SPORTFOGADÁSI ÉS PÉNZÜGYI PIACOK KAPCSOLATA

MSc Diplomamunka

Írta: Csikai Mátyás

Biztosítási és pénzügyi matematika MSc

Kvantitatív pénzügyek szakirány

Eötvös Loránd Tudományegyetem, Természettudományi Kar

Budapesti Corvinus Egyetem, Közgazdaságtudományi Kar

Témavezető: Badics Milán Csaba

PhD Hallgató, Budapesti Corvinus Egyetem,

Befektetési és Vállalati pénzügyek Tanszék

2016

NYILATKOZAT

Név: Csikai Mátyás

ELTE Természettudományi Kar, szak: Biztosítási és pénzügyi matematika MSc

NEPTUN azonosító: GGIW5M

Szakdolgozat címe: Sportfogadási és pénzügyi piacok kapcsolata

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

Budapest, 2016. 05. 17.____________________________ *a hallgató aláírása*

Tartalom

Bevezetés

A dolgozat témája a sportfogadási piac, szűkebben a futball mérkőzések, meccs közbeni (in-game) fogadásainak árazásának és ezek fedezésére alkalmas matematikai keretrendszer áttekintése. Ennek során több szempontból is vizsgáltam a hasonlóságokat a sportfogadási és a pénzügyi piacok hasonlóságait. A pénzügyi piacra is teljesülő *Eszközárazás Alaptétele* lehetővé teszi, hogy arbitrázs mentes árakat tudjunk adni a mérkőzéseken alapuló pénzügyi termékekre. A téma aktualitásának alapja, hogy a fogadások népszerűsége ugrás-szerűen nőtt az elmúlt években. Továbbá életre hívott egy korábban nem létező piacot, melynek lényege, hogy a fogadásokat már nem csak a fogadóirodákkal köthetjük, hanem a fogadók egymás között is kereskedhetnek. Az ezen belüli árazási módszerekkel kapcsolatban egyelőre kevés kutatás történt, így rengeteg módszertani kérdés merül fel. Az árazás alapja a lőtt gólok modellezése lesz, melyek esetében nem konstans intenzitású, Poisson eloszlást veszek alapul. Az induló intenzitás értékének és annak mérkőzésen belüli változásainak becslése historikus adatok és saját feltételezések alapján történt.

Az *in-game* fogadások népszerűsége rengeteget nőtt a közelmúltban. Azonban annak lehetősége, hogy a fogadásokat nem csak a fogadóirodákkal köthetjük, hanem akár a fogadók egymás között is kereskedhetnek ezekkel, sokaknak még ismeretlen. Az első ilyen lehetőséget kínáló nemzetközi fogadó-oldal a betfair.com volt. A honlap rengeteg fogadási lehetőséget kínál és „tőzsde” oldalán minden élő fogadás kereskedhető is. A lényeges különbség például a *bWin.com* és hasonló oldalakkal szemben, hogy bár ezeken is eladhatjuk élő fogadásainkat, de ezekben az esetekben csak a fogadóirodával kereskedhetünk egy általuk meghatározott áron. Ezzel szemben a *betfair.com*-on egy kereskedési könyvbe rakhatjuk az általunk eladni kívánt fogadásainkat, általunk meghatározott áron, a könyv tartalma ennek megfelelően dinamikusan változik. A honlapon mindig az aktuálisan legjobb eladási és vételi árak jelennek meg, melyekhez, a természetes piaci feltételek alapján egy bid-ask spread is tartozik. Ezen piacok rendszere sokban hasonlít a pénzügyi piacokhoz. Vannak azonban érdekes eltérések is.

Mivel gyakoriak a nagy ugrások az árakban a mérkőzések folyamán, ezért minden góleseményt követően rövid időre befagyasztják a könyvet, hogy az eredmény módosulásának megfelelően az összes fogadást újraárazhassa kiírójuk.

Később látni fogjuk, hogy a piac elméletben arbitrázsmentes, azonban ilyen esetekben a valóságban jelentős arbitrázslehetőségek jöhetnének létre. Mivel ez a befagyasztás csak rövid ideig (nagyjából 1 percig) tart, a kereskedőknek szükségük lehet gyors és hatékony árazó formulákra, melyek segítségével fogadásaikat minél pontosabb áron tudják visszatenni a kereskedési könyvbe a fagyasztás után. Természetesen a piacon magától is elég hamar visszaállnak a fair árak, melyek alapján később mindenki könnyebben tudja meghatározni, hogy hogyan érdemes fogadását értékelni, azonban ha valaki megfelelő árazási formulákkal is rendelkezik, piaci előnyre tehet szert. Emellett egyéb módokon is ellenőrzik az arbitrázsok kialakulását, a minél nagyobb technikai hatékonyság elérése érdekében.

Az oddsokat a mérkőzésen történő összes esemény módosíthatja, természetesen a legjelentősebb hatást persze a gólesemények okozzák. A mérkőzés állását tekinthetjük egyfajta underlying terméknek, melyre az egyes sportfogadási lehetőségek derivatívákként is kezelhetőek. Ennek megfelelően a legelterjedtebb sportfogadási termékeket megfeleltethetjük adott pénzügyi termékeknek. A következő fejezetben ezek közül néhányat fel is sorolok.

A sportfogadási piacok jellemzőit dolgozták fel Peter Divos, Sebastian del Bano Rollin, Tomaso Aste és Zsolt Bihary [2015], belátták, hogy teljesül rá az eszközárazás alaptétele, ez alapján pedig a fogadások kockázatmentes környezetben árazhatóak. Árazó modelljükben, mivel egy mérkőzésen esett gólok száma önmagában nem kereskedhető termék, olyan helyettesítő termékeket vezettek be, melyek értéke a mérkőzés végén megegyezik a csapatok által szerzett gólok számával. E termékek előállításának alapjául konstans intenzitású Poisson eloszlású valószínűségi változókat vettek. Ezen valószínűségi változók voltak hivatottak leírni a mérkőzés adott pillanatában a meccs végére vonatkozó gólok számának eloszlását. Azonban tapasztalati úton jogosnak tűnik az a feltételezés, hogy ez az intenzitásparaméter az idő múlásával változik, sőt általános jellemzője az, hogy a mérkőzés vége fele megnő. A paramétert historikus adatok alapján számították és nem csak a meccs időtartama alatt, hanem az adott csapat összes mérkőzésére is általánosnak tekintették.

Szakdolgozatom célja e két feltételezés finomítása továbbá egy nem konstans függvény illesztése. Először egy pozitív együtthatójú másodfokú polinomot szerettem volna illeszteni, azonban a vártakkal ellentétben a modell nem bizonyult szignifikánsnak. Az ezt követően alkalmazott lineáris regressziós modell - ahol a magyarázó változó az idő, a magyarázott változó pedig a csapat által az adott percben szerzett gólok száma több meccsre tekintve – már szignifikáns eredményt adott. A konstans tagot úgy választottam, hogy a függvény időskála szerinti integrálja megegyezzen a historikusan becsülhető konstans intenzitásfüggvény integráljával. Ennek a feltételezésnek az az alapja, hogy csak a gólok meccsen belüli eloszlásán szerettem volna finomítani, a több mérkőzésre vonatkozó intenzitás valószínűleg jól jellemzi az átlagos gólszámot csapatonként, így érdemes ehhez igazítanunk a becslésünket.

In-game sportfogadási termékek

Az elmúlt években nagyban nőtt az in-game fogadások népszerűsége, egyes fogadóirodáknál meghaladta a pre-game fogadások forgalmát, de már kifejezetten erre épülő fogadó irodák (mint például a korábban említett *betfair.com*) is jelen vannak a piacon. Az információs technológia robbanásszerű fejlődésével nem jelent gondot az, hogy az aktuális meccset akár mobiltelefonon, vagy interneten kövessük, és így bárhol bármikor fogadhassunk.

Hagyományosan a sportfogadások esetében, a mérkőzés ideje alatt történő, a mérkőzéshez kapcsolódó, szinte összes eseményre lehet fogadni. Ezek közül természetesen a meccs végeredményére szóló fogadások a legnépszerűbbek (azaz, hogy melyik csapat nyer illetve milyen arányban). A mérkőzésen esett gólok számától kezdve, a hosszabbítások időtartamán át, olyan egészen egyedi fogadásokat is lehet kötni, mint például a 2015-ös Super Bowlon, hogy Eli Manning legalább négyszer lesz-e mutatva a közvetítés során a lelátón, amíg bátyja Payton Manning a pályán játszik. Természetesen, mi azon jellegű fogadásokat fogjuk vizsgálni, melyek „alaptermékéül" a lőtt gólok száma szolgál valamilyen formában.

Az in-game fogadások egyik jellemző eltérése a pre-game fogadásokhoz képest, hogy a mérkőzésen történt események folyamatosan, nagymértékben változtathatják az aktuális oddsokat. Ez további két csoportra bontható: az „odds" fogadás és a „spread" fogadás. Az első típusba az olyan fogadások tartoznak, melyeknek egy adott esemény bekövetkezése alapján definiált a kifizetésük, azaz ha az esemény bekövetkezik, akkor egy bizonyos összeget fizetnek, ha nem akkor semmit, azaz a digitális opciókra hasonlítanak leginkább. Ezek közül is a fent említettek a legelterjedtebbek. A „spread" típusú fogadások olyanok, melyek esetén a nyeremény összege közel folytonosan változhat, példának okáért egy csapat labdabirtoklási százalékát forintban kifizető fogadás. Mi a továbbiakban árazás szempontjából csak „odds" típusú fogadásokkal foglalkozunk majd.

Ahogyan már említettem, az odds típusú fogadások digitális opciókra hasonlítanak. Minden esetben egy előre fixált kifizetés történik, amennyiben bizonyos feltételek teljesülnek, és az egyes fogadások ezekben a feltételekben térhetnek el. A teljesség igénye nélkül az ismertebb fogadási típusok:

Eredményre vonatkozó fogadások:

Játékidő eredménye: A legismertebb fogadás típus. A mérkőzés végeredményére, mint három értékű vektorra lehet fogadni, ahol az *1*-es általában a hazai, a *2*-es a vendéggyőzelmet, míg az *X* a döntetlent jelöli. Akármelyik fogadást is tesszük meg, lényegében a gólok relatív számára veszünk egy digitális opciót. *1*-es esetben arra fogadunk (azaz az opció akkor fizet), hogy a hazai gólok száma nagyobb a vendég gólokénál és hasonlóan a többi esetben is. Hasonló ehhez a **félidei eredmény**, ahol ilyen formában tippelhetünk a félidő eredményre.

Dupla esély: Technikailag ez az előző típus ellentettje. A három lehetséges kimenetel közül az egyik nem-bekövetkezésére fogadunk. Tehát *12, X1,* és *X2* fogadásokat tehetünk.

DNB (Draw No Bet): A döntetlent kizárják a fogadásból, csak valamelyik csapat győzelmére fogadhatunk. Döntetlen esetén ezért visszakapjuk a megjátszott tétet. Ez az előzőekhez képest abban tér el, hogy a kifizetési függvény három értékű: 0, ha a másik csapat nyer, az opció díja amennyiben döntetlen, és az előre meghatározott nyeremény, ha jól tippeltünk.

Pontos eredmény: Ahogyan a neve is mutatja, csak a pontos eredmény eltalálása esetén fizet.

Félidő/Vége: A két eredményt összekötve fogadjuk meg, így három helyett kilenc kimenetel közül kell eltalálnunk egyet. *(1/1; 1/2; 1/X; 2/1; 2/2; 2/X; X/1; X/2; X/X).*

Hendikep (Handicap): Ebben az esetben, valamelyik csapatnak 1 vagy több gólos előnyt adnak. Ezzel végeredmény mellett a gólkülönbségre teszünk még fogadást. Ha például a *Sporting – Benfica* meccsen a *Benfica (+1)* megjelölés alatt lehet fogadni az *{1,2,X}*-es kimenetelekre, ebben az esetben az *1*-es fogadás azt jelenti, hogy a *Sporting* legalább 2 góllal nyer, a *2*-es azt, hogy a *Benfica* döntetlent ér el, vagy nyer, az *X* pedig a *Sporting* 1 gólos győzelmére fogad. Így a megfelelő hendikepek mellett az eredmények és gólkülönbség párok széles körére fogadhatunk.

Gólokra vonatkozó fogadások:

Gólok alatt/felett: A fogadás tárgy, hogy egy előre meghatározott számnál több gól esik-e a mérkőzésen. Az odds fogadási típuson belül természetesen itt is előre meghatározott a kifizetés, ha a feltétel teljesül, de lehetséges lenne a gólok számának növekedésével növekvő kifizetésű fogadást is kiírni. Hagyományosan nem egész értékű számokat adnak meg határként (jellemzően $n + 0.5$ alakúakat) így biztosítva, hogy csak két lehetséges kimenetele legyen a fogadásnak.

Gólok száma pontosan/hazai-/vendég csapatnál: A meccsen esett összes gólok számára, illetve külön-külön a csapatok által szerzett gólokra is fogadhatunk.

Gólt szerző csapat: Ez a fogadás négy kimenetelű. Fogadhatunk arra, hogy mindkét, egyik sem, vagy egy adott csapat szerez gólt.

Következő gól szerzője: A fogadás arra vonatkozik, hogy a mérkőzésen eső következő gólt melyik csapat fogja szerezni.

Spread típusú fogadások:

Point spread: A csapatok közötti gól/pontkülönbségtől függ a kifizetés lineárisan. Első sorban nem a labdarúgásra, hanem nagyobb pontszámú játékokra jellemző. Pénzügyi vonatkozásában, ha az irányra is fogadunk, azaz, ha az egyik csapat szemszögéből nézve pozitív és negatív különbségeket nézünk, inkább hasonlít a fogadás egy európai opcióra. A kifizetés 0, ha a „csapatunk" veszít, és a kifizetés a szerint nő, hogy mennyivel nyer. Lehetséges kétirányú fogadást is tenni, azaz csak a különbség mértékére vonatkozik a kifizetés, arra nem, hogy melyik csapat nyert. Ekkor egy terpeszpozíciónak feleltethető meg a fogadásunk.

Goal minutes: A kifizetés a gólok időpontjainak (a meccs kezdetétől nézve, hányadik percben esnek) összegére vonatkozik. Például, ha *Sporting – Benfica* mérkőzésen a 32. és 68. percben esik gól, és úgy kötöttük a fogadást, hogy ezek összegének 100-szorosát kapjuk, akkor 10000 HUF lesz a kifizetésünk.

Az oddsok jelölésére két elterjedt konvenció létezik. Mindkettőre jellemző az, hogy a fogadás, amire kiírják, olyan jellegű, hogy az esemény bekövetkezése esetén fix összeget fizet, ellenkező esetben nem fizet semmit. Az egyik esetben a nettó kifizetés és a tét arányát adja meg ez a mérőszám (frakcionális), a másik esetben pedig a bruttó (tehát a teljes kézhez kapott összeg) és a tét hányadosa adott (decimális). A becslésekhez felhasznált adatok többsége a decimális jelölési rendszert veszi alapul. A dolgozat folyamán az X_t jelöli majd t időpillanatban azon aktuális fogadás megkötéséhez szükséges összeget, amely a hozzá tartozó esemény bekövetkezése esetén pontosan 1 forintot fizet, amennyiben nem következik be, akkor pedig 0-t. Tehát az opciónk kifizetését rögzítettük, ehhez fogunk egy fair árat meghatározni.

Matematikai keretrendszer

A dolgozat során egy kockázatmentes árazási modellt fogunk alkalmazni az in-game fogadásokra. A szokásos módszertan szerint az underlying termék dinamikáját fogjuk felírni, megfelelő valószínűségi mezőn. Korábban belátták, hogy a Termékárazás két alaptétele érvényes erre a piacra Divos et. al. [2015] alapján, ennek eredményeképp pedig létezik delta-hedge minden ezen a piacon kiírt derivatívára. E tételek finomításával és fejlesztésével korábban foglalkoztak még Cox és Ross [1976], Harrison és Kreps [1979], Harrison és Pliska [1981, 1983], Huang [1985], Duffie [1988], valamint Back and Pliska [1991].

Sajnos a gólok száma, mint alaptermék nem kereskedhető ezért olyan alapterméket kell találnunk mellyel hatásosan helyettesíthetjük a fentit (azaz értékük megegyezik a mérkőzés végén), és lehetséges a kereskedése. A következőkben felírt termékek jól alkalmazhatóak matematikailag, azonban bármilyen két lineárisan független fogadás alkalmas lenne alapterméknek.

Megjegyzés: A Poisson eloszlás illesztésének jóságát a mérkőzés végeredményére többen vizsgálták, bár korábban Moroney [1956], Reep and Benjamin [1968], Reep et. al. [1971] arra is találtak bizonyítékot, hogy a Negatív-binomiális eloszlás illesztése pontosabb. Maher [1982] azonban független Poisson, majd kétváltozós Poisson eloszlások illesztésével is igen jó eredményeket ért el. Az *in-game* fogadásokra vonatkozóan pedig Dixon és Robinson [1998] adtak meg állapotfüggő Poisson-modellt, mely figyelembe vette az aktuális állást vagy a hazai pálya előnyét is. Analitikai okokból standard Poisson eloszlást alkalmazott Fitt et. al. [2005], Spread típusú fogadások vizsgálatára.

Vegyünk tehát egy $(\Omega, \mathcal{F}, P)$ valószínűségi mezőt, valamint két Poisson folyamatot: N_t^1-et és N_t^2-t, amik intenzitásai rendre μ_t^1 és μ_t^2, ahol $\{\mathcal{F}_{t\in[0,T]}\}$ a folyamat által generált filtráció. Ezek a folyamatok hivatottak rendre a hazai és a vendégcsapat góljainak a számát reprezentálni. Jelölje T a játék végének időpontját, és t az aktuális időpillanatot. A valószínűségi mértéknek a fizikai valószínűségi mértéket tekintjük.

Tekintsük a következő három terméket:

- B_t: A kockázatmentes kötvény
- S_t^1: A hazai csapat góljainak számával megegyező értékű termék
- S_t^2: A vendégcsapat góljainak számával megegyező értékű termék

1. Definícó: A modell dinamikáját az alábbi összefüggések írják tehát le:

- $B_t = 1$
- $S_t^1 \ = N_t^1 \ + \ \lambda_t^1 \, (T - t)$
- $S_t^2 \ = N_t^2 \ + \ \lambda_t^2 \, (T - t)$

Ahol λ_t^1 és λ_t^2 valós determinisztikus függvények.

A fogadások árazása:

2. Definíció (stratégia): Egy kereskedési stratégia egy olyan $\varphi_t \ = \ (\varphi_t^0, \varphi_t^1, \varphi_t^2)$ vektor, amelyre $\int_0^t |\varphi_s^i| ds < \infty$, minden $i \in \{0, 1, 2\}$. A vektor elemei rendre azt jelölik, hogy adott t időpillanatban mekkora mennyiséget tartsunk az adott termékből a portfólióban. Ehhez természetesen szükséges, hogy t-ben ismerjük az értéküket, tehát $\mathcal{F}_t$-mérhetőnek kell lenniük. Tehát az értékfolyamat:

- $V_t^\varphi \ = \ \varphi_t^0 B_t \ + \ \varphi_t^1 S_t^1 \ + \ \varphi_t^2 S_t^2.$

Egy stratégia önfinanszírozó, ha:

- $V_t^\varphi = V_0^\varphi + \int_0^t \varphi_s^1 dS_s^1 + \int_0^t \varphi_t^2 S_t^2$

ahol a fenti integrálok Lebesgue-Stieltjes integrálok [Felt. 2.3.2. Brémaud, 1981]

3. Definíció (arbitrázs mentesség): Egy modell arbitrázs mentes, ha nem létezik olyan kereskedési stratégia, amely egy valószínűséggel nem veszteséges, és pozitív eséllyel profitot eredményez, azaz:

- $\boldsymbol{P}[V_t^\varphi - V_0^\varphi \geq 0] = 1, \text{illetve } \boldsymbol{P}[V_t^\varphi - V_0^\varphi > 0] > 0$

4. Definíció (fogadás): Egy *fogadás* egy, az $\mathcal{F}_T$ szerint mérhető valószínűségi változó, X_T.

5. Definíció (teljesség): Egy modell teljes, ha minden fent definiált X_T-hez létezik őt előállító, önfinanszírozó stratégia. Azaz erre a stratégiára $X_T = V_t^\varphi$.

6. Tétel (kockázat semleges mérték): Létezik olyan Q valószínűségi mérték (kockázat semleges mérték), amely ekvivalens a P mértékkel és teljesül rá:

a) *A termékek B_t, S_t^1, S_t^2 értékfolyamata Q martingál.*
b) *A gólok számát leíró N_t^1 és N_t^1 folyamatok Q mérték szerint is Poisson-folyamatok λ_t^1 és λ_t^2 intenzitásokkal (melyek általában eltérnek az eredetiektől).*
c) *Q és P ekvivalensek.*
d) *Q egyértelmű.*

Bizonyítás: A létezés bizonyítása Girsanov egyik tételén alapulva következik a 2., 3. és 8. tételekből (165., 166. és 64. old.), Brémaud [1981], az egyediség pedig a 27. tételen, Brémaud [1981] és a 9.5 feltételen, Tankov [2004] alapszik.

■

7. Megjegyzés: A mértékcsere során a Poisson-folyamatok esetében nem a drift tag változik, hanem az intenzitás.

8. Tétel (arbitrázs mentesség): A modell arbitrázs mentes és teljes.

9. Következmény: A fogadás t-beli értéke megegyezik a kockázat semleges mérték szerinti várható értékével a meccs végén, azaz T-ben.

- $X_t = \mathbf{E}^Q[X_T|\mathcal{F}_t]$

Bizonyítás: Egyenesen következik a Harrison és Pliska [1981] 3.31 feltételéből.

10. Következmény: Ez az érték továbbá megegyezik a hozzá tartozó önfinanszírozó stratégia *t* időpontbeli értékével is.

- $X_t = V_t^{\varphi} = V_0^{\varphi} + \int_0^t \varphi_s^1 dS_s^1 + \int_0^t \varphi_t^2 S_t^2$

Bizonyítás: Egyenesen következik a Harrison és Pliska [1981] 3.32 feltételéből.

11. Definíció (lineáris függetlenség): $Z_{\boldsymbol{T}}^{\mathbf{1}}$ és $Z_{\boldsymbol{T}}^{\mathbf{2}}$ fogadások *lineárisan függetlenek,* ha az őket előállító önfinanszírozó stratégiákhoz tartozó $\varphi_t^1 = (\varphi_t^{10}, \varphi_t^{11}, \varphi_t^{12})$ $és$ $\varphi_t^2 = (\varphi_t^{20}, \varphi_t^{21}, \varphi_t^{22})$ vektorok majdnem mindig lineárisan függetlenek. Azaz minden $t \in [0, T]$ –re és $c_1, c_2 \in \mathbb{R}$ konstansra:

- $c_1\, \varphi_t^1 \neq c_2\, \varphi_t^2,$ P mérték mellett.

12. Tétel: Minden X_T fogadás replikálható két lineárisan független fogadás ($Z_{\boldsymbol{T}}^{\mathbf{1}}$ és $Z_{\boldsymbol{T}}^{\mathbf{2}}$*)* dinamikus kereskedésével, az alábbi formában:

- $X_T = X_0 + \int_0^t \psi_s^1 dZ_{\boldsymbol{s}}^1 + \int_0^t \psi_s^2 dZ_{\boldsymbol{s}}^{\mathbf{2}},$

ahol:

- $\int_0^t \psi_s^1 dZ_{\boldsymbol{s}}^{\mathbf{1}} = \int_0^t \psi_s^1 \varphi_s^{11} dS_{\boldsymbol{s}}^1 + \int_0^t \psi_s^2 \varphi_t^{12} dS_s^2,$

és hasonlóan $\int_0^t \psi_s^2 dZ_{\boldsymbol{s}}^{\mathbf{2}}$, illetve a ψ_s^1, ψ_s^2 kielégítik a:

- $\begin{pmatrix} \varphi_t^{11} & \varphi_t^{12} \\ \varphi_t^{21} & \varphi_t^{22} \end{pmatrix} \begin{pmatrix} \psi_s^1 \\ \psi_s^2 \end{pmatrix} = \begin{pmatrix} \varphi_t^1 \\ \varphi_t^2 \end{pmatrix}$

egyenletet.

Bizonyítás: Behelyettesítve, hogy $dZ_{\boldsymbol{t}}^{\mathbf{1}} = \varphi_t^{11} dS_{\boldsymbol{t}}^1 + \varphi_t^{12} dS_{\boldsymbol{t}}^2$ és $dZ_{\boldsymbol{t}}^{\mathbf{2}} = \varphi_t^{21} dS_{\boldsymbol{t}}^1 + \varphi_t^{22} dS_{\boldsymbol{t}}^2$ valamint a 10. Következményből és a 12. Tétel első egyenletéből következik az állítás.

Európai típusú fogadások:

13. Definíció (Európai típusú fogadás): Az európai opciókhoz hasonlóan, olyan fogadás, melynek az értéke csak a mérkőzés végi gólok számától függ. Ilyenek például a *Végeredmény, Pontos végeredmény*, *Hendikep,* stb.

- $X_T = \prod(N^1(T), N^2(t))$

Ahol $\prod$ ismert,$\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$.

14. Feltétel (árazó formula): Az európai típusú fogadások *t*-beli értéke explicit formában megadható:

$$X_t = \sum_{n_1=N_t^1}^{\infty} \sum_{n_2=N_t^2}^{\infty} \prod(n_1, n_2) P\big(n_2 - N_t^1, \lambda_1(T-t)\big) P(n_2 - N_t^2, \lambda_2(T-t))$$

ahol $P(N, \Lambda) = \frac{e^{-\Lambda}}{N!} \Lambda^N$, ha $N \geq 0$, illetve $P(N, \Lambda) = 0$ egyébként.

Bizonyítás: Közvetlenül adódik a 13. definícióból és a 10. következményből. ■

15. Definíció (greeks): A *greekek* a következő egyenletekből adódnak:

- $\delta_1 X_t(N_t^1, N_t^2) = X_t(N_t^1 + 1, N_t^2) - X_t(N_t^1, N_t^2)$
- $\delta_2 X_t(N_t^1, N_t^2) = X_t(N_t^1, N_t^2 + 1) - X_t(N_t^1, N_t^2)$
- $\partial_t X_t(N_t^1, N_t^2) = \lim_{dt \to 0} \frac{1}{dt}[X_{t+dt}(N_t^1, N_t^2) - X_t(N_t^1, N_t^2)]$

Ebben az esetben a δ_1, δ_2 a Deltának, míg a ∂_t a Thetának megfelelő szerepet tölti a Black-Sholes modell szerint.

16. Tétel (Kolmogorov egyenlet): Egy európai fogadás $X_t(N_t^1, N_t^2)$, adott $\prod(N_t^1, N_t^2)$ kifizetéssel, kielégíti a következő *Feynman-Kac* reprezentációt $[0, T]$-n:

- $\partial_t X_t(N_t^1, N_t^2) = -\lambda_1 \delta_1 X_t(N_t^1, N_t^2) - \lambda_2 \delta_2 X_t(N_t^1, N_t^2)$
- $X_t(N_t^1, N_t^2) = \prod(N_t^1, N_t^2)$

Bizonyítás: A bizonyítás megtekinthető például Feller [1940] 13. egyenlete alapján.

Ez másnéven a *Kolmogorov-egyenlet.*

Megjegyzés: Ennek az egyenletnek következménye az is, hogy minden *delta-semleges* (nem változik az értéke, ha valamelyik csapat gólt szerez) európai fogadásokból álló portfólió *theta-semleges* (azaz a gólok között sem változik az értéke). Megjegyzem továbbá, hogy ez általánosan minden fogadásra érvényes (nem bizonyítjuk).

17. Következmény: Az $X_t(N_t^1, N_t^2, \lambda_1, \lambda_2)$ európai típusú fogadás értékére igaz az alábbi egyenlet:

- $\frac{\partial}{\partial \lambda_i} X_t = (T - t)\delta_i X_t$

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

Bizonyítás: Ez a 14. (árazó-formula) formulából következik.

18. Feltétel (portfólió súlyok): Az európai fogadást előállító stratégia komponenseire:

- $\varphi_t^1 = \delta_1 X_t(N_t^1, N_t^2)$
- $\varphi_t^2 = \delta_2 X_t(N_t^1, N_t^2)$
- *Bizonyítás:* Felhasználva a korábbiak alapján, hogy

$$X_T = X_0 + \int_0^t \varphi_s^1 dS_s^1 + \int_0^t \varphi_s^2 dS_s^{\mathbf{2}},$$

a $dS_s^{\boldsymbol{i}} = dN_t^i - \lambda_i dt$ helyettesítést követően:

- $X_t = X_0 + \int_0^t (\varphi_t^1 \lambda_1 + \varphi_t^2 \lambda_2) ds + \sum_{k=0}^{N_t^1} \varphi_{t_k^1}^1 + \sum_{k=0}^{N_t^2} \varphi_{t_k^2}^2,$

ahol $\int_0^t \varphi_s^i dN_s^i = \sum_{k=0}^{N_t^i} \varphi_{t_k^i}^i$, ahol $0 \leq t_k^i \leq t$ a k. ugrás ideje (azaz a k. gól időpontja) az N_t^i folyamat esetén $i \in \{1,2\}$-re.

A másik irányból pedig az Ito-formulát alkalmazhatjuk ugró folyamatokra (Tankov [2004], 8.15 feltétel), mivel az árazó-formula folytonosan differenciálható:

- $X_t = X_0 + \int_0^t \partial_s X_s(N_s^1, N_s^2)ds$

$$+\sum_{k=0}^{N_t^1} \delta_1 X_{t_k^1}\left(N^1_{t_k^1-}, N^2_{t_k^1-}\right)$$

$$+\sum_{k=0}^{N_t^2} \delta_2 X_{t_k^2}\left(N^1_{t_k^2-}, N^2_{t_k^2-}\right)$$

ahol $t_k^i -$ arra utal, hogy a folyamat értékét közvetlenül az ugrás előtt vesszük. Mivel e két egyenletet minden ugrás időpontra igaz, és így az összegek belseje is meg kell, hogy egyezzen.

A termékek fedezése

A korábbiakban bebizonyosodott, hogy az in-game sportfogadási piacra teljesül a Termékárazás két alaptétele. Ez alapján létezik fedezési stratégia a pozíciókra, az eredeti feltételeink mellett.

Minden X_t termék ezen a piacon előállítható két, lineárisan független termékből. Továbbá teljesülnie kell annak, hogy a replikáló portfólió értékének változása meg kell, hogy egyezzen a sportfogadás értékváltozásával, minden gól eseményt követően. Ebből adódnak a következők:

Legyen a két lineárisan független termék a piacon $Z_{\boldsymbol{t}}^1$ és $Z_{\boldsymbol{t}}^2$:

- $X_t = X_0 + \int_0^t \psi_s^1 dZ_{\boldsymbol{s}}^1 + \int_0^t \psi_s^2 dZ_{\boldsymbol{s}}^{\mathbf{2}},$

ahol a ψ_s^1, ψ_s^2 kielégítik a:

- $\begin{pmatrix} \delta_1 Z_{\boldsymbol{t}}^1 & \delta_1 Z_{\boldsymbol{t}}^2 \\ \delta_2 Z_{\boldsymbol{t}}^1 & \delta_2 Z_{\boldsymbol{t}}^2 \end{pmatrix} \begin{pmatrix} \psi_t^1 \\ \psi_t^2 \end{pmatrix} = \begin{pmatrix} \delta_1 X_t \\ \delta_2 X_t \end{pmatrix}$ egyenletet,

ahol a δ_1 és δ_2 a véges első és második parciális differenciáloperátorok a 15. Definícióból.

Érdemes a lineárisan független termékeinknek az egyik, illetve másik csapat győzelmét reprezentáló termékeket választani, mivel ezekre a függetlenség triviálisan teljesül, illetve az egyik legsűrűbben kereskedett termékek.

A fedezés lényegében analóg módon történik a Black-Scholes modellben alkalmazottakkal. Ennek megfelelően hasonló problémák is felmerülhetnek, mint a pénzügyi piacokon a digitális opciókkal. A mérkőzés végéhez közeledve, vagy nagy gólkülönbségek esetén a további gólokra, a fedezésül szolgáló $Z_{\boldsymbol{t}}^1$ és $Z_{\boldsymbol{t}}^2$ kevésbé lesznek érzékenyek, mint a fogadásunk. Ennek okán a ψ_t^i súlyok olyan nagyra nőhetnek, amelyek már lehetetlenné teszik a fedezést, ezáltal instabillá teszik a replikációt. Tekintsünk egy olyan példát, ahol a fogadásunk arra vonatkozik, hogy a gólok száma 3,5 fölött lesz-e a mérkőzés végén. A csapatok gólintenzitásai rendre 1,5 és 1, illetve a hazai csapat 3-0-ra vezet a 60. percben. Ebben az esetben igen kicsi a valószínűsége annak, hogy a hazai csapat vereséget szenvedjen akár esik gól akár nem, tehát a $Z_{\mathbf{60}}^{2}$ közel értéktelen. Azonban a jelenlegi fogadásunk kifizetése egy gólon múlik, akármelyik csapat is szerezze azt.

Modell kalibráció

A modellben használt intenzitásfüggvényt historikus adatok alapján határoztuk meg. Az intenzitás értékét úgy kapjuk majd meg, hogy adott csapat meccsenként átlagosan hány gólt szerez. Ezzel konstans intenzitást adnánk meg. A függvény alakjának becslését általánosnak tekintettük a csapatokra nézve. Lineáris becslést adtunk, majd ennek a becslésnek a konstans értékét határoztuk meg úgy, hogy az visszaadja egy fiktív, teljes mérkőzésre az adott csapatra megfigyelt gólintenzitást.

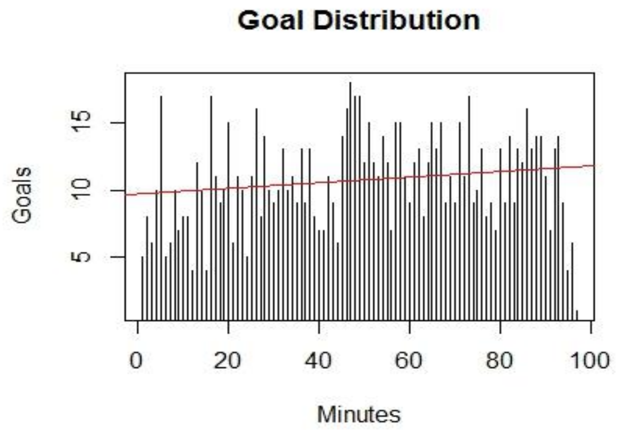
A függvény alakjának kalibrálása:

Ahhoz, hogy csapatonként egyedileg határozzuk meg az intenzitás függvény alakját, viszonylag kevés adat áll rendelkezésünkre (főleg olyan csapatok esetén, amelyek gólintenzitása eleve kisebb). Továbbá feltételezéseink alapján általánosan jellemző a növekvő gólintenzitás, így úgy ítéltük meg, hogy a minden csapatra azonos függvény is jól jellemzi a valóságot.

Mivel adott időpillanatra 0 valószínűséggel esik gól, továbbá kevés is az olyan statisztikai adat, amely másodpercre pontosan mutatná a gólesemények idejét, ezért az adott percre eső gólok számát vettük figyelembe. Ennek előnye továbbá az is, hogy a legtöbb elérhető adatbázisban ez szintén csak percre pontosan van megadva. Azaz minden a 32. percben esett gólt a 32. perchez rendeltünk. A kapott eredmények további kiegészítéseket vetnek fel, amelyeket most csak érintőlegesen tárgyalunk majd.

Az intenzitásfüggvény:

A függvény alakját tehát a Premier League 2013/14 összes bajnoki mérkőzése alapján határoztuk meg. Az eredményt az *1. ábrán* láthatjuk.



1. ábra – Egy teljes Premier League szezon (2013/2014) összes gólja, csoportosítva a szerint, hogy melyik percben estek. Valóban megfigyelhető egy enyhe trend, azonban a gólok sűrűsége a 45. perc környékén a legmagasabb. Az adatokat az InStat statisztikai cégcsoport biztosította (www.instatfootbal.com).

A becsült lineáris függvény, amelyhez természetesen a gólok számát újraskáláztuk, annak megfelelően, hogy egy szezonban összesen 380 mérkőzést játszanak:

```
Coefficients:
     (Intercept)  GoalDist$Minutes
       0.0254883         0.0000562
```

A kapott eloszlás több ponton is eltér a várttól, nem egyértelműen emelkedik. Az egyik jellemző ellentmondás az előzetes feltételezésekkel szemben, hogy a legmagasabb intenzitás nem a mérkőzés végén figyelhető meg, hanem a félidő környékén. Ennek hátterében két ok feltételezhető:

- A játékosok pszichéjére hasonlóan hat a félidő és a mérkőzés vége. Azaz az első félidő végéhez közeledve jobban hajtanak a csapatok, ezzel a hiba és a gólszerzés lehetősége is megnő.

- A valószínűleg erősebb hatást azonban az eredményezi, hogy az első félidő hosszabbításában esett gólokat hogyan könyvelik el. Esetünkben a hosszabbítás 2. percében esett gól 47. percbeli gólként van elkönyvelve, akárcsak egy, a második félidő második percében esett. Tehát ezt a 45. és 50. perc közötti intervallum (vagy egy része), gyakorlatilag nagyobb súlyt kap a historikus becslésben. Érdemes lenne ezt valamilyen formában szabályozni. Azonban a dolgozat keretein belül erre nem dolgoztunk ki lehetséges módszereket.

A másik hasonló effektus, hogy a második félidő hosszabbításában a gólok száma drasztikusan el kezd csökkenni. Ennek igen természetes magyarázata van. Sok mérkőzésen nem is játszanak már a 96-97. percekben.

Adott csapatra a konstans értéket úgy határoztuk meg, hogy az általuk mérkőzésenként elért átlagos gólszámot konstans intenzitásnak vettük. Ennek megfelelően választottunk olyan konstans értéket, hogy a kapott függvény integrálja megegyezzen a konstans függvény integrálértékével.

Az Arsenal csapata például átlagosan *1,816* gólt szerzett, ezt úgy tekintjük, hogy 90 perc alatt. A függvényünk meredeksége, pedig *0,000056*. Ez alapján tehát:

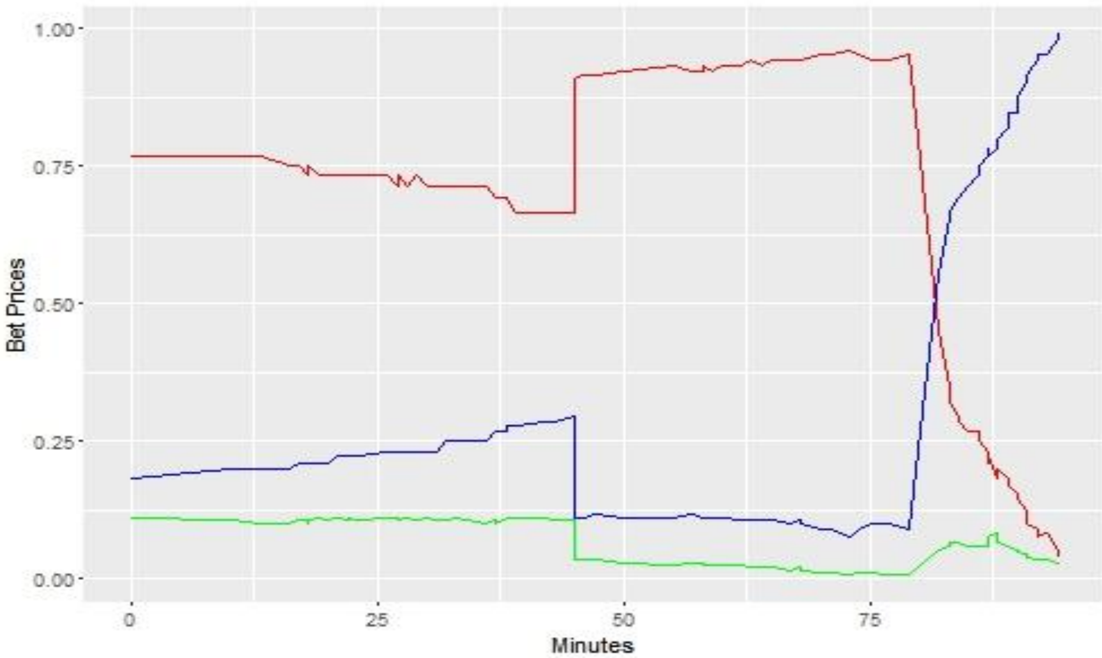
$\mu_A(t) = 0{,}01766 + 0{,}000056 \cdot t$ ahol $\mu_A(t)$ jelöli a csapat gólintenzitását.

Példamérkőzés:

A fenti képletet és kalibrált intenzitásparamétert egy adott példamérkőzésre alkalmaztam. A vizsgált meccs 2016. április 17-én zajlott, az Arsenal otthonában, ahová a Crystal Palace együttese látogatott. A mérkőzés végeredménye 1-1 lett, az Arsenal a 45. a Crystal Palace pedig a 82. percben szerzett gólt. A 4. ábrán látható a három vizsgált fogadástípus ára a mérkőzés folyamán, melyek a hazai, illetve vendég győzelem, valamint a döntetlen. Természetesen a mérkőzés végén a döntetlen ára 1 közeli, a másik két fogadásé pedig 0-ra zuhan.

A két csapat gólintenzitás függvényeit meghatározva az alábbiakat kapjuk:

- $\mu_1(t) = 0{,}01766 + 0{,}000056 \cdot t$ az Arsenal-é (hazai csapat), illetve
- $\mu_2(t) = 0{,}00713 + 0{,}000056 \cdot t$ a Crytal Palace-é (vendég csapat).

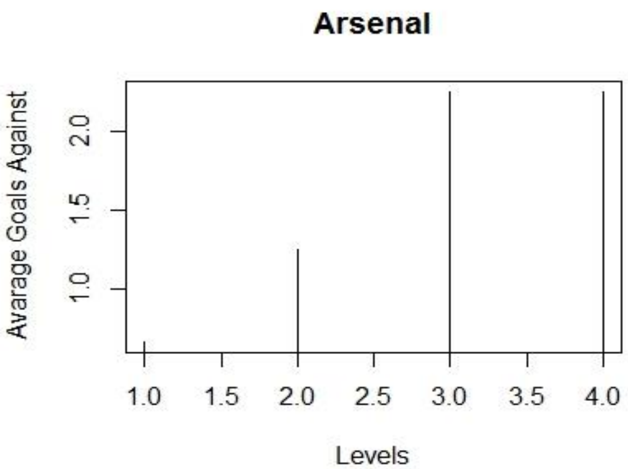


2. ábra – Az Arsenal – Crystal Palace mérkőzés hazai (piros) és vendég (zöld) győzelemre, illetve a döntetlenre (kék) vonatkozó fogadásai. Az oddsokat a bet365.com oldal alapján dolgoztam fel.

Újrakalibrálás

A második lépésben az előbb említett konstans értéken változtathatunk. Adott csapatok párba állítva más csapatokkal a rájuk jellemző játékminőségtől eltérőt produkálhatnak. Egyes esetekben, ha nagyon nagy a fölény az egyik csapat számára a játékerőt tekintve, akkor előfordulhat, hogy az átlagosnál jóval több gólt szereznek, míg ha a két fél nagyjából egy színvonalat képvisel, akkor esetleg mind a ketten a megszokottnál kevesebb gólt fognak rúgni. Ezt lehet megfogni úgy, hogy a mérkőzés kimenetelére (azaz adott csapatok közül ki győz, ki veszít) való becslést felhasználom a gólintenzitás függvény eltolására is. Ugyanis, ha valamely jól használható modell alapján, nagy valószínűséggel az „A" csapat nyer, akkor a valóságot kevésbé jól írja le, ha mégis a „B" csapat gólintenzitását tekintenénk magasabb paraméterűnek.

Ennek bemutatására, megvizsgáltam, hogy egy csapat hogyan teljesít a nála jobb, vagy egyszintű, illetve a nála gyengébben rangsorolt csapatokkal szemben. A rangsorolás az InStat adatai alapján történt, akik saját indexet alkalmaznak a csapatok teljesítményének mérésére és ez alapján 1-7-ig terjedően rangsorolják a csapatokat. Az Arsenal ez alapján a 2. csoportba esik, míg a Premier League 2013/14-es szezonjának minden csapata 1-4-ig besorolásra került. Adott évre megnézve, hogy az Arsenal átlagban hány gólt szerez a különböző besorolású csapatok ellen az eredmény a *3. ábrán* látható.



3. ábra – Az Arsenal átlagos gólszáma különböző szintű csapatok ellen. Az 1. szinthez az olyan csapatok tartoznak, mint például a Chelsea, vagy a Manchester City. A teljes mezőnyt tekintve a leggyengébb csapatok is legalább 4-es besorolásúak. Az Arsenal e szerint a rangsorolás szerint a 2. osztályba tartozik. A rangsorolás alapja az InStat teljesítményindexe.

Pontosabban megvizsgálva, hogy valóban eltér-e a 1. és 2. kategóriába eső csapatok ellen szerzett gólok átlaga a 3. és 4. kategóriába esők ellen szerzettekétől:

A független mintás t-próba eredménye szerint minden szignifikancia szint mellett elvethető az egyezés:

```
data:  Ars.GoalsAgainst$Goals by Ars.GoalsAgainst$Bin.Level
t = -3.5834, df = 26.112, p-value = 0.001365


mean in group 0 mean in group 1
       1.071429        2.250000
```

ahol a *group 0* alá esnek az 1. és 2. kategóriás csapatok.

Az végeredmény becslésére az SVM (Support Vector Machine) módszert alkalmaztuk, mely egy Machine Learning módszer, melyet számos esetben használtak már ilyen jellegű becslésekre. Például A. Tsakonas et al. [2002], illetve N. Vlastakis et al. [2007] foglalkoztak e módszer pontosságának vizsgálatával, a korábbi becslő módszerekhez (például Poisson-számláló regresszió) képest. Ezek a cikkek a mérkőzéseket 2 kimenetűnek választották: {1, -1}, ahol az 1 felel meg a hazai csapat győzelmének, a -1 pedig annak, hogy nem nyer, azaz ezen belül kezeljük a döntetlent és a vendégcsapat győzelmét is. Végül a mérkőzés kimenetelére vonatkozó becslés szerint, a győztesnek jósolt csapat gólintenzitását megemelhetjük, a vesztesét pedig lentebb tolhatjuk, szem előtt tartva azt is, hogy a végső esetben nem szerettük volna, hogy sérüljön az a feltétel, miszerint a győztes csapattól magasabb gólintenzitást várunk el, mint a vesztestől.

Az alapprobléma:

Több esetben foglalkoztak már korábban mérkőzések kimenetelére vonatkozó előrejelzésekkel, nem csak labdarúgás kapcsán. Kosárlabdával kapcsolatban J. B. Yang és C. H. Lu [2012] foglalkoztak Support Vector Machine használatával, a rájátszás mérkőzéseinek kimenetelének becslésére. A. Somboonphokkaphan, S. Phimoltares és C. Lursinsap [2009] pedig historikus adatokra illesztett neurális modellel, ezen belül is

Multilayer Perceptron (MLP) modellel próbálta teniszmérkőzések végeredményét jósolni. Természetesen e módszereket a labdarúgásnál is alkalmazták. A. Tsakonas, G. Dounias és S. Shtovba [2002], illetve N. Vlastakis, G. Dotsis és R. N. Markellos [2007] is SVM segítségével adtak előrejelzéseket meccsek kimenetelére vonatkozó fogadásokkal kapcsolatban. K. Y. Huang és K. J. Chen [2011] pedig MLP modellt illesztettek a 2006-os világbajnokság mérkőzéseinek végeredményeire. Emellett széles körben alkalmaztak még Bayes-i hálókat ilyen jellegű problémákra. N. E. Fenton, M. Niel és A. Joseph [2005], A. C. Constantinou, N. E. Fenton és M. Niel [2012] illetve F. Owramipur, P. Eskandarian és F. S. Mozneb [2013] is ezzel a módszerrel próbálták jósolni adott meccsek kimenetelét.

A fent említett kérdések kezelése szoros kapcsolatban van más valós problémákra vonatkozó döntéshozási technikákkal, illetve előrejelzési módszerekkel, melyek végrehajtásához csak korlátos minőségű és mennyiségű adat áll rendelkezésre. Az ehhez hasonló problémák megoldására alkalmazott „intelligens technikák” közé tartozik a következőkben alkalmazott SVN módszer is. Ez a módszer jól alkalmazható olyan esetekben, amikor kevés adat elérhető és az is bizonytalan, hogy milyen faktorok befolyásolják a végkimenetelt.

A probléma leegyszerűsíthető arra a feladatra, hogy találjunk egy megfeleltetést, mely az alábbiakat kielégíti:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \rightarrow y \in \{d_1, d_2, d_3\},$$

ahol az $\boldsymbol{x}$ jelöli azon faktorok vektorát, melyek hatással lehetnek a mérkőzés kimenetelére (pl.: játékosállomány, hazai/vendégcsapat, aktuális forma, stb.). Továbbá az y jelöli a mérkőzés kimenetelét, jelölés szerint:

- d_1: A hazai csapat nyeri a mérkőzést.
- d_2: Döntetlen.
- d_3: A vendégcsapat nyeri a mérkőzést.

Azonban az SVM alkalmazásához két dimenziós eredményvektorra van szükségünk, így a gyakorlatban az alábbi függvényt fogjuk keresni:

$$\boldsymbol{x} = (x_1, x_2, \dots, x_n) \to y \in \{-1,1\},$$

ahol az $\boldsymbol{x}$ ugyan az, mint korábban, azonban az y értékei az alábbiakat jelentik:

- -1: A hazai csapat **nem** nyeri meg a mérkőzést.
- 1: A hazai csapat nyeri meg a mérkőzést.

Faktorok kiválasztása:

A mérkőzés kimenetelének előrejelzésére használt befolyásoló tényezők kiválasztása szubjektív természetű. Természetesen vannak olyanok, melyek használata általánosnak tekinthető, ilyenek például a „hazai” faktor, vagy a csapatok rangsorolása. Esetünkben természetesen nem szerettünk volna olyan faktorokat a becslésben felhasználni, amelyek közvetlenül kötődnek a gólintenzitáshoz, mivel a mérkőzések kimenetele alapján szeretnénk ezt felülvizsgálni. Így például nem használtuk fel a korábbi mérkőzéseken szerzett gólok a számát, vagy azok különbségét. Az általunk választott faktorok így az alábbiak:

- $x_1 =$ Hazai/ Vendég faktor (mégpedig *HP/HM – VP/VM* alakban, ahol a *HP* a hazai csapat otthoni környezetben szerzett pontjainak a száma, a *HM* pedig az otthon játszott meccseinek a száma az idényben. Ugyan így a *VP* vendégcsapat, vendégként szerzett pontjainak a száma, a *VM* pedig azt jelöli, hogy ezt hány mérkőzésen érték el.)
- $x_2 =$ InStat index alapján való rangsorértékek különbsége *HI-VI*.
- $x_3 =$ Forma mutató – Az elmúlt 5 mérkőzésen szerzett pontok különbsége *HP5-VP5* alakban, ahol a *HP5* jelöli a hazaiak a *VP5* pedig a vendégek által szerzett pontokat.
- $x_4 =$ Nemzetközi, illetve hazai kupa szereplés. Itt 0 értéket kap egy csapat, ha nem szerepel aktuálisan kupában (így csak bajnokságban érdekelt), 1-et kap, ha az FA Kupábanban szerepel, 2-t, ha az UEFA Európa-ligában, illetve 3-at, ha az UEFA Bajnokok Ligájában. Amennyiben egy csapat az FA Kupában és valamelyik nemzetközi kupasorozatban is részt vesz, azok összegét kapja értékül. Ezt követően pedig a *HL-VL* értéket vesszük figyelembe.

Ezen faktorok az InStat indexet leszámítva könnyen hozzáférhetőek a mérkőzéseket megelőzően, de ennek helyettesítésére is számos elérhető alternatíva kínálkozik. Természetesen sok más faktort is alkalmazhatnánk, például a mérkőzés kezdő játékosainak InStat indexe, illetve amennyiben nem a gólintenzitás becslésére, hanem csak előrejelzésre alkalmazzuk, akkor bármelyik szerzett gólokkal kapcsolatos statisztika. Azonban a jelen dolgozatban ezek hatásával nem foglalkozunk.

Support Vector Machine módszer:

Az SVM egy machine learning típusú módszer, melynek lényege, hogy bizonyos visszatérő minták felismerésére alkalmas. Emellett regressziós becslésekre és lineáris operátorok inverzének meghatározására is használják. Nagy előnye más hasonló módszerekkel szemben, hogy jól értelmezhető geometriai interpretációja is létezik, illetve, hogy minden esetben képes globális minimumot találni. Emellett képes nagy mennyiségű adat feldolgozására is, úgy, hogy sebessége nem marad el a neurális hálókhoz képest sem. Az alap probléma, aminek a megoldását keressük, az alábbi f függvény meghatározása:

$$f : \boldsymbol{x} \rightarrow \{\pm 1\}.$$

Itt lényeges, hogy megfelelő mennyiségű faktoron és méretű adathalmazon keresztül próbáljuk meg megadni ezt a függvényt. A szakirodalom erre a két tulajdonságra úgy hivatkozik, mint a modell *kapacitása*. Ha a kapacitás kicsi, akkor összetettebb függvények meghatározása kevésbé hatékony, míg nagy kapacitás esetén overfitting léphet fel, melynek következtében ne kapunk jó általános megoldást. A neurális hálók esetén az overfitting elkerülésére „korai megállás” módszert szoktak alkalmazni, azonban az SVM overfitting-je korlátos a kis mintákra vonatkozó statisztikai tételek alapján (Tsakonas et. al. [2002]).

Lineáris függvények meghatározása esetén az SVM alapja egy felület keresése, amely alapján kettéválasztjuk a faktorok értékeit a kimenetelek alapján. Az elválasztó felület és a csoportokhoz tartozó pontok távolságát maximalizálja az implementáció. Ennek a *szegélynek* a maximalizálása egy konvex kvadratikus probléma. Kiegészítésképp a módszerbe egy további paraméter is be van építve, amely lehetővé teszi, hogy egyes kilógó pontokat rosszul soroljunk be, annak érdekében, hogy általános esetben jobb eredményt kapjunk.

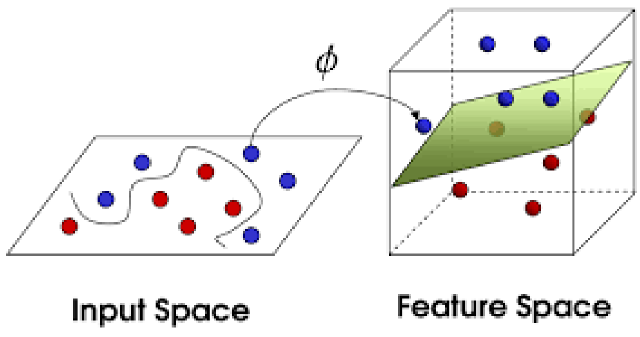
Az input adatokat ezután az alábbi módon, egy a jellemzők által meghatározott térre transzformálva:

$$\varphi: \boldsymbol{x} \rightarrow F,$$

az eredetileg lineáris problémát nem-lineárissá alakíthatjuk. Ezt implicit módon hajtjuk végre, nekünk ugyanis a jellemzők terében kell meghatároznunk két együttes mérkőzésének pozícióját, annak érdekében, hogy eldönthető legyen, hogy az adott mérkőzés milyen kimenetelű lesz. Ezt az alábbi skalárszorzat fogja megadni: $\varphi(\boldsymbol{x}_1) \cdots \varphi(\boldsymbol{x}_n) \in F$. Erre a következő jelölést vezetjük be:

$$K(\boldsymbol{x}_1, .., \boldsymbol{x}_n) = \varphi(\boldsymbol{x}_1) \cdots \varphi(\boldsymbol{x}_n),$$

ahol a K-ra a továbbiakban *kernel* függvényként fogunk hivatkozni. Ebben az esetben nem kell ismernünk a φ függvényünket, azt impliciten használjuk. Ez a kernel függvény megadja azt is, hogy az adott jellemzők közül melyek jelentenek tényleges hatást az elválasztásban. Ennek a megfelelő megválasztása nagyban elősegítheti a módszer hatékonyságát.



3. ábra – Az SVM algoritmusa szerint a bemeneti adatainkat a $K(\boldsymbol{x}_1, .., \boldsymbol{x}_n)$ kernel függvény segítségével úgy képezzük le, hogy a leképzést követően az adatok jól szeparálhatóak legyenek. A szeparációban természetesen megengedett a marginális hiba, amit jelen esetben a zöld sáv reprezentál. Ez után az egyenesünket visszaképezve az eredeti térbe, megkapjuk a szeparáló felületünket.

SVM algoritmus:

Adott $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, ahol minden $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ $\mathbf{R}^n$-beli vektor, mely egy adott mérkőzés paramétereit tartalmazza, illetve $y_i \in \{-1,1\}$ pedig az adott meccs kimenetele. Ez az S lesz az a tanuló adathalmaz, amely alapján a célunk meghatározni a következőt:

$$f(\mathbf{x}) = \mathbf{w} \cdot \varphi(\mathbf{x}) + b,$$

ahol $\mathbf{w} = (w_1, w_2, \dots, w_n)$ and b az elválasztó hipersík paraméterei, a φ pedig a korábban implicite definiált $\mathbf{R}^n$-ből $\mathbf{R}^m$-be való leképzés.

A módszer formalizálása az alábbi kvadratikus probléma megoldását adja (Tsakonas [2003]):

- **Alap** probléma:

  Minimalizálni: $\frac{\mathbf{w} \cdot \mathbf{w}}{2} + C \cdot \sum_{i=1}^{N} \xi_i,$

  feltéve, hogy: $y_i(\mathbf{w} \cdot \varphi(\mathbf{x}) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1,2, \dots, N,$

ahol a C a pozitív büntetésfüggvény a félreosztályozás esetére.

Ennek a problémának a $\mathbf{w}^*$ megoldását pedig az alábbi egyenlet adja:

$$\mathbf{w}^* = \sum_{i=1}^{N} a_i^* \, y_i \varphi(\mathbf{x}_i),$$

ahol $\mathbf{a}^* = (a_1, a_2, \dots, a_N)$ a megoldása a következő *duális problémának.*

- **Duális** probléma:

  Minimalizálni: $-\frac{1}{2}\mathbf{a}^T \mathbf{D} \mathbf{a} + \sum_{i=1}^{N} a_i,$

  feltéve, hogy: $\sum_{i=1}^{N} a_i y_i = 0,\ 0 \leq a_i \leq C, i = 1,2, \dots, N,$

ahol $\mathbf{D}$ egy $N \times N$-es mátrix, és az alábbi alapján áll elő:

$$D_{ij} = y_i y_j \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j).$$

Felhasználva az alapprobléma $\mathbf{w}^*$-re vonatkozó egyenletét és a $f(\mathbf{x}) = \mathbf{w} \cdot \varphi(\mathbf{x}) + b$-t, kapjuk az **Alap** probléma megoldását:

$$\sum_{i=1}^{N} a_i^* \, y_i \varphi(\mathbf{x}_i) + b^*.$$

Ebben az esetben azok a pontok, amelyre $a_i^* \geq 0$, vagy rossz besorolást kaptak, vagy a minimális távolságnál közelebb helyezkednek el a hipersíkhoz. Ezeket nevezzük Support Vectoroknak, SV-knek. Ezek többnyire a tanuló adathalmaz kis részét adják.

Megjegyzés: Belátható, hogy a **Duális** probléma komplexitása N^2-es, ezáltal független attól, hogy hány dimenziós teret alkotnak a faktorok. Ezáltal a módszer akár végtelen faktorra is kiterjeszthető, azonban a gyakorlatban a számítási igény miatt mutatkoznak korlátok a probléma megoldhatóságára, a tanuló adathalmaz méretének növekedésével.

SVM implementálása:

A felhasznált tanuló adatcsoport a Premier League 2013-14-es szezonjának első 34 fordulója, a teszthalmaz pedig a maradék 4 forduló volt. Ezen felül a teszthalmazba beillesztettem még a vizsgált idei *Arsenal – Crystal Palace* mérkőzést is, erre is megjósolva a meccs előtt a kimenetelt.

A modell illesztéséhez az R „kernlab" programcsomagját alkalmaztam. Ez a program csomag tartalmaz egy *ksvm* névre hallgató funkciót, melynek segítségével a tanuló adatokra és kimenetelekre könnyen illeszthetjük az SVM modellünket. A tanuló adattáblánk tartalmazta a mérkőzésekre vonatkozó {1;-1} kimeneteleket, valamint az $\mathbf{x}_i$, $i = 1,..,4$ változóinkat, melyek alapján a kimenetelt szeretnénk meghatározni. A konkrétan alkalmazott módszer *nu-regresszió* volt, melynek *nu* paramétere határozza meg a felső hibahatárt a tanulóadatokra vonatkoztatva. A kernel függvénynek pedig a Laplace-kernelt választottam. A tanuló adatokra ez adta a legjobb eredményeket a felhasznált kernel-függvények közül.

Az illesztett modell aztán alkalmaztam a tanulóadatokra. A *predict* függvény segítségével meghatározható az, hogy a modell alapján a bemenő paraméterekkel mekkora az esélye annak, hogy a mérkőzés kimenetele hazai győzelem legyen. Ebből a kimenetelt természetesen úgy adtam meg, hogy minden esetben ahol a kapott valószínűség $0{,}5$ fölé esett (szigorúan), abban az esetben a várható kimenetel a hazai győzelem. Lehetséges finomítása lehetne a modellnek, ha a valós kimeneteleknek megfelelően három valószínűségi intervallumot tekintenénk, és a $0{,}5$ környezetébe eső eredményeket döntetlennek tekintenénk, azonban ezzel részletesebben ebben a dolgozatban nem foglalkoztam.

Az eredmények a tanulóhalmazra kitűnőek, a mérkőzések $98{,}5\%$-nál is többet talált el a módszer. A tesztadatokra alkalmazva is jónak tűnnek a becslések, a mérkőzések $62{,}5\%$ találta el a módszer. Emellett a példamérkőzésünkre (*Arsenal – Crystal Palace*) is azt az eredményt jósolta, hogy a hazai csapat nem fog nyerni

$$P(y_{AvsCP} = 1) = 37{,}2\%,$$

ami helytálló. Ez érdekes abból a szempontból, hogy ha csak például az InStat rangsorolásukat figyelnénk, akkor az Arsenal-tól hazai pályán elvárnánk, hogy nyerjen. Ez alapján, az eredmény alapján, mégis azt javasolnám erre az adott mérkőzésre, hogy közelítsük a gólintenzitásokat egymáshoz.

Megj.: *Természetesen érdemes lenne itt is figyelembe vennünk a döntetlen és vendég győzelemmel végződő mérkőzések különválasztását, ha a valósághoz jobban közelítő modellt szeretnénk alkalmazni. Azonban ez az apró átalakítás is sok olyan megfontolást igényel, amelyekre e dolgozaton belül nem térnék ki.*

Az újrakalibrálásra a következő módszert alkalmazom majd. Első sorban a hazai csapat gólintenzitását módosítanám, még pedig ha a SVM kimenetele szerint a hazai csapat győzelme várható, de a gólintenzitásbeli különbség nem haladja meg az 1 gólt, akkor megemelem a teljes mérkőzésre vonatkozó gólintenzitását $0{,}5$ góllal. Ellenben, ha az előrejelzés szerint nem nyer a hazai csapat, de a gólintenzitásbeli különbség 1-nél nagyobb, akkor csökkenteném ugyan ezzel az értékkel, ha ez megtehető úgy, hogy a gólintenzitás sehol se legyen negatív (ilyenre azonban a jelenleg felhasznált adatok között nem találtam példát). Formalizálva:

- Ha a becsült $y_i = 1$ és $\int_0^{90} \mu_1(t) - \mu_2(t)\, dt < 1$ akkor $\mu_1^*(t) = \mu_1\ (t) + \frac{0,5}{T}$, ahol $\mu_1^*(t)$ jelöli az újrakalibrált gólintenzitást, a T pedig a mérkőzés teljes hosszát percekben (ált. 90)
- Ha $y_i = -1$, de $\int_0^{90} \mu_1(t) - \mu_2(t)\, dt > 1$, akkor pedig $\mu_1^*(t) = \mu_1\ (t) + \frac{0,5}{T}$.

A példamérkőzésünkre tehát $y_i = -1$, ez alapján pedig:

- $\mu_1^*(t) = 0{,}01366 + 0{,}000056 \cdot t$ az Arsenal (hazai csapat) becsült gólintenzitása, illetve
- $\mu_2^* = \mu_2(t) = 0{,}00713 + 0{,}000056 \cdot t$ a Crytal Palace-é (vendég csapat).

Implicit és historikus intenzitás:

A korábbi kutatásokban megfigyelt implicit intenzitás alapján javasolt exponenciális növekedés a ***P*** mérték alatt nem tűnik jó elképzelésnek. Ez jóval kevésbé meredek, ebben valóban lineáris jellegű trend figyelhető meg, míg implicit esetben inkább exponenciális növekedést feltételeztek Divos et. al. [2015]. Ennek két kézen fekvő oka lehet. Az implicit intenzitás ugyanis, a piacon megfigyelhető fogadási árakat vette alapul. Ebbe azonban valószínűleg belejátszik az is, hogy mérkőzés vége felé a fogadások intenzitása nő, ami megemeli az odds-okat. Ez a hatás pedig a gólintenzitásban, mint paraméterben fog megjelenni. Erre tehát piaci faktorok is hatnak, nem csak maguk a mérkőzés eseményei, továbbá mivel a piacon megfigyelhető árakhoz igazodnak, valószínűleg jobban el is találják azokat. A másik pedig természetesen a mértékcsere hatása. A valós mérték alatti gólintenzitást, mint megfigyel tulajdonságot, csak a mérték csere után hasonlíthatnánk össze az implicit intenzitás értékével.

Lehetséges általánosítások

A legkézenfekvőbb továbbvitele a dolgozat által tárgyalt témának természetesen azon Q mérték meghatározása, amelyre áttérve már árazhatnánk a fogadásainkat. Általánosan, ha tekintünk egy N $\{\mathcal{F}_t\}$-mérhető $(\Omega, \mathcal{F}, P)$-n Poisson-folyamat független növekményekkel, illetve egy Z nemnegatív $\{\mathcal{F}_t\}$-mérető folyamatot, akkor:

$$L(t) = \exp\{\int_0^t \ln Z(s-)\, dN(s) - \int_0^t Z(s) - 1 ds\}$$

kielégíti a

$$L(t) = 1 + \int_0^t (Z(s-) - 1) L(s-) d(N(s) - s)$$

egyenletet, és P-lokális martingál. Ha $E[L(t)] = 1$ és definiáljuk a $dQ = L(T)dP$ tulajdonságú Q mértéket $\mathcal{F}_t$-n, akkor $N(t) - \int_0^t Z(s)ds$ Q-lokális martingál lesz.

Érdemes lehet továbbá megfontolni azt, hogy a mérkőzésen történt események is módosíthatják a gólintenzitás értékét. Például egy csere is a csapat frissítésével és taktikai módosításokkal megnövelheti, illetve az edző elvárása szerint meg is kell, hogy növelje a csapat gólintenzitását. Ezt lehetséges lenne tovább finomítani azzal, hogy mikor milyen erősségű, vagy milyen formában lévő játékosok cserélnek egymással helyet a pályán. Rengeteg oldal foglalkozik manapság a játékosok meccsenkénti értékelésével és általánosan a képességeik megítélésével. Például az InStat.com korábban említett indexe is létezik nem csak csapatok, de játékosok értékelésére is. Egy nagyon jó képességű gólerős csatár sérülés következtében való lecserélése csökkentheti a csapata gólintenzitását, míg ha a védelem egyik biztos pontját kell lehozni a pályáról hasonló okok miatt, akkor az az ellenfél gólintenzitásának növekedésével járhat. Ugyan ezt akár a gólintenzitás becslésének azon részébe is beépíthetjük ahol a mérkőzés kimenetelére adunk előzetes becslést (esetünkben ez az SVM történő előrejelzés volt). A meghirdetett kezdő 11 játékos értékelése is bevehető a mérkőzés kimenetelét meghatározó fatorok közé.

Természetesen a legerősebb vonzattal a gólok járnak. Míg az oddsokat természetesen látványosan és azonnali jelleggel módosítják, ezzel beépülve az árba, de pszichológiai hatásuk a csapatok teljesítményére is lényeges. Hogy lenne érdemes ezek hatását modellezni? A feltételezés, amit alkalmazhatunk az, hogy minden gólesemény, ami nem növeli a két csapat közötti különbséget 2-nél nagyobbra, az az elszenvedő fél gólintenzitását növeli meg nagyban. Míg ha a gólkülönbség háromra emelkedik, akkor mind a két csapat nagy valószínűséggel már eldőltnek nyilvánítja magában a mérkőzést, így mindkettejük paramétere csökken. E szerint egy sztochasztikus ugrófolyamatot kapunk a gólintenzitás paraméterének értékére, melynek ugráseloszlását önmaga hajtja meg.

További lehetőség lehet még az, hogy a pénzügyi piacon jól árazható termékeknek megpróbáljuk a megfelelőit megalkotni a sportfogadás keretein belül is. Rengeteg derivatíva (egzotikus opciók, swapok stb.) árazására jól bevett módszereket alkalmazunk a pénzügyi piacokon. Miért ne lehetne akár ezeknek egy alternatíváját összerakni sportfogadások keretein belül is, amennyiben arra van megfelelő alaptermékünk, fedezésünk? A pénzügyi piacokon is rengeteg olyan termék kerül bevezetésre, amelyet nem feltétlenül előre ismert piaci igények hívnak életre, de bevezetésük után a kereslet növekedése velük szemben hatékonnyá teszi piacukat, kereskedésük általánossá válik. Mind az odds, mind a spread típusú fogadások esetén rengeteg lehetőség adódik arra, hogy létrehozzunk ilyeneket a pénzügyi piacokon széles körben kereskedett termékek alapján. A nevéből adódóan is rengeteg spread jellegű pénzpiaci termék átöltethető a sportfogadások keretein belülre.

Lehetséges általánosítások

A legkézenfekvőbb továbbvitele a dolgozat által tárgyalt témának természetesen azon Q mérték meghatározása, amelyre áttérve már árazhatnánk a fogadásainkat. Általánosan, ha tekintünk egy N $\{\mathcal{F}_t\}$-mérhető $(\Omega, \mathcal{F}, P)$-n Poisson-folyamat független növekményekkel, akkor ha tekintünk egy Z nemnegatív $\{\mathcal{F}_t\}$-mérető folyamatot, akkor:

$$L(t) = \exp\{\int_0^t \ln Z(s-)\, dN(s) - \int_0^t Z(s) - 1 ds\}$$

kielégíti a

$$L(t) = 1 + \int_0^t (Z(s-) - 1) L(s-) d(N(s) - s)$$

egyenletet, és P-lokális martingál. Ha $E[L(t)] = 1$ és definiáljuk a $dQ = L(T)dP$ tulajdonságú Q mértéket $\mathcal{F}_t$-n, akkor $N(t) - \int_0^t Z(s)ds$ Q-lokális martingál lesz (Kurtz [2001]).

Érdemes lehet továbbá megfontolni azt, hogy más, a mérkőzésen történt események is módosíthatják a gólintenzitás értékét. Például egy csere is a csapat frissítésével és taktikai módosításokkal megnövelheti, illetve az edző elvárása szerint meg is kell, hogy növelje a csapat gólintenzitását. Ezt lehetséges lenne tovább finomítani azzal, hogy mikor milyen erősségű, vagy milyen formában lévő játékosok cserélnek egymással helyet a pályán. Rengeteg oldal foglalkozik manapság a játékosok meccsenkénti értékelésével és általánosan a képességeik megítélésével. Például az InStat.com korábban említett indexe is létezik nem csak csapatok, de játékosok értékelésére is. Egy nagyon jó képességű gólerős csatár sérülés következtében való lecserélése csökkentheti a csapata gólintenzitását, míg ha a védelem egyik biztos pontját kell lehozni a pályáról hasonló okok miatt, akkor az az ellenfél gólintenzitásának növekedésével járhat. Ugyan ezt akár a gólintenzitás becslésének azon részébe is beépíthetjük ahol a mérkőzés kimenetelére adunk előzetes becslést (esetünkben ez az SVM történő előrejelzés volt). A meghirdetett kezdő 11 játékos értékelése is bevehető a mérkőzés kimenetelét meghatározó fatorok közé. Modellezhetnénk ezt olyan függvények megadásával, melyek az ilyen események hatását közvetlenül a gólintenzitás változására képeznék le.

Természetesen a legerősebb vonzattal a gólok járnak. Míg az oddsokat látványosan és azonnali jelleggel módosítják, ezek beépülnek az árba, de pszichológiai hatásuk a csapatok teljesítményére is lényeges. Hogy lenne érdemes ezek hatását modellezni? A feltételezés, amit alkalmazhatunk az, hogy minden gólesemény, ami nem növeli a két csapat közötti különbséget 2-nél nagyobbra, az az elszenvedő fél gólintenzitását növeli meg nagyban. Míg ha a gólkülönbség háromra emelkedik, akkor mind a két csapat nagy valószínűséggel már eldőltnek nyilvánítja magában a mérkőzést, így mindkettejük paramétere csökken. E szerint egy sztochasztikus ugrófolyamatot kapunk a gólintenzitás paraméterének értékére, melynek ugráseloszlását önmaga hajtja meg.

További lehetőség lehet még az, hogy a pénzügyi piacon jól árazható termékeknek megpróbáljuk a megfelelőit megalkotni a sportfogadás keretein belül is. Rengeteg derivatíva (egzotikus opciók, swapok stb.) árazására jól bevett módszereket alkalmazunk a pénzügyi piacokon. Miért ne lehetne akár ezeknek egy alternatíváját összerakni sportfogadások keretein belül is, amennyiben arra van megfelelő alaptermékünk, fedezésünk? A pénzügyi piacokon is rengeteg olyan termék kerül bevezetésre, amelyet nem feltétlenül előre ismert piaci igények hívnak életre, de bevezetésük után a kereslet növekedése velük szemben hatékonnyá teszi piacukat, kereskedésük általánossá válik. Mind az odds, mind a spread típusú fogadások esetén rengeteg lehetőség adódik arra, hogy létrehozzunk ilyeneket a pénzügyi piacokon széles körben kereskedett termékek alapján. A nevéből adódóan is, rengeteg spread jellegű pénzpiaci termék átültethető a sportfogadások keretein belülre.

Összefoglalás

A dolgozat során in-game sportfogadásokat vizsgáltam, melyek népszerűsége az elmúlt időszakban igen jelentősen nőtt. Továbbá e fogadások kereskedésére is lehetőségünk nyílik, aminek folyamán nő az igény a fogadások árazásának javítására. Első lépésben bemutattam a legnépszerűbb ilyen termékeket. Ezt követően vizsgáltam azt a matematikai keretrendszert melyen belül létezik az ilyen típusú fogadások fair árazása. Az árazás alapjául azok a Poisson-folyamatok szolgálnak, melyek a mérkőzés alatt eső gólok eloszlását jellemzik. A dolgozat további részében ezeknek a folyamatok intenzitásának a becslésére mutattam be lehetséges módszereket. Historikus adatok alapján feltételezhető, hogy ez az intenzitás nem konstans a mérkőzés során. A függvény alakjának becslésére egyszerű lineáris függvényt alkalmaztam, melyről feltételeztem, hogy a mérkőzés előrehaladtával monoton nő. A másik feltevés, amelyet vizsgáltam az volt, hogy bár az intenzitás függvény alakját csapatok között általánosnak tekintettem, azonban természetesen a szintje csapatonként változik, sőt továbbmenve attól is függ, hogy a csapat egy adott mérkőzésen kivel áll szemben. Ennek kapcsán a mérkőzések kimenetelének becslésének segítségével a két csapat gólintenzitás függvényét úgy módosítottam, hogy az jobban illeszkedjen ehhez a jósolt kimenetelhez. A mérkőzések kimenetelének előrejelzéséhez a Support Vector Machine módszert alkalmaztam, amelyet a dolgozat során részletesen bemutattam, majd historikus adatokon kalibráltam, és egy példamérkőzésre alkalmaztam is. Az e folyamat során becsült intenzitások önmagukban még nem elegendőek a fogadások beárazásához. A matematikai keretrendszer jellemzése során felmerül, hogy az ezen intenzitásokkal meghajtott folyamatokat egy új mérték szerint kell vizsgálnunk, és ezen új mértékre való áttérés megváltoztatja az intenzitásfüggvény paramétereit, és adott esetben alakját. Azonban ebben a dolgozatban nem foglalkoztam e mérték meghatározásával és a mértékcsere alkalmazásával. Végül néhány lehetséges továbbfejlesztését soroltam fel a modellnek, melyeket a labdarúgó mérkőzések dinamikájáról való elképzeléseim alapján fogalmaztam meg.

Hivatkozások

BACK, Kerry; PLISKA, Stanley R. On the fundamental theorem of asset pricing with an infinite state space. *Journal of Mathematical Economics*, 1991, 20.1: 1-18.

BRÉMAUD, Pierre. Point processes and queues. 1981.

CONSTANTINOU, Anthony C.; FENTON, Norman E.; NEIL, Martin. pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 2012, 36: 322-339.

COX, John C.; ROSS, Stephen A. The valuation of options for alternative stochastic processes. *Journal of financial economics*, 1976, 3.1-2: 145-166.

DIVOS, Peter, et al. Risk-Neutral Pricing and Hedging of In-Play Football Bets. *Available at SSRN 2598767*, 2015.

DIXON, Mark; ROBINSON, Michael. A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1998, 47.3: 523-538.

DUFFIE, Darrell. *Security markets: Stochastic models*. New York: Academic Press, 1988.

FELLER, Willy. On the integro-differential equations of purely discontinuous Markoff processes. *Transactions of the American Mathematical Society*, 1940, 48.3: 488-515.

FITT, A. D.; HOWLS, C. J.; KABELKA, M. Valuation of soccer spread bets. *Journal of the Operational Research Society*, 2006, 57.8: 975-985.

HARRISON, J. Michael; KREPS, David M. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 1979, 20.3: 381-408.

HARRISON, J. Michael; PLISKA, Stanley R. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 1981, 11.3: 215-260.

HARRISON, J. Michael; PLISKA, Stanley R. A stochastic calculus model of continuous trading: complete markets. *Stochastic processes and their applications*, 1983, 15.3: 313-316.

HUANG, Chi-Fu. Information structure and equilibrium asset prices. *Journal of Economic Theory*, 1985, 35.1: 33-71.

HUANG, Kou-Yuan; CHEN, Kai-Ju. Multilayer perceptron for prediction of 2006 world cup football game. *Advances in Artificial Neural Systems*, 2011, 2011: 11.

JOSEPH, A.; FENTON, Norman E.; NEIL, Martin. Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 2006, 19.7: 544-553.

KURTZ, Thomas G. Lectures on stochastic analysis. *Department of Mathematics and Statistics, University of Wisconsin, Madison, WI*, 2001, 53706-1388.

MAHER, Michael J. Modelling association football scores. *Statistica Neerlandica*, 1982, 36.3: 109-118.

MEYER, David; WIEN, FH Technikum. Support vector machines. *The Interface to libsvm in package e1071*, 2015.

MORONEY, Michael Joseph. *Facts from figures*. 1956.

OWRAMIPUR, Farzin; ESKANDARIAN, Parinaz; MOZNEB, Faezeh Sadat. Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team. *International Journal of Computer Theory and Engineering*, 2013, 5.5: 812.

REEP, C.; POLLARD, R.; BENJAMIN, B. Skill and chance in ball games. *Journal of the Royal Statistical Society. Series A (General)*, 1971, 623-629.

SOMBOONPHOKKAPHAN, Amornchai; PHIMOLTARES, Suphakant; LURSINSAP, Chidchanok. Tennis winner prediction based on time-series history with neural modeling. In: *Proceedings of the International Multi Conference of Engineers and Computer Scientists*. 2009.

TANKOV, Peter. *Financial modelling with jump processes*. CRC press, 2004.

TSAKONAS, A., et al. Soft computing-based result prediction of football games. In: *The First International Conference on Inductive Modelling (ICIM'2002). Lviv, Ukraine*. p 181-186. 2002.

YANG, Jackie B.; LU, Ching-Heng. Predicting NBA championship by learning from history data. *Proceedings of Artificial Intelligence and Machine Learning for Engineering Design*, 2012.

VLASTAKIS, Nikolaos; DOTSIS, George; MARKELLOS, Raphael N. Nonlinear modelling of European football scores using support vector machines. *Applied Economics*, 2008, 40.1: 111-118.

http://betfair.com/ 2015.10.12.

http://bwin.com/ 2015.10.12.

http://instatscout.com 2016.04.01.

http://www.soccerstats.com/latest.asp?league=england_2014 , 2016.05.14.

http://www.bet365.com/ 2016.04.17.

Appendix

A1.

A következő táblázat foglalja össze a legnépszerűbb európai típusú sportfogadási termékek kifizetésfüggvényeit, illetve az azokhoz tartozó fogadások értékeit:

Fogadás típusa	**Kifizetésfüggvény:** $\prod(\boldsymbol{N_T^1}, \boldsymbol{N_T^2})$	**Fogadás értéke:** $\boldsymbol{X_t(N_t^1 N_t^2 \lambda_1 \lambda_2)}$
Hazai csapat nyer	$\chi(N_T^1 > N_T^2)$	$\sum_{k_1>k_2} \prod_{i=1,2} P(k_i - N_t^i, \Lambda_i)$
Vendég csapat nyer	$\chi(N_T^1 < N_T^2)$	$\sum_{k_1<k_2} \prod_{i=1,2} P(k_i - N_t^i, \Lambda_i)$
Döntetlen	$\chi(N_T^1 = N_T^2)$	$\sum_{k_1=k_2} \prod_{i=1,2} P(k_i - N_t^i, \Lambda_i)$
Pontos eredmény	$\chi(N_T^1 = K_1, N_T^2 = K_2)$	$\prod_{i=1,2} P(K_i - N_t^i, \Lambda_i)$
K-nál több gól	$\chi(N_T^1 + N_T^2 > K)$	$\sum_{k=K+1}^{\infty} P(k - N_t^1 - N_t^2, \Lambda_1 + \Lambda_2)$
K-nál kevesebb gól	$\chi(N_T^1 + N_T^2 < K)$	$\sum_{k=0}^{K-1} P(k - N_t^1 - N_t^2, \Lambda_1 + \Lambda_2)$

Ahol $\prod(N_T^1, N_T^2)$ kifizetésfüggvény a fogadás mérkőzés végi értékét adja meg. A $P(k, \Lambda)$ adja meg a Poisson eloszlást, $P(k, \Lambda) = \frac{1}{k!} e^{\Lambda} \Lambda^k$. Ezen belül pedig a $\Lambda_i(t) = \frac{1}{T-t} \int_t^T \lambda_i(s) ds$ minden $i \in \{1,2\}$.

A nem európai típusú fogadások köréből pedig az alábbiakat tekintjük:

- **Következő gól**: A korábbi definíció alapján tehát annak a fogadásnak az értéke, amely szerint a következő gólt a hazai csapat szerzi:

  $X_t = \frac{\Lambda_1(t)}{\Lambda_1(t)+\Lambda_2(t)} [1 - e^{-(\Lambda_1(t)+\Lambda_2(t))(T-t)}],$

illetve hasonlóan, a vendég csapatra:

$$X_t = \frac{\Lambda_2(t)}{\Lambda_1(t)+\Lambda_2(t)}\left[1 - e^{-(\Lambda_1(t)+\Lambda_2(t))(T-t)}\right].$$

- **Félidő/Vége:** A „félidő döntetlen, végeredmény pedig vendég győzelem” fogadás értéke:

  $$X_t = \sum_{k_1=k_2}\sum_{l_1<l_2} P(k_1 - N_t^1, \Lambda_1(t)\cdot(T_{\frac{1}{2}} - t))P(k_2 - N_t^2, \Lambda_2(t)\cdot(T_{\frac{1}{2}} - t)) \times P(l_1 - N_t^1, \Lambda_1(t)\cdot(T - T_{\frac{1}{2}}))P(l_2 - N_t^2, \Lambda_2(t)\cdot(T - T_{\frac{1}{2}})).$$

A második félidőre ez a fogadás vagy elértéktelenedik (amennyiben a félidei eredmény nem döntetlen), vagy pontosan meg fog felelni a **Vendégcsapat nyer** fogadásnak.

A2.

A dolgozat készítése során, a szükséges függvények alkalmazásához, számításokhoz, ábrázoláshoz, valamint az SVM-modell illesztéséhez is az R nyelvet használtam fel, amely ingyenessége, széleskörű alkalmazása és programcsomagjai miatt ideális választásnak bizonyult. A felhasznált lényegesebb kódrészletek az alábbiak:

```
##Felhasznált csomagok

install.packages("plyr")

install.packages("reshape")

install.packages("ggplot2")

install.packages("kernlab")


library(ggplot2)

library(grid)

library(reshape)

library(plyr)

library(kernlab)


##Megigyelt gólintenzitás kiszámolása és ábrázolása
```

```
GoalDist<-read.table("Goaldistribution.txt")

GoalDist<-rename(GoalDist, c("V1"="Minutes", "V2"="GoalCount"))

GoalDist$GoalPerMatch<-GoalDist$GoalCount/380


plot(main= "Goal Distribution", x=GoalDist$Minutes, y=GoalDist$GoalPerMatch, type="h",
xlab="Minutes", ylab="GoalsIntensity")


LinearIntensity<-lm(GoalDist$GoalPerMatch~GoalDist$Minutes)

abline(LinearIntensity, col=2)

mean(GoalDist$GoalPerMatch)


##Az Arsenal teljesítménye a különböző szintű csapatok ellen.

Performance<-read.table("PerformanceAgainst.txt")

Performance<-rename(Performance, c("V1"="Levels", "V2"="GoalCount"))

Performance<-data.frame(Performance)

Performance<-aggregate(Performance, by=list(Performance$Levels), mean)

round(Performance$GoalCount,2)

plot(Performance, type="h", xlab="Levels", ylab="Avarage Goals Against", main="Arsenal")


##Az Arsenal - Crystal Palace mérkőzés oddsainak reciprokának alakulása a mérkőz során

ArsVsCrystal<-read.table("ArsenalvsCrytalPalace20160417.txt")

str(ArsVsCrystal)

ArsVsCrystal<-rename(ArsVsCrystal, c("V1"="Minutes", "V2"="Arsenal", "V3"="Draw",
"V4"="CrystalPalace"))


ggplot(ArsVsCrystal, aes(ArsVsCrystal$Minutes))+
```

```
  geom_line(aes(y=1/ArsVsCrystal$Arsenal),colour="Red")+

  geom_line(aes(y=1/ArsVsCrystal$Draw),colour="Blue")+

  geom_line(aes(y=1/ArsVsCrystal$CrystalPalace),colour="Green")+

  ylab("Bet Prices")+

  xlab("Minutes")


##Az SVM modell illesztése a teljes szezonra.

##Az Arsenal - CP meccs paraméterei szerepeltek a TestData utolsó soraként

SVM_TrainData<-read.table("SVMPL_Learning.txt")

SVM_TrainData<-rename(SVM_TrainData, c("V1"="Result", "V2"="X1", "V3"="X2", "V4"="X3",
"V5"="X4"))



SVM_TestData<-read.table("SVMPL_Testing.txt")

SVM_TestData<-rename(SVM_TestData, c("V1"="Result", "V2"="X1", "V3"="X2", "V4"="X3",
"V5"="X4"))



set.seed(88888)



SVM_Predictor <- ksvm(factor(Result) ~ X1+

                 X2+

                 X3+

                 X4,

                SVM_TrainData,

                kernel='laplacedot',

                type="nu-svc",

                nu=0.3,
```

```
        #scaled=T,
        prob.model = TRUE)

SVM_TrainData$pred_svm <- predict(SVM_Predictor, SVM_TrainData, type="probabilities")[,2]
write.table(SVM_TrainData, file="SVM_result.csv", sep=";")

SVM_TestData$pred_svm <- predict(SVM_Predictor, SVM_TestData, type="probabilities")[,2]
write.table(SVM_TestData, file="SVM_Testresult.csv", sep=";")
```