\log L(Z, \beta)|X, \beta^{(t)})$$

Ez a log-likelihood függvény várható értéke.

2. M-lépés: Ha az előbbi $Q(\beta, \beta^{(t)})$ függvényt maximalizáljuk, úgy kapjuk az új paramétervektort, $\beta^{(t+1)}$-et. Tudjuk, hogy $L(X, \beta^{(t)})$ likelihood-függvény az algoritmus során monoton növekszik, és konvergál az L^* értékhez, ha a likelihood-függvény felülről korlátos. De nem biztos, hogy ez az L^* a globális maximum is, ugyanis a $\beta^{(t)}$ sorozat likelihood-függvény nyeregpontjához is konvergálhat. A kovergencia sebessége a hiányzó adat hányadosától függ.

Az EM-algoritmus előnyei: monoton növekszik a likelihood, könnyű beprogramozni, kicsi a számításigénye, gyorsan fut.

2.3. Az MM-algoritmus

A β a paramétervektort szeretnénk becsülni maximum-likelihood becsléssel, X mintából. (Itt nincs teljes, és hiányos minta.)

A $\beta^{(0)}$ kezdőértékből indulva az algoritmus egy minorizáló, majd egy maximalizáló lépést végez el. E két lépés szerint kell iterálni. Ezért nevezzük MM-algoritmusnak. Az alábbi két lépés tehát:

1. M-lépés: Minorization-lépés: előállít egy olyan $Q_t(\beta)$ függvényt, melyre

$$Q_t(\beta) \leq \log L(X, \beta) \quad \forall \beta$$

és

$$Q_t(\beta^{(t)}) = \log L(X, \beta^{(t)})$$

2. M-lépés: Maximization-lépés: $Q_t(\beta)$ függvényt kell a β-ban maximalizálni, így megkapjuk az új $\beta^{(t+1)})$ paramétervektort.

Az $L(X, \beta^{(t)})$ likelihood monoton nő. A Q_t függvényt jól kell megválasztani. Az a jó, ha Q_t függvény szétválasztja a β paramétervektor koordinátáit. Ez azért kell, hogy a β-ban vett maximalizálást koordinátánként tudjuk elvégezni.

2.4. Az MM-algoritmus az EM-algoritmus vonatkozásában

Az EM-algoritmusok valójában speciális MM-algoritmusok, így vizsgálhatjuk a kettőt egyszerre is.

Az EM-algoritmus (Dempster, Laird és Rubin, 1977), mint már említettem, egy nagyon gyakran használt, általános statisztikai módszer a hiányos adatrendszereknél, a likelihood maximalizálására.

Legyen itt a h a megfigyelt, az y a hiányzó adat. $z = (h; y)$

Jelölje $f(z|x)$ a z teljes adathalmazból való mintavétel sűrűségét, és x jelöljön egy ismeretlen paramétervektort. A $z = (h, y)$ vektor adatait kombináljuk a ténylegesen megfigyelt h *hiányzó* adatokkal. (Itt most h a hiányos adathalmaz, mert h-ból hiányzik az y.)

Legyen $g(h|x)$ a hiányos adatok likelihoodja, ezért ezt akarjuk maximalizálni. Legyen $k(z|h, x)$ az $f(z|x)/g(h|x)$ feltételes sűrűsége. Az alapelvünk az, hogy szétszedjük a célfüggvényt úgy, hogy:

$$\log g(h|x) = E\left[\log f(z|x)|h, \underline{x}\right] - E\left[\log k(z|h, x)|h, \underline{x}\right] \tag{1}$$

Ez az egyenlet következik a $k(z|h, x)$ átrendezéséből, a logaritmus definíciójából, majd ha átrendeztük, tekintjük a várható értéket.

Adottak a h megfigyelt értékek, és az $\underline{x}$ becslések, és a második tagban az összes adat sűrűségének feltételes várható értéke. Így a második tagot a következőképpen tudjuk minorizálni:

$$E\left[\log k(z|h,x)|h,\underline{x}\right] \leq E\left[\log k(z|h,\underline{x})|h,\underline{x}\right]$$

Így kaptunk egy egyenlőtlenséget, a Jensen-egyenlőtlenség felhasználásával, a feltételes várható értékre.

Tekintsük úgy, mintha egyszerű, véges eset lenne! Tegyük fel, hogy kifejezhetjük m elemmel a $k(z|h,x)$ feltételes sűrűséget, azaz mivel z diszkrét eloszlású, ezért fel tudjuk írni két feltételes, de egy-egy diszkrét eloszlással. Ekkor z felvehet m értéket. Az x paraméter mellett a megfelelő valószínűségek így alakulnak: $\{p_1,...,p_j,...,p_m\}$, és $\underline{x}$ paraméterek mellett a következőképpen: $\left\{\underline{p}_1,...,\underline{p}_j,...,\underline{p}_m\right\}$. Mivel ezek a feltételes valószínűségek, eleget tesznek annak, hogy $\sum_j p_j = \sum_j \underline{p}_j = 1$.

Tekintsünk egy olyan valószínűségi változót, amely $\underline{p}_j$ valószínűséggel $p_j/\underline{p}_j$ értéket veszi fel, azaz valamely z valószínűségi változó j-edik értékét $p_j/\underline{p}_j$ lehetséges értékkel veszi fel.

A logaritmus függvény konkáv, ezért $p_j/\underline{p}_j$ konvex kombinációjára a következő egyenlőtlenség teljesül:

$$\log \sum_j \underline{p}_j (p_j/\underline{p}_j) \geq \sum_j \underline{p}_j \log(p_j/\underline{p}_j)$$

ami akkor és csak akkor egyelő, ha $p_j = \underline{p}_j \quad \forall j$-re.

(Tehát, ha elvégezzük a legutóbbi egyenlőtlenségben a beszorzást, a baloldalon $\underline{p}_j$ kiesik, és csak p_j marad. Ezeknek az összege 1. Egynek pedig a logaritmusa 0, tehát pont akkor maximális $\underline{p}_j$ szerinti várható értéke $\log p_j$-nek, ha $p_j = \underline{p}_j$.)

$$\sum_j \underline{p}_j \log p_j \leq \sum_j \underline{p}_j \log \underline{p}_j$$

Akkor és csak akkor tudunk egyenlőséghez jutni, ha $x = \underline{x}$. A maximumot pedig akkor érjük el, ha $\underline{p}_j$ pont p_j, mert a logaritmus várható értéke akkor lesz maximális, ha mindegyik p egyenlő egymással.

Az (1) jobboldalát tudtuk minorizálni. Amivel minorizálunk, már nem függ x-től.

Ezért lehet maximalizálni a minorizált $E\left[\log f(z|x)|h, \underline{x}\right]$ -t, vagyis a képletben az első tag maximumát keressük a teljes adatokra, a várható értéknek megfelelően. Ez pont az EM-algoritmus.

Az E-lépés határozza meg az EM-algoritmusban z várható értékét, és egy olyan becslést $\underline{x}$ -ra, amely egy megfelelő választás a minorizáló függvények családjából, és ez elegendő a minorizáló függvény maximumának megtalálásához.

3. Bevezetés a Bradley-Terry modellbe

Már a szakdolgozatom bevezetésében is említettem, hogy nagy általánosságokban miről is van szó. Vizsgáljuk meg most matematikus szemmel!

A Bradley-Terry modell párosított összehasonlításokon alapul. Ez egy olyan egyszerű, és sokat, sokak által vizsgált eszköz, mely képes leírni a lehetséges eredmények valószínűségeit. Ha két dolgot hasonlítunk egymással össze - jelen esetben sportolókat, de akár piacvezető újságokra is felállíthatjuk a modellt - melyik jobb, melyik kevésbé, esetleg mindkettő egyformán, stb.

A modellnek számos többirányú általánosítása is született az elmúlt 75 évben, melyekben iteratív algoritmust használtak az általánosítás maximum-likelihood becslésének elérésére. Ilyen algoritmus az EM- algoritmus is, mely az MM- algoritmusnak speciális esete. Előbb minorizálja, majd maximalizálja a már minorizált függvényünket.

Egyszerű feltételek mellett kijelenthetjük, hogy minden algoritmus garantáltan előállít egy olyan sorozatot, mely konvergál az egyetlen maximum-likelihood becsléshez.

3.1. A modell

A következő modellt javasolta Bradley és Terry a fenti problémára 1952-ben.

$$P(i \text{ játékos megveri a } j \text{ játékost}) = \frac{\gamma_i}{\gamma_i + \gamma_j}. \tag{2}$$

Párosított összehasonlításokat vizsgálunk. Az (2) képletben a γ_i egy pozitív értékű paraméter, amely az i játékos teljesítményének előzetesen becsült paramétere (az addigi versenyeik eredményei alapján), míg a γ_j pozitív értékű paraméter, a j játékos teljesítményének előzetesen becsült paramétere. Ha csapatokra alkalmazzuk ezt a képletet, akkor a csapat átlagos képességét nézzük, akkor ezt jelölhetjük γ_i-val és γ_j-val.

Bradley-Terry problémája 1929-re nyúlik vissza, ugyanis ezt Zermelo már széles körben alkalmazta, de nem általánosította különböző esetek problémakörére.

A modellt felrajzolhatjuk irányított gráffal is. Ekkor i-k és j-k a csomópontok, és minden i és j között megy él, ha ők játszottak egymással. Ha többször is, súlyozzuk az éleket nemmegatív számokkal. Az irányítás mindig a felé mutat egységesen, aki a párharcból nyertesen, vagy vesztesen került ki. Ha például a vesztes felé mutatunk, akkor ráírjuk az irányított élre, hogy az i játékos (ha ő győzött) hányszor verte meg j játékost. Minden i és minden j között mindig van él, ha ők játszottak egymással.

Zermelo, Bradley és Terry után Davidson, és még sokan mások is foglalkoztak a modellel, általánosították, illetve történetét is megírta Simons és Yao.

Régóta ismert egy egyszerű, iteratív algoritmus a Bradley-Terry modellben a maximum-likelihood becslés megtalálására, de mióta Lange, Hunter és Yang bizonyította, hogy ez az algoritmus egy speciális esete az algoritmusok általános osztályának, azóta említjük MM- algoritmus néven. 30 év alatt sokat vizsgálták, különböző nevek alatt, de 2000-ben Hunter és Lange megadta a választ a problémára. Heiser használja a a kezdeti IM -et (iteratív majorizációt) az algoritmusok osztályának leírására, ahol IM ugyanaz, mint az MM, de az MM elnevezés jobban hangsúlyozza az MM- és az EM-algoritmus közötti kapcsolatot, minthogy ismeretes, hogy az EM az MM speciális esete. Megvizsgáljuk, hogyan tudunk az általánosított Bradley-Terry modellekre MM-algoritmusokat felépíteni, elégséges feltételek mellett, melyek garantálják az egyetlen maximum likelihood becsléshez való konvergenciát.

3.2. A modell alkalmazása

Tegyük fel, hogy megfigyelünk tetszőleges számú párosítást m egynén, csapat, vagy verenyző között, és becsülni szeretnénk a $\gamma_1, ..., \gamma_m$ paramétereket a maximum-likelihood becslés felhasználásával. Ha a különböző párosítások kimeneteleiről azt feltételezzük, hogy függetlenek, a Bradley-Terry modellben a log-likelihood a következő:

$$\ell(\gamma) = \sum_{i=1}^{m}\sum_{j=1}^{m}[w_{ij}\ln\gamma_i - w_{ij}\ln(\gamma_i+\gamma_j)] \tag{3}$$

Ahol w_{ij} azt fejezi ki, hogy hányszor veri meg i a j játékost, ha például sporteseményeket nézünk. Értelemszerűen $w_{ii}=0$, $\ell(\gamma)=\ell(a\gamma)$, $a>0$. A paraméterteret úgy kell tekintenünk, mint $\mathbb{R}_+^m$ ekvivalencia-osztályainak halmazát. Két vektor egyenlő, ha az egyik skalászorosa a másiknak. Ez könnyedén teljesül, ha korlátossá tesszük a paraméterteret. Ezért feltehetjük, hogy $\sum_i \gamma_i = 1$, és minden ekvivalenciaosztályból egy elemet kiválasztunk.

Szétbontjuk a versenyzők halamzát két diszjunkt részhalmazra. Valakik az A halmazba kerülnek, míg mások a B-be. Tegyük fel, hogy az A halmazbeli elemeket csak az A halmazbeli elemekkel, míg a B halmazbeli elemeket csak a B halmazbeli elemekkel hasonlítjuk össze. De így az a probléma, hogy az A halmazbeli versenyzőket sehogy sem tudjuk összehasonlítani a B halmazbeli versenyzőkkel.

Probléma még akkor adódhat, ha A és B elemeit ugyan össze tudjuk hasonlítani egymással, de a versenyeket például mindig A halmazbeli versenyző nyeri meg. Ekkor az A-beli paramétereket megduplázzuk, és újra normalizálunk. Úgy, hogy $\sum_i \gamma_i = 1$ lesz. A likelihood nőni fog, ezért nincs maximum-likelihood becslés.

A következő feltételezéssel kiküszöbölhetjük a problémák fennállásánal lehetőségét.

Ford feltevése:

A versenyzők minden lehetséges felosztásában két nemüres részhalmazt nézünk. Valamelyik versenyző a második halmazból megveri az első halmaz valamelyik tagját, legalább egyszer.

Gráfelméleti értelmezés szempontjából az egyének (versenyzők) a gráf csomópontjai (csúcsai), és irányított éllel (i,j) jelöljük azt, ha i győzött j felett. Ez a feltételezés egyenértékű azzal az állítással, hogy minden i - j párra van út i -től j -be.

Ez azt jelenti, hogy többek között létezik egyértelmű maximuma a log-likelihood függvénynek.

4. Bradley-Terry modell általánosításai

A Bradley-Terry modellre számos általánosítás született. Pl. Agresti (1990) felteszi, hogy a versenyzők bármely párosított összehasonlítása sorrendben történik, és megköveteljük, hogy annak valószínűsége, hogy i játékos megveri j játékost, attól függ, hogy milyen képességekkel rendelkezik a vesenyzők között az, aki az első helyen szerepel a listán.

Nem muszáj feltétlenül egyéni játékosokat tekinteni. Csapatot is tekinthetünk egynénnek, verenyzőnek. Ekkor a csapat átlagos képességét mérjük.

4.1. "Hazai pálya modell"

Sportban nagyon gyakran előfordul, hogy egy csapat valami fontos mérkőzésén, esetleg világversenyen hazai pályán játszik. Ez vajon gátolja, vagy segíti a győzelemben? Erre írhatunk fel egy matematikai modellt.

$$P(i \text{ játékos megveri a } j \text{ játékost}) = \begin{cases} \theta\gamma_i/(\theta\gamma_i + \gamma_j) & \text{ha } i \text{ otthon van} \\ \gamma_i/(\gamma_i + \theta\gamma_j) & \text{ha } j \text{ játszik otthon} \end{cases} \tag{4}$$

Ahol $\theta > 0$ méri a hazai pálya erősségét. Azt, hogy a hazai pálya inkább előny, vagy hátrány a versenyzők számára.

Hogy a hazai pálya egy versenyzőnek előnyt vagy hátrányt jelent, attól függ, hogy a θ paraméter 1-nél kisebb, vagy nagyobb.

Ha $\theta > 1 \Rightarrow$ a hazai pálya ***előny***t jelent az otthon játszó versenyzőnek.

Ha $\theta < 1 \Rightarrow$ a hazai pálya ***hátrány***t jelent az otthon játszó versenyzőnek.

4.2. A Rao-Kupper-féle döntetlen esete

A modellt kiterjeszthetjük több irányba úgy, hogy feltesszük, hogy a döntetlen is megengedett legyen a csapatok között.

$$\begin{aligned}
&P(i \text{ játékos legyőzi a } j \text{ játékost}) = \gamma_i/(\gamma_i + \theta\gamma_j)\\
&P(j \text{ játékos legyőzi az } i \text{ játékost}) = \gamma_j/(\theta\gamma_i + \gamma_j) \qquad (5)\\
&P(i \text{ és } j \text{ játékosok döntetlent játszanak}) = (\theta^2 - 1)\gamma_i\gamma_j\,[(\gamma_i + \theta\gamma_j)/(\gamma_j + \theta\gamma_i)]
\end{aligned}$$

A $\theta > 1$ egy küszöbparaméter. Minden párosításnál felmerülhet, hogy a bíró az $\ln\gamma_i - \ln\gamma_j$-t hibával becsli , és kijelenti a döntetlent, ha ennek az értéke kisebb, mint $\ln\theta$ értéke abszolútértékben. Ez azt jelenti, hogy a folyamatban lévő mérkőzés során a bíró meg tudja becsülni egyik, illetve másik versenyző erősségét. Nagyobb különbség esetén egyértemű́en ki tudja hirdetni a győztest, míg kisebb, vagy alig eltérő különbségnél nagyobb a valószínűsége annak, hogy hibával becsli meg a játékosok képességét, így 9-10-nél nagyobb a valószínűsége, hogy döntetlent jelent ki, mint például egy 20-10-es állásnál.

4.3. A Davidson-féle döntetlen esete

Davidson (1970) különböző beállításokat ad meg a Bradley-Terry modellben a döntetlen esetére, melyben a valószínűségek egymással arányosak.

$$\begin{aligned}
&P(i \text{ játékos legyőzi a } j \text{ játékost}) : P(j \text{ játékos legyőzi az } i \text{ játékost}) :\\
&\qquad : P(i \text{ játékos döntetlent játszik a } j \text{ játékossal}) = \gamma_i : \gamma_j : \theta\sqrt{\gamma_i\gamma_j}. \qquad (6)
\end{aligned}$$

A döntetlen valószínűsége a két versenyző nyerési valószínűségének mértani közepével arányos. A pozitív értékű θ paraméter mutatja meg ezt az arányossági tényezőt.

Davidson(1970) a *mértani közép* használatát javasolja. Az egyéni érdemeket logaritmikus skálán képzeljük el, és $\log\gamma$-kat hasonlítjuk össze.

4.4. A modell három személyre

A Bradley-Terry modellt kiterjeszthetjük úgy, hogy nem csak kettő személyt, versenyzőt, csapatot vizsgálunk, hanem mondjuk hármat, majd később többet is egyszerre. Ha hármat nézünk, mindhárom versenyző eredményeit rangsoroljuk a legjobbtól a legrosszabbig. Felírjuk, hogy ki a legjobb, a közepes, és ki a legrosszabb az adott játékosok, versenyzők között. Pendergrass és Bradley javasolta a következő modellt erre az esetre:

$$P(i \text{ a legjobb}, j \text{ a közepes}, k \text{ a legrosszabb}) = \frac{\gamma_i \gamma_j}{(\gamma_i + \gamma_j + \gamma_k)(\gamma_j + \gamma_k)} \tag{7}$$

Ez az általánosítás tetszőleges számú egyén összehasonlítására alkalmas. Ez az úgy nevezett Plackett-Luce modell.

5. Minorizáló függvény és az MM-algoritmus

A logarimus függvény szigorú konkáv voltából következik pozitív x-re és y-ra, hogy:

$$-\ln x \geq 1 - \ln y - (x/y) \tag{8}$$

ahol egyenlőség akkor és csak akkor áll fenn, ha $x = y$.

Most nézzük a sima Bradley-Terry modellt, és a (3)-es képletre alkalmazzuk a fenti egyenlőtlenséget. $\ln \gamma_i - \ln(\gamma_i + \gamma_j)$, ahol $(\gamma_i + \gamma_j) = x$. Továbbá $-\ln x \geq 1 - \ln y -$ $-(x/y)$. Ebbe visszaírjuk a gammákat, akkor: $-\ln(\gamma_i + \gamma_j) \geq 1 - \ln(\gamma_i^{(k)} + \gamma_j^{(k)}) - (x/y)$. $-\dfrac{x}{y} = -\dfrac{\gamma_i + \gamma_j}{\gamma_i^{(k)} + \gamma_j^{(k)}}$. Így kaphatjuk a következőt:

$$Q_k(\gamma) = \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij} \left[\ln \gamma_i - \frac{\gamma_i + \gamma_j}{\gamma_i^{(k)} + \gamma_j^{(k)}} - \ln(\gamma_i^{(k)} + \gamma_j^{(k)}) + 1 \right] \tag{9}$$

Majd ezt kell maximalizálni. Ez az iteráció növeli a likelihood-ot.

$Q_k(\gamma)$ függvény iménti meghatározása megkönnyíti a maximalizálást. Ekkor az eredeti log-likelihood ténylegesen elkülöníti a γ paramétervektor összetevőit, $Q_k(\gamma)$-ban γ komponensei szétválnak. Így a $Q_k(\gamma)$ maximalizálása egyenlő azzal, ha minden egyes komponenst külön-külön maximalizálunk.

Ha ciklikus esetre nézzük, a ciklikus algoritmus maga is egy MM-algoritmus, mivel $\gamma_i^{(k+1)}$ a maximalizálója $Q_k(\gamma_i^{(k+1)}, ..., \gamma_{i-1}^{(k+1)}, \gamma_i, \gamma_{i+1}^{(k)}, ..., \gamma_m^{(k)})$ -nak, amely minorizálja $\ell(\gamma)$ -át a $\gamma = (\gamma_i^{(k+1)}, ..., \gamma_{i-1}^{(k+1)}, \gamma_i, \gamma_{i+1}^{(k)}, ..., \gamma_m^{(k)})$ pontokban.

Ciklikus esetben nem mindig egyértelmű, hogy mit értük egy algoritmus iterációján.

5.1. Iteratív algoritmus a $\ell(\gamma)$ maximalizálására

Vezessünk be egy kezdeti paramétervektort: $\gamma^{(1)}$ /Dykstra(1956) taglal néhány lehetőséget/ Habár a kezdőpont jó megválasztása csökkenti az általános számítási igényt, mi most feltesszük, hogy $\gamma^{(1)}$ megválasztás tetszőleges.

Ha minden egyes komponensre külön-külön elvégezzük a maximalizálást, a következőhöz jutunk. Tehát a maximalizálásnak a megoldása:

$$\gamma_i^{(k+1)} = W_i \left[\sum_{j \neq i} \frac{N_{ij}}{\gamma_i^{(k)} + \gamma_j^{(k)}} \right]^{-1} \tag{10}$$

W_i jelöli az i játékos nyeréseinek számát. $N_{ij} = w_{ij} + w_{ji}$ a párosítások száma i és j között. Ha az eredő $\gamma^{(k+1)}$ vektor nem felel meg a $\sum_i \gamma^{(k+1)} = 1$ korlátnak, egyszerűen újra kell normalizálni.

Amelyiknek már megvan a $(k+1)$. értéke, azt használhatom, frissíthetek vele. Ez vezet a ciklikus MM-algoritmus előállításához, melyet ha maximalizálunk, a következőhöz jutunk:

$$\gamma_i^{(k+1)} = W_i \left[\sum_{j<i} \frac{N_{ij}}{\gamma_i^{(k)} + \gamma_j^{(k+1)}} + \sum_{j>i} \frac{N_{ij}}{\gamma_i^{(k)} + \gamma_j^{(k)}} \right]^{-1} \tag{11}$$

Mindkét algoritmus előállít egy olyan $\gamma^{(1)}, ..., \gamma^{(n)}$ sorozatot, amely garantálja a konvergenciát az egyetlen maximum likelihood becsléshez. Emellett $\ell(\gamma^{(1)}), ..., \ell(\gamma^{(n)})$ monoton növekedő. Az $\left\{\ell(\gamma^{(k)})\right\}$ sorozat monotonitása minden MM algoritmusnak karakterisztikus tulajdonsága.

Az MM-algoritmus ciklikus változata is örökli a konvergencia tulajdonságokat.

5.2. MM-algoritmus hazai pályára

A már előzőekben ismertetett *"Hazai pálya modell"*- nél az egyenlőtlenség felhasználásával felépíthetünk egy egy minorizáló függvényt a log-likelihood függvényre.

$$\ell(\gamma,\theta)=\sum_{i=1}^{m}\sum_{j=1}^{m}\left[a_{ij}\ln\frac{\theta\gamma_i}{\theta\gamma_i+\gamma_j}+b_{ij}\ln\frac{\gamma_j}{\theta\gamma_i+\gamma_j}\right] \tag{12}$$

ahol a_{ij} jelöli, hogy i hányszor verte meg hazai pályán j-t, és b_{ij} jelöli azt, hogy i hányszor kapott ki hazai pályán j-től. Legyen $H=\sum_i\sum_j a_{ij}$ az hazai pályán aratott győzelmek száma és W_i az i csapat összes győzelmének száma. Ezeket figyelembe véve a következőt írhatjuk föl:

$$Q_k(\gamma,\theta)=H\ln\theta+\sum_{i=1}^{m}W_i\ln\gamma_i-\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\frac{(a_{ij}+b_{ij})(\theta\gamma_i+\gamma_j)}{\theta^{(k)}\gamma_i^{(k)}+\gamma_j^{(k)}}\right]$$

Ez $\ell(\gamma,\theta)$-t egy additív konstans erejéig minorizálja, így a következőhöz jutunk:

$$Q_k(\gamma,\theta)+\left[\ell(\gamma^{(k)},\theta^{(k)})-Q_k(\gamma^{(k)},\theta^{(k)})\right]\le\ell(\gamma,\theta)$$

A $\theta\gamma_i$ szorzat előfordulása azt jelenti, hogy a paramétereket nem tudja teljesen elkülöníteni a minorizáló függvény, ami a függvény közvetlen maximalizálását némileg problematikussá teszi. Habár, könnyű maximalizálni $Q_k(\gamma,\theta^{(k)})$-t, mint a γ függvényét és $Q_k(\gamma^{(k+1)},\theta)$-át, mint θ függvényét. Így konstruálhatunk egy ciklikus algoritmust erre az esetre.

5.3. MM-algoritmus a Rao-Kupper-féle döntetlen esetére

Itt a log-likelihood:

$$\ell(\gamma,\theta)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\{2w_{ij}\ln\left(\frac{\gamma_i}{\gamma_i+\theta\gamma_j}\right)+t_{ij}\ln\left(\frac{(\theta^2-1)\gamma_i\gamma_j}{(\theta\gamma_i+\gamma j)(\gamma_i+\theta\gamma_j)}\right)\right\} \tag{13}$$

Itt $t_{ij}=t_{ji}$ az a szám, ahányszor az i és j versenyzők döntetlent játszottak egymás ellen.

Használjuk az előző fejezet legelső egyenlőtlenségét. Ebből megkonstruálhatjuk a következőt:

$$Q_k(\gamma,\theta)=\sum_{i=1}^{m}\sum_{j=1}^{m}\left\{(w_{ij}+t_{ij})\left(\ln\gamma_i-\frac{\gamma_i+\theta\gamma_j}{\gamma_i^{(k)}+\theta^{(k)}\gamma_j^{(k)}}\right)+t_{ij}\ln(\theta^2-1)\right\}$$

Ez minorizálja $\ell(\gamma,\theta)$-t a $(\gamma^{(k)},\theta^{(k)})$-ban.

A paraméterek nem teljesen szeparáltak, de felváltva maximalizálhatjuk $Q_k(\gamma,\theta^{(k)})$ -t, mint γ függvényét, és $Q_k(\gamma^{(k+1)},\theta)$-t, mint θ függvényét. Ezzel egy ciklikus MM-algoritmushoz jutunk.

$Q(\gamma,\theta^{(k)})$ maximalizálása γ-ra vonatkozóan adja:

$$\gamma_i^{(k+1)}=\left[\sum_{i=j}s_{ij}\right]\left[\sum_{j\neq i}\left(\frac{s_{ij}}{\gamma_i^{(k)}+\theta^{(k)}\gamma_j^{(k)}}+\frac{\theta^{(k)}s_{ji}}{\theta^{(k)}\gamma_i^{(k)}+\gamma_j^{(k)}}\right)\right]^{-1}\tag{14}$$

Ahol $s_{ij}=w_{ij}+t_{ij}$ az a szám, ahányszor az i versenyző megverte, vagy döntetlent játszott j versenyzővel.

Másodfokú egyenlet megoldásával maximalizálhatjuk $Q_k(\gamma^{(k+1)},\theta)$-t, ami θ-ra vonatkozólag adja:

$$\theta^{(k+1)}=\frac{1}{2C_k}+\sqrt{1+\frac{1}{4C_k^2}}$$

ahol

$$C_k=\frac{2}{T}\sum_{i=1}^{m}\sum_{j\neq i}\frac{\gamma_j^{(k+1)}(s_{ij})}{\gamma_i^{(k+1)}+\theta^{(k)}\gamma_j^{(k+1)}}$$

T a döntetlenek teljes száma az összes megfigyelt összehasonlítás között. Az előbbi egyenletet Rao és Kupper javasolta, bár ők nem tártak föl minden ebből származó konvergencia tulajdonságot.

A fenti egyenletet módosíthatjuk úgy, hogy γ paramétert állandóan frissítjük, így elkészíthetjük a ciklikus változatát.

5.4. MM-algoritmus a Davidson-féle döntetlen esetére

Erre a modellre alkalmazva a log-likelihood-ot, a következőt kapjuk:

$$\ell(\gamma,\theta) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[2w_{ij}\ln\frac{\gamma_i}{\gamma_i+\gamma_j+\theta\sqrt{\gamma_i\gamma_j}} + t_{ij}\ln\frac{\theta\sqrt{\gamma_i\gamma_j}}{\gamma_i+\gamma_j+\theta\sqrt{\gamma_i\gamma_j}}\right] \tag{15}$$

minorizált az irreleváns konstansig az (8) egyenlőtlenségen keresztül:

$$Q_k^*(\gamma,\theta) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[2w_{ij}\ln\gamma_i + t_{ij}\ln(\theta\sqrt{\gamma_i\gamma_j}) - \frac{(2w_{ij}+t_{ij})(\gamma_i+\gamma_j+\theta\sqrt{\gamma_i\gamma_j})}{\gamma_i^{(k)}+\gamma_j^{(k)}+\theta^{(k)}\sqrt{\gamma_i^{(k)}\gamma_j^{(k)}}}\right]$$

által.

Habár a második $\sqrt{\gamma_i\gamma_j}$ miatt $Q_k^*(\gamma,\theta)$ maximalizálása nem könnyű, még akkor sem, ha θ-t rögzítjük a $\theta^{(k)}$ pontban, ezért a továbbiakban egy jól ismert egyenlőtlenséget hívunk segítségül. A *számtani-mértani közép egyenlőtlenség* által fel tudjuk építeni $Q_k^*(\gamma,\theta)$ egy minorizációját.

Ebben az általános formában az a *számtani-mértani közép egyenlőtlenség*ből következik, hogy $\prod_i x_i^{w_i} \leq \sum_i w_i\ x_i \geq 0$ -ra és $w_i > 0$ -ra és $\sum_i w_i = 1$, ahol egyenlőséget akkor és csak akkor engedünk meg, ha minden x_i egyenlő. Ha $w_1 = w_2 = \frac{1}{2}$-del elérjük:

$$-\sqrt{\gamma_i\gamma_j} \geq -\frac{\gamma_i}{2}\sqrt{\frac{\gamma_j^{(k)}}{\gamma_i^{(k)}}} - \frac{\gamma_j}{2}\sqrt{\frac{\gamma_i^{(k)}}{\gamma_j^{(k)}}} \tag{16}$$

Egyenlőség akkor van, ha $\gamma = \gamma^{(k)}$. Ezért $Q_k^*(\gamma,\theta)$ -t minoralizálja $Q_k(\gamma,\theta)$, a $(\gamma^{(k)},\theta^{(k)})$ -ban.

$$Q_k(\gamma,\theta) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\Bigg[2w_{ij}\ln\gamma_i + t_{ij}\ln(\theta\sqrt{\gamma_i\gamma_j}) - \frac{(2w_{ij}+t_{ij})(\gamma_i+\gamma_j)}{\gamma_i^{(k)}+\gamma_j^{(k)}+\theta^{(k)}\sqrt{\gamma_i^{(k)}\gamma_j^{(k)}}} - \frac{\theta(2w_{ij}+t_{ij})}{\gamma_i^{(k)}+\gamma_j^{(k)}+\theta^{(k)}\sqrt{\gamma_i^{(k)}\gamma_j^{(k)}}}\left(\frac{\gamma_i}{2}\sqrt{\frac{\gamma_j^{(k)}}{\gamma_i^{(k)}}}+\frac{\gamma_j}{2}\sqrt{\frac{\gamma_i^{(k)}}{\gamma_j^{(k)}}}\right)\Bigg]$$

A minorizáció egy tranzitív reláció. $Q_k(\gamma,\theta)$ minorizálja $Q_k^*(\gamma,\theta)$-t $(\gamma^{(k)},\theta^{(k)}$ -ban-ban, és $Q_k^*(\gamma,\theta)$ minorizálja $\ell(\gamma,\theta)$-t $(\gamma^{(k)},\theta^{(k)}$ -ban. Ekkor $Q_k(\gamma,\theta)$ is minorizálja $\ell(\gamma,\theta)$-t $(\gamma^{(k)},\theta^{(k)})$ -ban.

γ összetevői most szeparáltak, és $Q_k(\gamma,\theta^{(k)})$ maximalizációja γ-ra nézve:

$$\gamma_i^{(k+1)} = \frac{2W_i+T_i}{\sum_{j=1}^{m} g_{ij}(\gamma^{(k)},\theta^{(k)})}$$

Itt W_i az i versenyző összes nyerésének száma, T_i pedig az i versenyző összes döntelen játékának száma, és

$$g_{ij}(\gamma,\theta) = \frac{(w_{ij}+w_{ji}+t_{ij})(2+\theta\sqrt{\gamma_j/\gamma_i})}{\gamma_i+\gamma_j+\theta\sqrt{\gamma_i\gamma_j}} \tag{17}$$

Természetesen γ komponensei lehetnek ciklikusan frissítettek, ha a $\gamma_i^{(k+1)}$ nevezőjét $\sum_{j<i} g_{ij}(\gamma^{(k+1)},\theta^{(k)}) + \sum_{i<j} g_{ij}(\gamma^{(k)},\theta^{(k)})$. Végül maximalizáljuk $Q_k(\gamma^{(k+1)},\theta)$ -át, mint θ függvényét.

$$\theta^{(k+1)} = 4T\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{(2w_{ij}+t_{ij})(\gamma_i^{(k+1)}+\gamma_j^{(k+1)})}{\gamma_i^{(k+1)}+\gamma_j^{(k+1)}+\theta^{(k)}\sqrt{\gamma_i^{(k+1)}\gamma_j^{(k+1)}}}\right]$$

Ebben a modellben T az összes döntetlen számát jelöli.

Davidson (1970) közel azonos okoskodást használ, mint Ford az első feltételezés alapján vett bizonyításnál, a ciklikus verziónál, csak egy enyhe különbséggel frissíti θ-t. Ez garantálja az egyetlen maximum likelihood becslést.

Láttuk, hogyan tudjuk alkalmazni az MM-algoritmust a Bradley-Terry modell néhány általánosítására. Ugyanazt a technikát alkalmazhatjuk arra a modellre is, ahol három versenyzőt hasonlítunk össze. (Ezt később tárgyaljuk.)

Ezek az MM-algoritmusok, mint minden MM-algoritmus, garantáltan növelni fogják a log-likelihood-ot minden egyes iterációs lépésben, de az MM-algoritmusnak ez a monotonitási tulajdonsága nem garantálja még, hogy ez az algoritmus elvezet minket a maximum-likelihood becsléshez. A következő részben a konvergencia vizsgálatával foglalkozunk.

6. Az MM-algoritmus konvergenciájának tulajdonságai

Van némi bizonytalanság azt illetően, hogy mit is értünk egy algoritmus konvergenciáján.

Mi itt most azt mondjuk, hogy egy algoritmus konvergens, ha:

$$\gamma^* = \lim_k \gamma^{(k)}$$

Ez sokkal szigorúbb definíció a konvergenciára nézve ahhoz képest, amit néha látunk az irodalomban. Például Hastie és Tibshirani (1998) csak azt jegyzik meg, hogy $\lim_k \gamma^{(k)}$ létezik, és véges, ebből következik, hogy az algoritmus konvergens.

Mi itt most két okból is az erősebb definíciót fogjuk használni. Először is γ végső értéke sokkal érdekesebb, mint $\ell(\gamma)$ végső értéke; másodszor pedig $\lim_k \ell(\gamma^{(k)})$ határértéke véges, ha $\ell(\gamma)$ felülről korlátos. Ha γ^* egyértelműen létezik, az érdekel minket, ez hogyan tudná maximalizálni $\ell(\gamma)$-t.

Általánosságban ezt nem mindig lehet bizonyítani, hogy egy MM-algoritmus által meghatározott paraméterek sorozata konvergál. Nem is beszélve a globális maximumról. McLachlan és Krishnan(1997) példát mutat olyan EM-algoritmusra, amely vagy a nyeregponthoz konvergál, vagy egyáltalán nem konvergál. Ford (1957) az első feltételezés alapján mutat (11) egy olyan algoritmusát, ami konvergál az egyetlen maximum-likelihood becsléshez, és korábban Zermelo (1929) származtatott már egy hasonló eredményt. Erre az eredményre úgy tekinthetünk, mint egy sokkal általánosabb tétel következményére. [Lange(1995)].

Ljapunov tétele

6.1. Tétel. *Tegyük fel, hogy $M : \Omega \to \Omega$ folytonos és $\ell : \Omega \to \mathbb{R}$ differenciálható, és $\forall\gamma \in \Omega$ -ra $\ell[M(\gamma)] \geq \ell(\gamma)$. Egyenlőség csak akkor áll fenn, ha γ egy stacionárius pontja ℓ-nek, azaz ha a gradiens 0 a γ -ban. Ekkor tetszőleges $\gamma^{(1)} \in \Omega$-ra a $\left\{\gamma^{(k+1)} = M(\gamma^{(k)})\right\}_{k\geq 1}$ sorozat bármely torlódási pontja egy stacionárius pontja $\ell(\gamma)$-nak.*

Bizonyítás. $\ell(\gamma^{(k_n)}) \leq \ell(M(\gamma^{(k_n)})) = \ell(\gamma^{(k_n+1)}) \leq ... \leq \ell(\gamma^{(k_{n+1})})$. Ha $n \to \infty$ esetén $\ell(\gamma^{(k_n)})$ tart $\ell(\gamma^*)$-hoz, $\ell(M(\gamma^{(k_n)}))$ pedig $\ell(M(\gamma^*))$-hoz, de $\ell(\gamma^{(k_{n+1})})$ is $\ell(\gamma^*)$-hoz tart, akkor $\ell(\gamma^*) = \ell(M(\gamma^*))$, tehát ebből következik, hogy γ^* stacionárius pont. ■

Egy MM-algoritmusra, a tételben szereplő $M(\gamma)$ leképezés adott az algoritmus egy iterációja által, amely garantálja, hogy $\ell[M(\gamma)] \geq \ell(\gamma)$ legyen. Minden egyes MM-algoritmusról azt állítjuk, hogy $M(\gamma)$ folytonossága világos. Az $\ell[M(\gamma)] = \ell(\gamma)$ azt jelenti, hogy γ egy stacionárius pont. Ez abból következik,hogy a minorizáló függvény differenciálható, és ez az érintője a log-likelihood függvénynek az aktuális iterációban. Tehát a minorizáló függvény deriváltja/érintője megegyezik a log-likelihood függvény érintőjével a stacionárius pontban.

Ebben az esetben a ciklikus MM-algoritmusnál $M(\gamma^{(k)}) = \gamma^{(k+1)}$, ahol a parciális derivált nulla, ha csak az aktuálisat változtatjuk. Mindazonáltal M folytonossága világos, és csak akkor lehet, hogy $\ell[M(\gamma)] = \ell(\gamma)$ legyen, ha számos MM iterációban γ-t változatlanul hagyjuk , ami azt jelenti, hogy γ stacionárius pontja ℓ-nek. Ha egyik iterációban mindent változtatunk, akkor minden parciális derivált nulla lesz. Így a *Ljapunov-tétel* alapján a ciklikus MM-algoritmusra is az MM-algoritmus konvergencia tulajdonságai vonatkoznak.

A Bradley-Terry modell MM-algoritmusának konvergencia igazolására a következő a stratégiánk:

Először is, megadunk egy elégséges feltételt a log-likelihood függvény felső kompaktságára.

Az ℓ felülről kompakt, ha minden konstans c-re a $\{\gamma \in \Omega : \ell(\gamma) \geq c\}$ halmaz egy kompakt részhalmaza Ω paramétertérnek.

Másodszor, újra paraméterezzük a log-likelihoodot, és megaduk egy elégséges feltételt a újra-paraméterezett log-likelihood függvény szigorú konkávságára.

Míg a felső kompaktság azt jelenti, hogy legalább egy torlódási pont megléte szükséges, addig a szigorú konkávság azt jelenti, hogy legfeljebb egy stacionárius pont kell, hogy legyen. Nevezetesen a maximumhely.

*Ljapunov tétel*éből arra következtethetünk, hogy az MM-algoritmus konvergens, független a kezdőponttól, és konvergál az egyetlen maximum-likelihood becsléshez. Szemben más algoritmusokkal (pl. Newton-Raphson algoritmus), az MM-algoritmusban az átparaméterezés után az iterációk sorrendje nem változik. Az újra paraméterezés nem teszi tönkre a minorizációs tulajdonságokat vagy nem változtat a maximumon.

Szinte minden log-likelihood függvény, mely az előző fejezetben adott, felső kompakt, ha az első feltételezés teljesül. Kivételt képez a *Hazai pálya modell*, amire erősebb feltevést kell alkalmaznunk.

Második feltevés

(Az első feltevés Ford feltevése volt)

Vesszük a csapatok két lehetséges partícióját A halmazba és B halmazba soroljuk őket. Van olyan csapat, amelyik A halmazból megveri valamely B halmazbeli csapatot. Méghozzá olyanokat, akik hazai pályán jászanak, és néhány A-beli csapat megver néhány B-beli csapatot úgy, hogy ekkor A van otthon.

A következő lemma elégséges, de akár néhány esetben szükséges is lehet.

Feltételek a likelihood függvény felső kompaktságára:

Lemma 1. Legyen $\Omega = \{\gamma \in R^m : \forall \gamma_i > 0 \ \sum_{i=1}^{m} \gamma_i = 1\}$ A paramétertér Ω a (3) és a (7) log-liklelihoodjára, $\Omega \times \{\theta \in R : \theta > 0\}$ a (15) és a (12) log-likelihoodjára, és $\Omega \times \{\theta \in R : \theta > 1\}$ a (13) log-likelihoodjára.

Az első feltevés alapján azt mondjuk, hogy i megveri j-t egy hármas összehasonlításban, ha i előrébb áll a rangsorban, mint j.

(a) (3) és (7) likelihoodja felülről kompakt akkor és csak akkor, ha az első feltevés teljesül.

(b) (13) és (15) kilelihoodja felülről kompakt, ha teljesül az első feltevés, és legalább egy döntetlen van.

(c) (12) log-likelihoodja felülről kompakt, ha teljesül rá a második feltevés.

Az elégséges feltétel a felső kompaktságra a *Hazai pálya modell*ben, nevezetesen a második feltevés, amely szokatlanul erős. Ez azt jelenti, hogy minden csapat legalább négyszer játszik. Otthon és idegenben. Otthon nyer és veszít, majd idegenben nyer és veszít. Ez négy mérkőzést jelent minden egyes csapat számára.

A (b) és a (c) részben arról nincsen tudomásunk, hogy az elégséges feltételek egyben szükséges feltételek is lennének.

Mint ahogy már korábban tettük, most is újra paraméterezzük a modellt adott feltételek mellett, úgy, hogy a log-likelihood függvény szigorúan konkáv volta megmaradjon.

Legyen $\beta_i = \ln\gamma_i - \ln\gamma_1$, i megy 1-től m-ig.

Az inverz függvény:

$$\gamma_i = \frac{e^{\beta_i}}{\sum_{j=1}^{m} e^{\beta_j}}$$

létrehoz egy egy-egy értelmű megfeleltetést $\left\{\gamma \in \mathbb{R}_+^m : \sum_i \gamma_i = 1\right\}$ és $\{\beta \in \mathbb{R}^m : \beta_1 = 0\}$ között.

A modellek további paramétere θ. Legyen $\phi = \ln\theta$

Megjegyzés: Az első lemmában az újraparaméterezés után az állítások igazak maradnak (a)-tól (c)-ig. Mivel a paramétervektorokból előállított minden olyan sorozat, mely közelít az eredeti paramétertér határához, az közelít az újra paraméterezett tér határához is.

Újra paraméterezés után az eredeti, azaz az (2) -es Bradley-Terry modell a következővé válik:

$$\text{logit}\ [P(i \text{ játékos megveri a } j \text{ játékost})] = \beta_i - \beta_j \tag{18}$$

A "logit" kifejezés p és $1-p$ hányadosának logaritmusát jelenti.

Az (2)-re elvégezzük az ellenőrzést:

$P_{ij} = \dfrac{\gamma_i}{\gamma_i + \gamma_j}$ és $1 - P_{ij} = \dfrac{\gamma_j}{\gamma_i + \gamma_j}$

Ha ennek a kettőnek a hányadosát vesszük, $\dfrac{\gamma_i}{\gamma_j}$-t kapjuk. Ha ennek a hányadosnak vesszük a logaritmusát, az pont a logitP_{ij}-vel lesz egyenlő.

A (3)-as képlet, ha újraparaméterezzük, a következővé válik:

$$\ell(\beta) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left[w_{ij}\beta_i - w_{ij} \ln(e^{\beta_i} + e^{\beta_j}) \right] \tag{19}$$

Mint, ahogy Bradley és Terry a (18) -ban javasolja, a modellre illetszthetünk logisztikus regressziót, ami annyit jelent, hogy $0-1$ megfigyelésünk van, az alapján, hogy nyert, vagy nem nyert az általunk megfigyelt egyén. 0, ha nem nyert, 1, ha nyert. Mindezek alapján a nyerés valószínűségét szeretnénk felírni úgy, hogy logit $(p_{ij}) = c + \beta_i - \beta_j$. Agresti (1990)-ben leírja, hogyan is történik mindez.

Ha konstans tagot is tartalmaz a modell, akkor a modell speciális esetét, nevezetesen a *Hazai pálya modell*t kapjuk, melyben *hazai pálya* paraméterét a következőképp írhatjuk fel: $\phi = \log\theta$ az (4) -es képletből mindaddíg, amíg a kiszámítása helyesen definiált úgy, hogy: logit $(p_{ij}) = \log\theta + \beta_i - \beta_j$. A regresszióban a független változók a β_i és a β_j, melyek úgy nevezett prediktorok.

A logisztikus regresszió nem alkalmazható a Bradley-Terry modell bármelyik más általánosítására azok közül, melyeket most itt tárgyalunk.

A (19) -es képlet log-likelihood-jának konkáv volta azonnal következik, mert log-konvex függvények halmaza (azok a függvények, melyek logaritmus függvénye konvex) zártak az összedaásra nézve.

A konkávitást a Hölder-egyenlőtlenség felhasználásával tudjuk bizonyítani. Ennek a megközelítésnek a további előnye az, hogy elégséges feltételeket szolgáltat a a szigorú konkávitásra.

Hölder-egyenlőtlenség:

$$||fg||_1 \leq ||f||_p ||g||_q,$$

ahol $\dfrac{1}{p} + \dfrac{1}{q} = 1$

Tekintsük a logaritmust a Hölder-egyenlőtlenség egyik formájában, pozitív számokra $c_1, ..., c_N$ és $d_1, ..., d_N$ és $p \in (0,1)$, ekkor

$$\ln \sum_{k=1}^{N} c_k^p d_k^{1-p} \leq p \ln \sum_{k=1}^{N} c_k + (1-p) \ln \sum_{k=1}^{N} d_k \tag{20}$$

Bizonyítás.

$$\sum c_k d_k \leq (\sum c_k^p)^{\frac{1}{p}} \cdot (\sum d_k^q)^{\frac{1}{q}}$$

$$||cd||_1 \leq ||c||_p ||d||_q$$

$$\log \sum c_k d_k \leq \frac{1}{p} \log \sum c_k^p + \frac{1}{q} \log \sum d_k^q$$

$$||c_k^p d_k^{1-p}||_1 \leq ||c_k^p||_p ||d_k^{1-p}||_q$$

$$\log \sum c_k^p d_k^{1-p} \leq \frac{1}{p'} \log \sum (c_k^p)^{p'} + \frac{1}{q'} \log \sum (d_k^{1-p})^{q'}$$

$p' = \frac{1}{p}$ és $q' = \frac{1}{1-p}$, $\frac{1}{p'} + \frac{1}{q'} = 1$, így a következőt kapjuk:

$$\log \sum c_k^p d_k^{1-p} \leq p \log \sum (c_k^p)^{\frac{1}{p}} + (1-p) \log \sum (d_k^{1-p})^{\frac{1}{1-p}}$$

■

Egyenlőség akkor és csak akkor áll fenn, ha $\exists$ olyan $\xi > 0$, melyre $c_k = \xi d_k$ $\forall k$ -ra.

Egy log-likelihood függvény λ paraméterrel definíció szerint konkáv, ha minden paramétervektorára teljesül az, hogy α, β és $p \in (0,1)$,

$$\ell[p\alpha + (1-p)\beta] \geq p\lambda(\alpha) + (1-p)\lambda(\beta) \tag{21}$$

Szigorú konkávitásról beszélünk, ha $\alpha \neq \beta$ esetén, és ettől a feltételtől függ az is, hogy a (21) -es képletben szigorú egyenlőtlenségünk van-e, vagy nem.

A (20) pedig a következőt jelenti:

$$-\ln[e^{p\alpha_i+(1-p)\beta_i} + e^{p\alpha_j+(1-p)\beta_j}] \geq -p\ln(e^{\alpha_i} + e^{\alpha_j}) - (1-p)\ln(e^{\beta_i} + e^{\beta_j}) \tag{22}$$

Így megszorozzuk a (22) -es egyenlőtlenséget w_{ij}-vel és i-re és j-re összegzünk. Ez bizonyítja (19) log-likelihoodjának konkávitását.

A (20)-as képletben a Hölder egyenlőtlenségére is lehet használni az egyenlőség feltételeit. Ezekből a származtatott feltételekből következtethetünk az újra paraméterezett függvény szigorú konkávságára.

Harmadik feltevés

Ez egy enyhébb feltétel, ami garantálja, hogy a log-likelihood függvényünk konkáv legyen.

Két nemüres halmazaba soroljuk a versenyzőket. Valamely versenyzőt a második halmazból összehaonslítjuk valamely első halmazbelivel legalább egyszer.

2. Lemma

Az újra paraméterezésből adódóan $(\gamma, \theta) \rightarrow (\beta, \phi)$, melyben $\beta_i = \ln \gamma_i - \ln \gamma_1$ és $\phi = \ln \theta$, és legyen $\Omega' = \{\beta \in \mathbb{R}^m : \beta_1 = 0\}$, a következőket kapjuk:

(a) Az (3) és az (7) log-likehoodjainak újra paraméterezett változata szigorúan konkáv az Ω' paramétertéren akkor és csak akkor, ha a harmadik feltevés teljesül.

(b) A (13) újra paraméterezett változata szigorúan konkáv a $\Omega' \times \mathbb{R}_+$ -on, és a (15) újra paraméterezett változata szigorúan konkáv az $\Omega' \times \mathbb{R}$ -en akkor és csak akkor, ha a harmadik feltevés teljesül, és legalább egyszer volt döntetlen is.

(c) A (12) -as újra paraméterezett változata is szigorúan konkáv az $\Omega' \times \mathbb{R}$ -en, ha a harmadik feltevés teljesül, és van benne egy olyan hurok, hogy $(i_0, i_1, ..., i_s = i_0)$, úgy, hogy i_{j-1} otthon játszik, és legalább egy összehasonlítás van közötte, és i_j között úgy, hogy $1 \leq j \leq s$.

Mivel a feltételezés biztosítja az első lemmában adott felső kompaktságot, ezért ez erősebb, mint azok a feltételek, melyek biztosítják a szigorú konkávságot. A Ljapunov-tétel magában foglalja azt, hogy minden MM-algoritmus (ciklikus, vagy nem) garantáltan előállítja a paraméter vektoroknak olyan sorozatát, mely konvergál a maximum likelihood becsléshez, az első feltételezései mellett.

7. Több versenyző összehasonlítása

Nem csak kettő, vagy három versenyzőt hasonlíthatunk össze egymással, hanem egyszerre többet is.

Tekintsünk a Bradley-Terry modellnek egy olyan kiterjesztését, melyben $k \geq 3$ versenyzőt hasonlítunk össze. Majd az összehasonlításokat véve alapul, eredményként felállítunk egy rangsort, a legjobbtól egészen a legrosszabbig. Ez a szituáció merülhet fel például akkor, minden bíró csak néhány bejegyzést lát a versenyzőkről, majd rangsorolja a látott bejegyzéseket. Marden (1995) készített egy alapos felmérést az ilyen típusú modellről.

Tegyük fel, hogy adott m versenyző, és őket címkézzük 1-től m-ig. $A \subset \{1, ..., m\}$ és $A = \{1, ..., k\}$ $k \leq m$. Tegyük fel, hogy a versenyzők indexeltek az A halmazbeli rangsorral.

Jelölje $\rightarrow$ a kapcsolatot két versenyző között. A nyíl a *"Jobb helyen áll a rangsorban, mint..."* relációt jelenti. Például, ha az egyes játékos jobb, mint a kettes, az egyestől, a kettes felé mutat a nyíl. A rangsorban nyilván mindig a kisebb sorszámútól mutat a nagyobb sorszámú felé, hiszen az első mindig jobb, mint a második, a második mindig jobb, mint a harmadik, és így tovább.

Jelölje $\wp_k$ k versenyző permutációjának halmazát. Adott A és néhány $\pi \in \wp_k$. A valószínűség, amit pedig hozzárendelünk a $\pi(1) \rightarrow \pi(2) \rightarrow ... \rightarrow \pi(k)$ eseményhez, a következő:

$$P_A \left[\pi(1) \rightarrow \pi(2) \rightarrow ... \rightarrow \pi(k)\right] = \prod_{i=1}^{k} \frac{\gamma_\pi(i)}{\gamma_\pi(i) + ... + \gamma_\pi(k)} \tag{23}$$

Ezt az általánosítását a Bradley-Terry modellnek Marden (1995) Plackett-Luce modellnek nevezte, mivel először Plackett vezette be, 1975-ben.

Ha csak három versenyzőre tekintjük az összehasonlítást, a (23)-as képletben, az pont a Pendergrass-Bradley (1960) modellhez, azaz esetünkben a (7)-es képlethez vezet.

Az A minden részhalmazára, például $\{1, 2\}$-re értelmezhetünk olyat, hogy $P_A(1 \rightarrow 2)$ minthogy

$$\sum_{\pi \in \wp_k : \pi^{-1}(1) < \pi^{-1}(2)} P_A \left[\pi(1) \rightarrow ... \rightarrow \pi(k)\right]$$

Az összeg az $\{1, ..., k\}$ halmazból kapott minden rangsor valószínűsége, melyben $1 \rightarrow 2$, vagyis, az első versenyző legyőzi a másodikat, illetve jobb nála, tehát a rangsorban előrébb szerepel. Ideálisan, ennek a modellnek koherensnek kellene lennie ebben az esetben, különösen, hogy a rangsorolás valószínűsége nem függ attól, hogy a versenyzőket melyik részhalmazból vettük. Feltételezzük, hogy így is el tudjuk készíteni a modellt. Más szóval, ha (23) koherens, akkor az A indexelése $P_A\left[\pi(1) \to ... \to \pi(k)\right]$-ben nem szükséges. Tehát azt akarjuk, hogy A-tól ne függjön a valószínűség. Ekkor a (23) valószínűsége k versenyző összes olyan permutációja, hogy ha k-adik versenyző bármelyik helyen állhat, akkor az elsőtől a $(k-1)$-edik versenyzőig a többi milyen sorrendben állhat. (23) valószínűségét úgy kapjuk, ha k darab permutációt összeadunk. A számláló szorzata az összes γ szorzata, amit kiemelhetünk, így marad a nevezők szorzata, összesen k darab, amit összeadunk.

A koherencia bizonyítása:

Legyen $A = \{1, ..., k\}$, mint korábban, és kiértékelése a következő:

$$P_A(1 \to ... \to k-1) = \gamma_1 ... \gamma_{k-1}\gamma_k \left[\frac{1}{(\gamma_1 + ... + \gamma_k)...(\gamma_{k-1} + \gamma_k)\gamma_k} + \right. \\ \left. + \frac{1}{(\gamma_1 + ... + \gamma_k)...(\gamma_k + \gamma_{k-1})\gamma_{k-1}} + ...\right] \quad (24)$$

ahol az összeg k szempontjából megfelelő a k különböző permutációra $\wp_k$-ban. Az $(1, ..., k-1)$ sorrend változatlan marad. A (24)-es képlet leegyszerűsítve a következő lesz:

$$P_A(1 \to ... \to k-1) = \frac{\gamma_1 ... \gamma_{k-1}}{(\gamma_1 ... \gamma_{k-1})...(\gamma_{k-2} ... \gamma_{k-1})\gamma_{k-1}} = P_{(1,...,k-1)}(1 \to ... \to k-1). \quad (25)$$

A $P_A(1 \to ... \to k-1)$ részhalmazt helyettesíthetjük A bármely részhalmazával, $k-1$ elemmel, így a (25) -ös képlet használatánál az ismétlődés szükségszerű. Minden $B = \{b_1, ..., b_l\} \subset A$ -ra fennáll a következő:

$$P_A(b_1 \to ... \to b_l) = P_B(b_1 \to ... \to b_l). \quad (26)$$

Így a modell koherens, ezért felhagyhatunk az A és a B halmaz indexelésével, és egyszerűen csak $P(b_1 \to ... \to b_l)$ -t, vagy a rövidség kedvéért $P(b)$-t írunk. A versenyzők számát tartalmazza egy adott rangsor, de nem minden rangsorban kell az összes versenyzőnek szerepelnie. A találkozókon mindig más és más csapatokat, versenyzőket hasonlítunk össze, melyből a teljes szezon eredményeit össze tudjuk kombinálni, így minden csapatra, versenyzőre kapunk egy becslést, akárki nyer.

Röviden megemlítjük a (23) és Luce választási axiómája közötti kapcsolatot. Az axióma kimondja, hogy minden modellre, melyben i játékos pozitív valószínűséggel megveri j játékost, és a páronkénti összehasonlításban $i \neq j$, akkor

$$P_B(i \text{ nyer }) = P_A(i \text{ nyer })P_B(A \text{ részhalmazból valaki nyer }), \forall i \in A \subset B. \qquad (27)$$

Luce (1959) megmutatta, hogy a (27)-es aximóma egyenlő a következő állítással:

$$P_B(i \text{ nyer }) = \frac{\gamma_i}{\sum_{j \in B} \gamma_j} \qquad (28)$$

pozitív értékű γ_i paraméterekre. Nem nehéz látni, hogy a (23)-as képlet egyelnő a (28)-as állítással. Marden rámutat, hogy a (23)-as képlet igazából a (28)-as állításból ered. Ha elképzelünk egy rangsorolási folyamatot úgy, hogy elsőnek választjuk a győztest, aztán a második helyezettet úgy, hogy a megmaradt játékosok között nézzük a legjobbat, és így tovább. Az ellenkezője azért következik, mert (23) esetén:

$$P_A(i \text{ nyer }) = \sum_{\pi:\pi(1)=i} P_A\left[\pi(1) \to ... \to \pi(k)\right] =$$
$$= \sum_{\pi:\pi(1)=i} \frac{\gamma_i}{\gamma_1 + ... + \gamma_k} \prod_{j=2}^{k} \frac{\gamma_{\pi(j)}}{\gamma_{\pi(j)} + \gamma_{\pi(j+1)} + ... + \gamma_{\pi(k)}} = \frac{\gamma_i}{\gamma_1 + ... + \gamma_k}$$

Így a (23)-as képlet ekvivalens Luce választási axiómájával, ami magában foglalja a koherenciát, a fent meghatározott értelemben.

A (23)-as modell illesztéséhez használjuk a maximum-likelihoodot, ismét konstruálhatunk egy minorizáló függvényt a (8) egyenlőtlenség felhasználásával.

Tegyük fel, hogy N rangsorból állnak az adatok, ahol a j-edik rangsor magában foglalja m_j-t, ahol m_j-vel azt fejezzük ki, hogy hány versenyzőt hasonlítottunk össze. $1 \leq j \leq N$. Rendeljünk a versenyzőkhöz indexeket a j-edik rangsorolásban, melyeket a következőképpen jelöljünk: $a(j,1),...,a(j,m_j)$, úgy, hogy $a(j,1) \to a(j,2) \to a(j,m_j)$, és e szerint építsük fel a j-edik rangsort. Feltesszük, hogy a rangsorolások függetlenek, a log-likelihood a következőképp írható fel:

$$\ell(\gamma) = \sum_{j=1}^{N} \sum_{i=1}^{m_j-1} \left[\ln \gamma_{a(j,i)} - \ln \sum_{s=i}^{m_j} \gamma_{a(j,s)} \right]$$

A (8)-es egyenlőtlenséggel:

$$Q_k(\gamma) = \sum_{j=1}^{N} \sum_{i=1}^{m_j-1} \left[\ln \gamma_{a(j,i)} - \frac{\sum_{s=i}^{m_j} \gamma_{a(j,s)}}{\sum_{s=i}^{m_j} \gamma_{a(j,s)}^{(k)}} \right]$$

minorizálja a log-likelihood $\ell(\gamma)$-t $\gamma^{(k)}$-ban konstans együttható erejéig. A paraméterek szétválasztásával, és

$$\gamma_t^{(k+1)} = \frac{w_t}{\sum_{j=1}^{N} \sum_{i=1}^{m_j-1} \delta_{jit} \left[\sum_{s=i}^{m_j} \gamma_{a(j,s)}^{(k)} \right]^{-1}} \tag{29}$$

kifejezéssel érhetjük el $Q_k(\gamma)$ maximalizálását. $t = 1,...,m$, ahol w_t azoknak a rangsoroknak a száma, melyekben a t-edik versenyző előrébb áll a rangsorban, mint az utolsó, és

$$\delta_{jit} = \begin{cases} 1 & \text{ha } t \in \{a(j,i),...,a(j,m_j)\} \\ 0 & \text{máskülönben} \end{cases}$$

Más szóval δ_{jit} fejezi ki azt az lehetőséget, hogy t versenyző jobb rangot kap-e, mint i a j-edik rangsorban. A (29)-es a (10), sima Bradley-Terry modell általánosítása. A γ összetevőit lehet ciklikusan frissíteni.

Ebben az összefüggésben az első feltételezésnek akkor van értelme, ha tudjuk értelmezni azt, hogy i versenyző megveri a j játékost, és így i rangja magasabb, mint j rangja egy olyan rangsorban, mely mindkét játékost tartalmazza. Az 1(a) és a 2(a) lemmához az első feltételezés szükséges és elégséges, a log-likelihood függvény felső kompaktságára nézve. Mivel a harmadik feltételezés szükséges és elégséges a log-likelihood függvény szigorú konkávságságára nézve, így újra paraméterezhetünk. $\beta_i = \ln \gamma_i - \ln \gamma_1$. Arra a következésre juthatunk, hogy az MM-algoritmus garantálja a konvergenciát az egyetlen maximum-likelihood becsléshez, ha az első feltételezés fennáll. A Plackett-Luce modell likelihoodjának meghatározására nem ismerünk más algoritmust, Plackett szerint csak numerikus módszerekkel lehet meghatározni a likelihood maximumát.

8. Bradley-Terry modell R-ben

Egy valós példát tekintek, és ezt elemzem az R, statisztikai program segítségével. Az R-ben a Bradley-Terry modellt könnyen installálhatjuk, és akár beépített adatokra is működtethetjük. (Újságok összehasonlítása, baseball meccsek eredményei találhatók a beépített változatban, de most nem ezekre hagyatkozom.)

Jelen esetben férfi vízilabda mérkőzéseket nézek, Európa köztudottan négy élvonalbeli csapatára, azaz Magyarországra, Szerbiára, Horvátországra és Montenegróra. Az utolsó húsz mérkőzés eredménye szolgál alapul mindegyik csapatnak(2010. 09. 11-től visszamenően 2008. 08. 10-ig), mivel például a 2010-es zágrábi Európa Bajnokságon nem játszottak egymással olyan sokszor, hogy érdemleges modellt fel tudjunk állítani rájuk, így belekerült a 2009-es római világbajnokság, és a 2008-as pekingi olimpia is a megfigyelések közé. A 2010-es Európa Bajnokságon Zágrábban Horvátország lett az aranyérmes csapat. Tehát a hazai pálya minden számolás nélkül is nagy valószínűséggel előnyt jelentett az otthoniaknak.

De vizsgáljuk meg részletesebben a modellt!

A program működtetése:

A R programba úgy kell beírnunk az adatokat, illetve betöltetnünk a vizsgálandó txt vagy xls fájlt, hogy az első oszlpoba írjuk a nyertes nevét, a második oszlopba a vesztes nevét, a harmadik oszlopba pedig azt, hogy azokon a mérkőzéseken, mikor az adott két versenyző játszott egymással, az, amelyik a nyertes oszlpoban van, hányszor nyert.

Mivel az R program Bradley - Terry modelljébe nincs beépítve a döntetlen lehetősége, ezért a döntetlent úgy adjuk meg, hogy $1/2 - 1/2$ meccs megnyerését számítjuk azoknál a csapatoknál, melyek döntetlent játszottak egymással.

Ha figyelembe vesszük, hogy Magyarország és Montenegró egyszer játszott döntetlent egymással, és Magyarország csapata egyszer legyőzte Montenegró csapatát, akkor a következőképpen alakul a felírásunk, és a modellünk:

```
>  vl <- read.table("G:/vizilabda.txt")
> vl
        winner      loser Freq
1.     Hungary     Serbia  1.0
2.      Serbia    Hungary  2.0
3.     Hungary    Croatia  0.0
4.     Croatia    Hungary  0.0
5.     Hungary Montenegro  1.5
6.  Montenegro    Hungary  0.5
7.      Serbia    Croatia  1.0
8.     Croatia     Serbia  2.0
9.      Serbia Montenegro  2.0
10. Montenegro     Serbia  0.0
11.    Croatia Montenegro  1.0
12. Montenegro    Croatia  2.0

>
>
> library(BradleyTerry)
> vlModel <- BTm(vl ~ ..)
>  vlModel

Call:  BTm(formula = vl ~ ..)

Coefficients:
```

```
  ..Hungary  ..Montenegro      ..Serbia
    0.09098      -0.44107       0.44107

Degrees of Freedom: 5 Total (i.e. Null);  2 Residual
Null Deviance:      4.315
Residual Deviance: 3.449        AIC: 16.03
```

Az országokhoz rendelt együtthatók a β-kat határozzák meg a log-likelihood függvényben. A csapatok egyéni képességeit, vagyis a γ értékeket úgy kaphatjuk meg, ha e-t értelemszerűen a β-adik hatványokra emeljük. Azaz, ha az előbbi modellben Horvátországot veszi alapul, ehhez illeszti a modellt, neki a β értéke a log-likelihood modellben 0 lesz, míg a γ e^0, azaz 1. Magyarországnak $e^{0.09098}$, Montenegrónak $e^{-0.44107}$, és Szerbiának $e^{0.44107}$ a γ értéke, tehát a csapat képessége, ha nem a csapaton belüli versenyzők egyéni teljesítményeiből tevődik össze a csapat képessége. (Arra a modellre nem térünk ki.)

A devianciák a modell illeszkedését mérik. Az illeszkedést χ^2 próbával vizsgáljuk. Az eloszlás közelítőleg χ^2 eloszlású. Jelen esetben a szabadsági fok: 5 , ami világos, hiszen a szabadági fokot a χ^2 próbánál úgy számoljuk ki, hogy karakterek száma mínusz egy, vagyis $f = n - 1$. Jelen esetben a négy csapatból kiválasztunk kettőt, amit egymással hasonlítunk. Ezt négy alatt a kettő féleképpen tehetjük meg, ami hat lehetőség, majd kivonunk belőle egyet. Így adódik az eredmény. Tehát ez egy 5 szabadsági fokú, χ^2 eloszlású a vizsgált modellünk..

Mivel a Bradley-Terry modellből becsült paraméterekkel egy χ^2 próbát illesztünk a modellünkre, ezért az illeszkedésvizsgálat becsléses, a próbastatisztika szabadságfoka a becsült paraméterek számával csökken. A reziduális 2 lesz.

Null Deviance éa a Residual Deviance jelentésével később, a modell összesített vizsgálatánál foglalkozunk részletesebben.

Ugyanez a modell egy másik paraméterezés szerint, ahol a "refcat" paranccsal mi határozhatjuk meg azt, hogy melyik csapat β-ja legyen 0 a log-likelihood függvényben, vagyis, hogy melyikhez szeretnénk illeszteni.

```
> update(vlModel, . ~ ., refcat = "Hungary")

Call:  BTm(formula = vl ~ .., refcat = "Hungary")

Coefficients:
   ..Croatia  ..Montenegro      ..Serbia
    -0.09098      -0.53205       0.35010

Degrees of Freedom: 5 Total (i.e. Null);  2 Residual
Null Deviance:      4.315
Residual Deviance: 3.449        AIC: 16.03
>
>
> update(vlModel, . ~ ., refcat = "Serbia")

Call:  BTm(formula = vl ~ .., refcat = "Serbia")

Coefficients:
   ..Croatia     ..Hungary  ..Montenegro
     -0.4411       -0.3501       -0.8821

Degrees of Freedom: 5 Total (i.e. Null);  2 Residual
Null Deviance:      4.315
Residual Deviance: 3.449        AIC: 16.03
>
> update(vlModel, . ~ ., refcat = "Montenegro")

Call:  BTm(formula = vl ~ .., refcat = "Montenegro")

Coefficients:
..Croatia  ..Hungary   ..Serbia
   0.4411     0.5321     0.8821
```

```
Degrees of Freedom: 5 Total (i.e. Null);  2 Residual
Null Deviance:      4.315
Residual Deviance: 3.449        AIC: 16.03
>
```

Ha az R-be beépített újságokra vizsgálnánk Bradley-Terry modellt, nézhetnénk, hogy például az Amerikában kiadott lapok jobbak, vagy az Angliában kiadottak, de jelenlegi modellünkre, ha nagyon akarnánk, maximum azt vizsgálhatnánk, hogy a szlávok jobbak-e, mint a magyarok, vagy fordítva. A modell mindenesetre működik rá, csak nem mindegy, hogy Magyarországot hányadik helyre írjuk. A program betűrendben kéri be az adatokat.

```
> countryNames <- levels(vl$winner)
> countryData <- data.frame(origin =
c("Slavic", "HUN", "Slavic", "Slavic"), row.names = countryNames)
>  vlModel2 <- BTm(vl ~ origin, data = countryData)
> vlModel2

Call:  BTm(formula = vl ~ origin, data = countryData)

Coefficients:
originSlavic
           0

Degrees of Freedom: 5 Total (i.e. Null);  4 Residual
Null Deviance:      4.315
Residual Deviance: 4.315        AIC: 13.43
>
```

Így meglehetősen érdekes eredmény jött ki, ami azt mutatja, hogy a magyarok ugyanolyan jók, mint a szlávok.

Az egész modellre is elvégezhetjük az illeszkedésvizsgálatot.

```
> summary(vlModel)

Call:
BTm(formula = vl ~ ..)

Deviance Residuals:
   Hungary vs Croatia  Montenegro vs Croatia  Montenegro vs Hungary
          0.0000                 0.9621                -0.3621



    Serbia vs Croatia      Serbia vs Hungary   Serbia vs Montenegro
         -0.9621                  0.2849                 1.1770



Coefficients:
             Estimate Std. Error z value Pr(>|z|)
..Hungary     0.09098    1.24446   0.073    0.942
..Montenegro -0.44107    0.96771  -0.456    0.649
..Serbia      0.44107    0.96771   0.456    0.649

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4.3152  on 5  degrees of freedom
Residual deviance: 3.4489  on 2  degrees of freedom
AIC: 16.032

Number of Fisher Scoring iterations: 3
```

Itt z-ben standardizáljuk a becslést. A becsült paramétereket elosztjuk a szórással, vagyis a standard hibával, tehát az első oszlopot a másodikkal, így kapjuk meg z értékét. Ez standard normális eloszlású. Milyen eséllyel lesz negyedik oszlopnál nagyobb abszolútértékű? Itt u-próbát alkalmazunk, kétoldali ellenhipotézissel. Ebből azt látjuk a negyedik oszlop alapján, hogy egyik sem szignifikáns.

A Residual Deviance a Deviance Residuals négyzetösszege, a Null Deviance pedig az, ha mindnek nulla lenne az a bizonyos paramétere. A Deviance Residulas értékeket úgy számoljuk, hogy először egy általunk kiválasztott csapatpárra alkalmazzuk a sima Bradley-Terry modellt. Megnézzük, hogy ez a két csapat hányszor győzött a másik felett, illetve, hogy hányszor játszott egymással. Például, ha Magyarországot és Szerbiát hasonlítjuk össze, nézzük azt, ha Magyarország nyer. Ekkor a következőt írjuk be a Bradley-Terry modellbe: $\frac{e^{0.53}}{e^{0.53}+e^{0.88}} = 0.4134$. Ennyi az esélye, hogy Magyarország nyer, ha a szerbekkel játszik. Ezt megszorozzuk hárommal, mert ez a két csapat ennyiszer játszott egymással. Méghozzá úgy, hogy Magyarország egyszer, Szerbia kétszer nyert, így 1.24-et kapunk. Mivel most Magyarország nyerési esélyeit vizsgáljuk Szerbiával szemben, ezért felírhatunk rá egy olyan számítást, hogy: $\frac{1-1.24}{\sqrt{3 \cdot 0.41 \cdot (1-0.41)}} = -0.28$. Ha Szerbiára írtuk volna fel, ugyanez jött volna ki, csak pozitív előjellel. Ha a többire is alkalmazzuk ezt a számítást, megkapjuk a Deviance Residuals közelítő értékeit. Majd ha ezeknek vesszük a négyzetösszegét, akkor a Residual Devaince értékhez jutunk. Ezek megmutatják, hogy a reziduális lineáris trenddel becsült értéke a valós értéktől átlagosan mennyire tér el. Ahol nulla, ott jól illeszkedik. Ahol nagy, ott eléggé eltér.

Fisher Scoring: egy logisztikus regresszió van beépítve az R-be, mely itt most három iterációs lépés alatt konvergál. Három lépés kellett ahhoz, hogy hozzájussunk ezekhez a paraméterbecslésekhez.

Az egyes országok vízilabda csapatainak képességei, a versenyen nyújtott teljesítményük, azaz a γ paraméterek, ha a hazai pályát nem vesszük figyelembe, egyszerre meghatározhatók az R program által:

```
> BTabilities(vlMod)
              ability      s.e.
Croatia     0.0000000 0.0000000
Hungary     0.0909773 1.2444622
```

```
Montenegro -0.4410731 0.9677065
Serbia      0.4410731 0.9677065

> residuals(vlMod)
   Hungary vs Croatia Montenegro vs Croatia Montenegro vs Hungary
            0.0000000             0.9620876            -0.3620745

    Serbia vs Croatia     Serbia vs Hungary  Serbia vs Montenegro
           -0.9620876             0.2848893             1.1770257
```

Hazai pálya modell R-ben

Az adatainkat ugyanúgy betöltjük, mint az előzőekben. Az első oszlop a győztesé, a második a vesztesé, a harmadik az, hogy a győztes hányszor verte meg azt, aki tőle kikapott, de itt jön be egy negyedik oszlop is, mégpedig az, ami azt mutatja, hogy a győztes hazai pályán játszott-e, avagy nem. Ha hazai pályán játszott, és nyert, akkor egy 1-est írunk a negyedik oszlopunkba, az adott sor mellé, ha viszont idegenben tudott nyerni, akkor egy -1-est. (Ekkor a vesztes csapat játszott otthon.) Ha egyik csapat sem játszott otthon, a nevük mellé 0-t írunk.

A mi modellünk akkor így alakul:

```
>  vlM <- read.table("G:/vldontetlenhazai.txt")
> vlM
        winner      loser Freq home.adv
1.     Hungary     Serbia  1.0        0
2.      Serbia    Hungary  2.0        0
3.     Hungary    Croatia  0.0        0
4.     Croatia    Hungary  0.0        0
5.     Hungary Montenegro  1.5        0
6.      Serbia    Croatia  1.0        0
7.     Croatia     Serbia  1.0        1
8.      Serbia Montenegro  2.0        0
9. Montenegro     Serbia  0.0         0
10.    Croatia Montenegro  1.0        1
```

```
11. Montenegro    Croatia  1.0       -1
12. Montenegro    Hungary  0.5        0
13.    Croatia     Serbia  1.0        0
14. Montenegro    Croatia  1.0        0
```

Ha erre az adatsorra összegzünk, a hazai pálya szerint, a következőt kapjuk:

```
> vizilM <- update(vizilM, order.effect = vl$home.adv)
> summary(vizilM)

Call:
BTm(formula = vl ~ .., order.effect = vl$home.adv)

Deviance Residuals:
    Hungary vs Croatia 0   Montenegro vs Croatia 0   Montenegro vs Hungary 0
                  0.0000                    1.0950                   -0.4056
     Serbia vs Croatia 0       Serbia vs Hungary 0    Serbia vs Montenegro 0
                 -0.6742                    0.3204                    1.2313
Montenegro vs Croatia -1      Serbia vs Croatia -1
                  0.6742                   -1.0950

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
..Hungary      0.6637     1.5418   0.430    0.667
..Montenegro   0.1970     1.3854   0.142    0.887
..Serbia       0.9717     1.3006   0.747    0.455
.order         1.1687     1.7747   0.659    0.510

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.4082  on 7  degrees of freedom
Residual deviance: 5.0904  on 3  degrees of freedom
```

```
AIC: 19.268

Number of Fisher Scoring iterations: 4
```

Itt a program becsli a θ paraméter logaritmusát is, azaz $\phi = \log \theta$-t, az országok β-ján kívül. Az alap *Hazai pálya modell*be visszaírjuk e^{ϕ}-t, és az országok e^{β}-aidikon egyéni képességeit. Minden számítás hasonlóképp történik, mint abban az esetben, ha nem vesszük figyelembe a hazai pályán játszott meccseket.

9. Összefoglalás

Láttuk, hogy a Bradley-Terry statisztikai modellt párosított összehasonlításoknál alkalmazhatjuk, annak eldöntésére, hogy ki a jobb, melyik csapatra érdemesebb fogadni egy sportesemény előtt, melyik az esélyesebb az előzetes eredményei alapján. De nem csak sportra alkalmazhatjuk a modellt, hanem akár újságokra, könyvekre is. Ott az olvasottságot, a népszerűséget néznénk. Most a szakdolgozatomban csak sporttal foglalkoztam, de számos területen tudnánk alkalmazni.

A *Maximum-likelihood becslés*, az *EM-algoritmus* és az *MM-algoritmus* bemutatása után kezdtem a modellel részletesebben foglalkozni, mivel az EM-algoritmus az MM-algoritmus speciális esete, és az MM-algoritmust használjuk a Bradley-Terry modellnek, és általánosításainak egyetlen maximum likelihood becslésének megtalálására.

A Bradley-Terry modellnek számos általánosítása született az elmúlt 75-80 évben. Ilyen a *hazai pálya* esete, és a döntetlenek, melyekkel többen is foglalkoztak, és így több modell is született rá. Úgy, mint a *Rao-Kupper-féle*, vagy *Davidson-féle*. Az MM-algoritmussal ezekre mindre felírtuk a log-likelihood maximalizálását, majd láttuk, hogy nem csak kettő csapatot lehet összehasonlítani egymással, hanem hármat, vagy esetleg háromnál többet is.

Egy valós példát vettem, és az R, statisztikai program segítségével vizsgáltam, hogy Európa négy, kétségtelenül élvonalbeli vízilabda csapata ha egymással játszik, a nem olyan régen egymással játszott eredményeik alapján ki mennyire esélyes az első, a második, a harmadik, és a negyedik helyre. (Természetesen a becsléseket valamennyi hibával végezzük, illetve végzi a program, de a hibát is kiszámítja.) Láttuk, hogy egyik csapat sem szignifikáns a becsült eredmények alapján, de Szerbia csapata a legerősebb. Az R beépített Bradley-Terry csomagot használ, amit egy tükörgépről installáltam. A *Hazai pálya modell*t nem teljesen úgy tudtam alkalmazni, ahogy szerettem volna, mert nem állt rendelkezésemre elég adat, illetve a csapatok az utóbbi időben inkább olyan helyen mérkőztek meg egymással, ahol egyik sem volt otthon. Kivételt képez a 2010-es zágrábi kontinensviadal, ahol Horvátország nyert.

10. Köszönetnyilvánítás

Köszönöm Csiszár Villő tanárnőnek, hogy idejét rám áldozva, hétről hétre rengeteget segített, hogy a szakdolgozatom elkészülhessen.

Hivatkozások

[1] Lukács, O.: Matematikai statisztika, Műszaki Kiadó (2006).

[2] Csiszár, V.: Véletlen permutációk statisztikai vizsgálata. PhD disszertáció (2009).

[3] Csiszár, V.: Statisztika jegyzet. http://www.cs.elte.hu/∼villo/esti/stat.pdf (2009).

[4] Firth, D.: Bradley-Terry Models in R. *Journal of Statistical Software* **12** (2005).

[5] Hunter, R.D.: MM-algorithms for generalized Bradley-Terry models. *Ann. Statist.* **32** (2004).

[6] Heiser, J.W.: Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. W. J. Krzanowski, ed.: Recent Advances in Descriptive Multivariate Analysis. Oxford University Press, Oxford, UK (1995).