Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Gertis Aru

Statistiline mudel kihlveokontoritega konkureerimiseks *English Premier League* näitel

Matemaatilise statistika eriala
Magistritöö (30 EAP)

Juhendaja: Prof. Kalev Pärna

Tartu 2017

Statistiline mudel kihlveokontoritega konkureerimiseks *English Premier League* näitel

Magistritöö

Gertis Aru

Lühikokkuvõte. Iga päev vaatavad miljonid inimesed mänge ja arutlevad jalgpalli tulemuste üle. Üheks hobiks on paljudel jalgpallifännidel tulemuste ennustamine. Käesoleva magistritöö eesmärk on leida statistiline mudel, mis aitaks meil võita kihlveokontoreid. Mudeli eesmärgiks on parimat panustamise strateegiat kasutades jääda konstantselt kasumisse. Lisaks vaadeldakse mudeli täpsuse statistikuid, mille abil saame objektiivselt öelda, milline mudel sobib jalgpalli tulemuste prognoosimiseks kõige paremini.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: statistiline mudel, kihlveokontor, panustamise strateegia, jalgpall, RPS

Statistical model to compete against bookmakers on a dataset from English Premier League

Master's thesis

Gertis Aru

Abstract. Every day many people are watching football matches and discuss about the outcome of it. One of the most popular hobies for many football fans is predicting the result of a football game. The aim of this master's thesis is to find a statistical model which helps us compete against bookmakers. The purpose of the model is to gain profit by using the best betting strategy and therefore being constantly efficient. Different statistics are considered for measuring the adequacy of models thus helping us determine the best model.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics

Keywords: statistical modelling, bookmaker, betting strategy, football, rank probability score

Sisukord

Sissejuhatus

Jalgpall on paljude hinnangul kõige suurem ja populaarseim spordiala maailmas. Inglise meistriliigas (EPL) osaleb kokku 20 meeskonda Inglismaalt ja Walesist, kes omavahel mängivad hooaja peale kokku 380 mängu. Kõik meeskonnad mängivad omavahel kaks korda läbi nii, et iga meeskond saab pidada ühe mängu kodus ja ühe mängu võõrsil. Seega mängitakse omavahel läbi kõik võimalikud kombinatsioonid. Mängu võit annab võitjameeskonnale kolm punkti ja kaotuse puhul null punkti. Kui aga mäng lõppeb viigiliselt, siis saab kumbki meeskond endale ühe punkti.

Kui kõik mängud on peetud, loetakse punktid kokku ja reastatakse meeskonnad hooaja jooksul kogutud punktide põhjal. See, kellel on hooaja lõpus kõige rohkem punkte, ongi EPL võitja antud hooajal. Esimesed kolm saavad otsepääsme UEFA Meistrite Liigasse ja neljandaks jäänud võistkonnal on samuti võimalus sinna kvalifitseeruda võisteldes teiste Euroopa riikide liigade madalama asetusega võistkondadega. Sarnaselt saab viies koht otsepääsme järgmise tugevusega Euroopa sarja, Euroopa Liigasse. Lisaks on kuuendal ja seitsmendal võimalus samuti sinna kvalifitseeruda, võisteldes teiste Euroopa liigade nõrgemate võistkondadega.

Iga päev vaatavad miljonid inimesed mänge ja arutlevad jalgpalli tulemuste üle. Üheks hobiks on paljudel jalgpalli fännidel tulemuste ennustamine. Motivatsiooniks võib erinevatel inimestel olla lihtsalt lõbu või soovitakse sellega ka reaalselt raha teenida. Igapäevaselt kasvab ka erinevate statistiliste mudelite arv, mis on mõeldud tulemuste prognoosimiseks. Antud töö esimeses peatükis tutvume nende mudelitega põgusalt.

Teises peatükis antakse ülevaade hea statistilise mudeli omadustest, millisega võiks võistelda kihlveokontoritega. Kolmandas peatükis vaadeldakse erinevaid headuse näitajaid ja tutvustatakse headuse näitajat RPS.

Neljandas ja viiendas peatükis võrreldakse omavahel erinevaid kihlveokontoreid ja antakse ülevaade erinevatest panustamisviisidest. Kuuendas peatükis valime välja parima mudeli. Lisaks viime läbi simulatsiooni ning võrdleme oma mudelit kihlveokontori omaga.

Töö on vormistatud tekstitöötlusprogrammiga MS Word ning praktiline pool on teostatud statistika tarkvaras R.

Töö valmimisele on kaasa aidanud juhendaja prof. Kalev Pärna, kellele töö autor soovib tänu avaldada nõustamise, paranduste ja täienduste eest.

1. Tulemuste ennustamine

Olgu meil lõplik arv võimalikke tulemusi (jalgpalli mängus on võimalikeks tulemuseks H, D, A[1] mis on vastavalt kodumeeskonna võit, viik või võõrsilmeeskonna võit). Olgu jalgpalli mängude korral sellised tõenäosused p_H, p_D, p_A. Mängude tõenäosuslik mudel aitab meil neid suuruseid määrata nii, et nad oleks kooskõlas mängude tulemustega.

1.1. Ülevaade tehtud mudelitest

Üheks esimeseks mudeliks, mis aitab ennustada jalgpalli mängu tulemusi, on loodud Maheri (1982) poolt. Ta modelleeris, kui mitu väravat meeskond i lööb meeskonna j vastu, kui meeskond i mängib kodus. Tähistame selle X_{ij}. Tegemist on Poissoni jaotusega juhusliku suurusega. Analoogselt formuleeris ta, kui palju meeskond i laseb endale samas mängus väravaid vastu lüüa. Ka siin on tegemist Poissoni jaotusega juhuliku suurusega (teistsuguse parameetriga) ja tähistame selle Y_{ij}. Lisaks eeldas ta veel, et X_{ij} ja Y_{ij} on sõltumatud.

Maher eeldas veel, et igal meeskonnal i on rünnaku efektiivsus α_i ja kaitse efektiivsus β_i. Suur rünnaku efektiivsus tähendab seda, et meeskond lööb palju väravaid. Kuna väärtused on normeeritud, siis summeeruvad need üle liiga meeskondade kokku üheks. Analoogselt väike kaitse efektiivsus tähendab seda, et meeskond laseb endale lüüa vähe väravaid. Lisaks arvestati veel koduväljaku eelisega γ, mis eelduse kohaselt oli kõigil meeskonnal võrdne. Saadud mudel aitas prognoosida iga mängu tulemust meeskonna i ja meeskonna j vahel järgmise tõenäosusega:

$$P(X_{ij} = x, Y_{ij} = y | \alpha, \beta, \gamma) = Poisson(x | \gamma \cdot \alpha_i \beta_j) \cdot Poisson(y | \alpha_j \beta_i), \quad (1)$$

kus $Poisson(x|\lambda)$ tähistab Poissoni jaotusfunktsiooni parameetriga λ, mis on arvutatud kohal x. Maher leidis suurima tõepära hinnangud oma mudeli parameetritele.

[1] Vigade vältimiseks kasutan vastavate ingliskeelsete sõnade esitähti: H=home, D=draw, A=away.

Küll aga on Maheri mudelil kaks suuremat probleemi. Esiteks, kui X on Poissoni jaotusest, siis vastavalt jaotuse tingimustele peavad kaks esimest momenti olema võrdsed, st $\mathrm{E}(X) = \mathrm{Var}(X)$. Karlis ja Ntzoufras (2000) leidsid oma analüüsis, et dispersioon on suurem kui keskväärtus ja seetõttu ollakse Poissoni jaotuse eeldustega vastuolus.

Teiseks ütleb loogika, et kui palju võimekam meeskond mängib palju nõrgema meeskonnaga, siis peaks tugevam meeskond lööma väga palju väravaid ja samal ajal laseb endale lüüa minimaalselt.

2. Mudelist

2.1. Mudeli omadused

Kuna eesmärk on välja arendada mudel, mis aitaks meil kihlveokontoreid võita, siis peavad mõned tingimused olema täidetud. Näiteks:

- Mudel peab arvesse võtma mängu mõlema meeskonna võimekused;
- Kodus mängivatel meeskondadel on koduväljaeelis;
- Üks olulisemaid suuruseid peab olema vorm, st kuidas vastav meeskond viimasel ajal mänginud on;
- Meeskonna võimekuse aitavad moodustada rünnaku efektiivsus (väravate löömine) ja kaitse efektiivsus (väravate ärahoidmine);
- Meeskondliku vormi arvestamisel tuleb arvesse võtta ka vastasmeeskonna võimekus.

Ilmselt ei oleks kõige mõistlikum arvestada kõiki mängude tulemusi. Antud suuruste hindamiseks moodustame statistilise mudeli, mis kõiki ülaltoodud eeldusi arvesse võtab. Loodava mudeli baasiks võtame Maheri loodud mudeli, kusjuures lubame vaadelda erinevate divisjonide andmeid koos ning eeldame, et meeskondlik tegutsemine ei ole kogu aeg konstantne.

Kuigi Maheri mudel on lihtsustatud ja tema eeldused ei pruugi alati kehtida, siis üldine struktuur peaks olema piisavalt täpne, et saaksime välja arendada strateegia, mis aitaks meil kihlveokontoreid võita.

Vaatleme kõigepealt sõltumatuse eeldust. Maher soovitas oma mudeli täiendamisel kasutada kahe muutujaga Poissoni jaotust, kuid antud jaotuste pere ei võimalda meil kirjeldada madalaskooriliste mängude tulemuste sõltumatust. Küll aga pakkusid Dixon ja Coles (1997) välja alljärgneva modifikatsiooni mudelist (1):

$$P(X_{ij} = x, Y_{ij} = y) = \tau_{\lambda,\mu}(x,y)\frac{\lambda^x \exp(-\lambda)}{x!}\frac{\mu^y \exp(-\mu)}{y!}, \qquad (2)$$

kus

$$\lambda = \alpha_i \beta_j \gamma ,$$
$$\mu = \alpha_j \beta_i$$

ning

$$\tau_{\lambda,\mu}(x,y) = \begin{cases} 1 - \lambda\mu\rho & kui\ x = y = 0, \\ 1 + \lambda\rho & kui\ x = 0, y = 1, \\ 1 + \mu\rho & kui\ x = 1, y = 0, \\ 1 - \rho & kui\ x = y = 1, \\ 1 & mujal \end{cases}$$

kus $\tau_{\lambda,\mu}(x,y)$ on funktsioon, mis teeb madalaskoorilised viigid (0:0 ja 1:1) veidi tõenäoselimaks ja tulemused 1:0 ja 0:1 veidi vähem tõenäolisemaks, võrreldes Maheri mudeliga.

Uus parameeter ρ, mis rahuldab tingimust

$$\max\left(-\frac{1}{\lambda}, -\frac{1}{\mu}\right) \leq \rho \leq \min\left(\frac{1}{\lambda\mu}, 1\right),$$

on sõltuvuse parameeter. Väärtus $\rho = 0$ vastab sõltumatuse olukorrale.

Teine piirang Maheri mudeli juures on see, et see on staatiline. See tähendab, et iga meeskonna rünnaku ja kaitse efektiivsused on ajas konstantsed.

2.2. Mudeli tuletamine

Mudeli (2) põhjal tuleb meil *n* meeskonna puhul hinnata rünnaku parameetrid $\{\alpha_1,...,\alpha_n\}$, kaitse parameetrid $\{\beta_1,...,\beta_n\}$, sõltuvuse parameeter ρ ja koduväljakueelise parameeter γ. Et hoiduda mudeli üleparametriseerimisest, lisame mudelile juurde kitsenduse:

$$n^{-1}\sum_{i=1}^{n}\alpha_i = 1.$$

Inglismaa liigasüsteem sisaldab Meistriliigat ja madalamaid divisjone. Meie vaatleme aga kuni esiliigani, kus mängib vastavalt 20 ja 24 meeskonda. Seega on meie mudelil 89 leitavat parameetrit (44 rünnaku-, 44 kaitseparameetrit ja 1 koduväljakueelis).

Peamine abivahend parameetrite hindamiseks on tõepärafunktsioon. Olgu meil mängud tähistatud $k=1,\dots, N$ ja neile vastavad tulemused (x_k, y_k). Siis avaldub tõepärafunktsioon kujul:

$$L(\alpha_i, \beta_i, \rho, \gamma, i = 1, \dots, n) = \prod_{k=1}^{N} \tau_{\lambda_k,\mu_k}(x_k, y_k) \exp(-\lambda_k) \, {\lambda_k}^{x_k} \exp(-\mu_k) \, {\mu_k}^{y_k}, \qquad (3)$$

kus

$$\begin{aligned} \lambda_k &= \alpha_{i(k)}\beta_{j(k)}\gamma, \\ \mu_k &= \alpha_{j(k)}\beta_{i(k)}, \end{aligned} \qquad (4)$$

kus $i(k)$ ning $j(k)$ tähistavad vastavalt kodu- ja võõrsilmeeskonda mängus k.

Valemisse (3) on tõepärafunktsiooni lisatud meistriliiga ja esiliiga meeskonnad. Sellel on aga kaks tagajärge: esiteks iga meeskonna parameetrid peaksid peegeldama erinevate divisjonide suhtelist kvaliteeti ja teiseks on parameetrid ainult siis hinnatavad, kui omavahel on kohtunud erinevate divisjonide meeskonnad. Õnneks on meeskondadel võimalik pärast hooaja lõppu liikuda kõrgemale või madalamale divisjoni. Seega on see probleem lahendatud. Lisaks aitaks seda lahendada ka karikamängude lisamine, kus omavahel kohtuvad erinevate divisjonide meeskonnad. Kuna parameetrid on normeeritud üle divisjonide, siis saab mudeli abil hinnata ka mängu tulemust, kus omavahel kohtuvad erinevate divisjonide meeskonnad, näiteks karikamängus.

2.3. Mudeli täiustus

Eelnevalt konstrueeritud mudeli (3) puuduseks on veel, et selle parameetrid on staatilised, ehk meeskondadel on ajas konstantne vorm suurustega α_i ja β_i. Reaalses elus on aga meeskonna vorm rohkem dünaamilisem. Näiteks võib meeskonnas olla vahetunud peatreener. Aastal 2016 sai Liverpooli lootsiks Jürgen Klopp, millest tulenevalt hakkas meeskond palju paremini esinema või kui selle aasta alguses sai Leicester City uueks peatreeneriks Craig Shakespeare, kelle käe all teenis meeskond alguses viis võitu järjest. Lisaks võib olla vormi muutuse põhjuseks põhimängija vigastus või hooaja keskel mõne uue staarmängija soetamine. Sellest tulenevalt tuleks vormi vaadata viimaste mängude tulemuste pinnalt. Antud aspekti arvesse võtmiseks eeldame, et varasema minevikuga informatsioon on väiksema kaaluga, kui hilisem informatsioon. Seega hindame parameetrid igal ajahetkel t võttes arvesse mineviku mängude tulemused. Sellest tulenevalt, konstrueerime võrrandist (3) "pseudo-tõepära" iga ajahetke t jaoks:

$$L_t(\alpha_i, \beta_i, \rho, \gamma, i = 1, \dots, n) = \prod_{k \in A_t} \left\{ \tau_{\lambda_k, \mu_k}(x_k, y_k) \exp(-\lambda_k) {\lambda_k}^{x_k} \exp(-\mu_k) {\mu_k}^{y_k} \right\}^{\phi(t - t_k)}, \quad (5)$$

kus t_k on aeg, millal kohtumine k mängiti. $A_t = \{k\colon t_k < t\}$, λ_k ja μ_k on samad nagu võrrandis (4) ja ϕ on mittekahanev ajafunktsioon.

Maksimeerides nüüd avaldist (5) ajahetkel t, saame parameetrite hinnangud, mis tuginevad mängudele kuni ajahetkeni t. Nüüd võtab mudel arvesse ka meeskondade vormi muutusi. Muutes ϕ väärtusi saame määrata, kui palju mineviku tulemused mõjutavad prognoose.

2.4. Kaalufunktsiooni ϕ valik

Võrrandis (5) on ϕ defineerimiseks mitu erinevat võimalust. Üheks selliseks on:

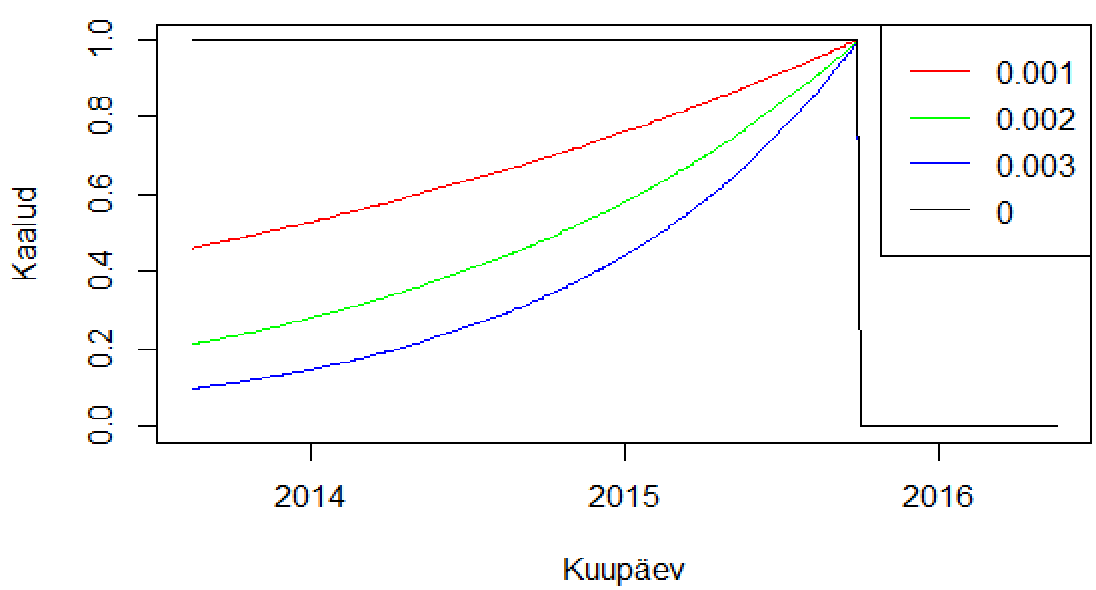
$$\phi(t) = \begin{cases} 1 & t \leq t_0, \\ 0 & t > t_0 \end{cases}$$

Sellisel juhul antakse ajahetkel t iga viimase t_0 ajaühiku jooksul olevale tulemusele võrdne kaal. Parem oleks vaadelda alljärgnevat mudelit:

$$\phi(t) = \exp(-\xi t), \tag{6}$$

Sellisel juhul oleks kõik eelnevad tulemused eksponentsiaalselt (vastavalt parameetrile $\xi > 0$) väiksema kaaluga. Meie eelnevalt loodud staatiline mudel (3) on selle mudeli erijuht tingimusel $\xi = 0$ ning kui ξ võtab väga suure väärtuse, siis on vastavalt kõige hilisemad tulemused kõige suurema kaaluga.

Oma mudeli loomisel võtame me funktsiooni (6) arvesse ja kujutame võimalikud kaalude väärtused joonisel 1. Nagu näha, juhul $\xi = 0$ saame igale mängule sama kaalu. Lisaks on hetke kuupäevaks pandud 2015. aasta oktoober, et illustreerida, et tuleviku kuupäevade kaalud on kõik mitteinformatiivsed, ehk väärtusega 0.



Joonis 1. Erinevate kaalude muutused ajas

2.5. Tulemuste „sisse“ vaatamine

Kui meeskond ei saavuta väga häid tulemusi, siis paljude treenerite ja meeskonna poolehoidjate peamiseks argumendiks on, et tuleb vaadata sügavamale tulemuse sisse. See tähendab seda, et mäng oleks tulnud võita, kui oleks olnud rohkem õnne. Antud idee tundub meie uuringu aspektis oluline, kui oleme loomas mudelit, mis aitab tulemusi paremini prognoosida.

Kuna tänapäeval avaldatakse palju erinevat statistikat erinevate mängude kohta, siis oleks patt neid mitte kasutada või vähemalt kaaluda nende kasutamist. Antud teema valguses on oma uuringus toonud Rue ja Salvesen (2000) välja, et arvesse tuleks võtta lööke väravale ja lööke värava „raamidesse“. Mõistlik oleks uurida ka erinevate kaartide mõju. Võib olla alust arvata, et kui meeskond teenib mängus punase kaardi ja jätab oma meeskonna vähemusse, siis sellel on mängu edasisele kulgemisele negatiivne mõju, samas jälle vastasmeeskonnal on suurem võimalus värava löömiseks.

3. Mudeli headuse hindamine

Antud peatükk baseerub (Constantinou, Fenton, 2012) artiklile, kus analüüsiti erinevaid mudeli headuse näitajaid. Mudeli täpsuse hindamine on üks olulisemaid aspekte mudeli juures. Selle puudumisel ollakse mudelile hinnangu andmisel võrdlemisi subjektiivsed ning selle tulemusena võivad jääda vastuseta küsimused: a) Kas antud mudel on piisavalt täpne? või b) milline kahest mudelist on parem. Konkreetselt selle töö raames on oluline, et me leiaksime kõige objektiivsema hinnangu jalgpalli tulemusi prognoosivale mudelile. Kuna mõningad statistikud eelistavad mudelit α mudelile β ja teised vastupidi, siis on äärmiselt oluline, et mudeli headuse hindaja valimisel langetaksime oma otsuse just sellisele, mille korral saaksime objektiivselt vastata paragrahvi alguses püstitatud küsimustele.

3.1. Ülevaade mudeli headuse statistikutest

Mudeli headuse hindamisel võtsid Jose jt (2009) aluseks, et statistik peab olema tehtud selliselt, et võimalikud mängu tulemused H, D, A oleksid mõõdetud järjestusskaalal ning mitte nominaalsel skaalal. See tähendab, et tulemus D on lähemal tulemusele H, kui A on tulemusele H. Seda seetõttu, et kui kodu meeskond juhib ühe väravaga, siis on vaja vaid ühte tabamust, et tablool oleksid viiginumbrid, st mudeli mõttes olekust H liikumist olekusse D ning võõral võistkonnal on vaja lüüa kaks tabamust, et toimuks liikumine olekusse A. Tuleb välja, et nii loogiline kui see ka tundub, ometi paljud varasemad uuringud ei ole seda arvestanud. Et seda illustreerida, vaatleme viit võimalikku stsenaariumi, et näidata, et mitte ükski varem loodud statistik ei ole suutnud kõiki stsenaariumeid rahuldada. Seega on paljud mudelid kasutanud ebaadekvaatseid statistikuid mudeli headuse hindamisel.

Enne kui me kirjeldame varem loodud statistikuid, teeme järgmised eeldused:

1. Igas nähtuse jaoks (näiteks jalgpalli mäng) on meil r võimalikku tulemust, st jalgpalli mängu puhul r = 3;

2. Mudeli hinnangud on vastavuses tõenäosustega $\{p_1, p_2, \dots, p_r\}$;

3. Vastavad võimalikud tulemused on $\{e_1, e_2, \dots, e_r\}$. Seega jalgpalli korral on iga e_i kas 1 või 0 (kas läks „täppi“ või mitte). Olgu w mängu lõppresultaat. Seega saame kirjutada:

$$e_i = \begin{cases} 1, & kui\ i = w \\ 0, & kui\ i \neq w \end{cases}$$

Mudeli headuse statistiku saab seega defineerida kaheastmeliselt:

1. Statistiku väärtus iga üksiku mängu korral, kui on teada selle sündmuse tõenäosus (näiteks üks jalgpalli mäng);

2. Meetod, mis agregeerib kõik mikrostatistikute väärtused üle kõikide mängude. Näiteks võib kasutada kas mikrostatistikute aritmeetilist keskmist üle kõigi mängude või mikrostatistikute summat üle kõigi mängude. Geomeetriline keskmine kasutab koondstatistiku määramiseks korrutustehet ja seega „karistab“ igat valesti tehtud ennustust rangemalt.

Varem loodud mudelite headuse näitajad üksikmängude korral (mikrostatistik) on defineeritud järgnevalt:

1. Binaarne otsus: Mikrostatistik saab väärtuseks 1, kui $p_w > p_i$, kui $i \neq w$ ja 0 vastasel juhul. Antud meetod otsib maksimaalset tõenäosust võttes arvesse kõik võimalikud mängude tulemuste tõenäosused. Kokkuvõtvalt, statistiku väärtus ei sõltu üksikutest tõenäosustest;

2. Brieri statistik: Antud mikrostatistik on defineeritud järgnevalt:

$$\sum_{i=1}^{r} (p_i - e_i)^2$$

3. Geomeetriline keskmine: Iga mängu jaoks on statistik lihtsalt p_w;

4. Informatsiooni kadu: Iga mängu jaoks on see defineeritud kui $-log_2 p_w$;

5. Maksimaalne log-tõepära hinnang (MLLE): Maksimaalne tõepära hinnang on statistikas standardne meetod parameetrite hindamiseks. See sobitab sellise tõenäosusjaotuse, mille korral antud mängutulemused on „kõige tõenäolisemad“. Informatsiooni kadu ja maksimaalne log-tõepära hinnangute kujud on võrrandites erinevad, kuid annavad tõenäosustele identseid hinnanguid. Binoomjaotuse puhul kasutame parameetreid n (katsete arv) ja t (õnnestumiste arv). Kuna meil on aga ainult üks vaatlus iga mängu tõenäosuste hindamiseks (n ja t on võrdsed ühega), siis on log-tõepära defineeritud kui $\ln(p_w)$. Seega, et me ei teeks ebavajalikke arvutusi, siis defineerime iga mängu jaoks MLLE kui $\ln(p_w)$.

Varem tehtud uuring (Gneiting, Raftery, 2007) näitab, et mitte ükski ülalnimetatud mikrostatistik ei ole sobilik, kui võimalikud tulemused on järjestusskaalal. Seega ei sobi need jalgpalli tulemusi prognoosivasse mudelisse. Me näitame seda kasutades viit erinevat stsenaariumi (Tabel 1), kus iga stsenaariumi jaoks on meil kaks mudelit (α ja β) ning nende hinnangud tegelikele tulemustele. Väljatoodud stsenaariumid näitavad, et varem defineeritud mikrostatistikud ei arvesta järjestusskaalat oma tõenäosusjaotustes. Tegelikkuses on võimalik välja tuua ka veel mitmeid stsenaariume, kuid üldiselt võiksid need olla piisavad, et näidata eelnevalt loodud statistikute puuduseid. Iga näite korral peaks olema intuitiivselt arusaadav, milliseid hinnanguid peaks pidama paremaks.

Tabel 1. Mudelite võrdlus viie stsenaariumi põhjal (hüpoteetiline näide)

Mäng	Mudel	*p(H)*	*p(D)*	*p(A)*	Tulemus	„Parim mudel“
1	α	1	0	0	*H*	α
	β	0.9	0.1	0		
2	α	0.8	0.1	0.1	*H*	α
	β	0.5	0.25	0.25		

3	α	0.35	0.3	0.35	D	α
	β	0.6	0.3	0.1		
4	α	0.6	0.25	0.15	H	α
	β	0.6	0.15	0.25		
5	α	0.57	0.33	0.1	H	α
	β	0.6	0.2	0.2		

Seda seetõttu, et:

a) Mäng 1: (Tuleks arvesse võtta ideaalne täpsus) Mudel α ennustab õige tulemuse täieliku kindlusega ja seega peaks saama parema headuse näitaja, kui ükskõik mis teine vähem täpsem mudel;

b) Mäng 2: (Tuleks arvesse võtta tegeliku tulemuse hinnanguline väärtus) Mõlemad mudelid α ja β annavad suurima tõenäosuse tegelikule tulemusele H ning teiste tulemuste hinnangud on võrdsed. Kuna mudeli α hinnang tegelikule tulemusele on suurem, siis peaks selle mudeli headuse näitaja olema parem;

c) Mäng 3: (Tuleks võtta arvesse teiste tulemuste prognoose) Kuna siin on tegelikuks tulemuseks D, siis on mõlemad valed hinnangud on võrdsel kaugusel tegelikust tulemusest. Seega siin pole meil järjestuse eeldust vaja täita. Küll aga peaks headuse näitaja tuvastama, et mudeli α puhul on tõenäosusjaotus täpsem ja tema hinnang viigile on parem kui mudelis β, mis hindab tugevalt kodumeeskonna võitu;

d) Mäng 4: (Tuleks arvesse võtta järjestust, kui valede ennustuste tõenäosuste summa on võrdne) Mõlemad mudelid α ja β annavad võidu tulemusele H võrdse tõenäosuse. Seekord annavad nad samad tõenäosused (0.25 ja 0.15) ka valedele ennustustele, aga teises järjekorras. Seega peab mudeli headuse statistik tagama selle, et α on täpsem, kuna tema tõenäosusjaotus annab täpsema hinnangu kodumeeskonna võidule;

e) Mäng 5: (Tuleks arvesse võtta terve tõenäosusjaotus) Kuigi α annab tegelikule tulemusele H väiksema tõenäosuse kui β, siis on α hinnang kodumeeskonna võidule parem kui mudelis β. Antud mäng on kõige vastuolulisem, aga saame seda seletada ka sellise nurga alt, et kui ennustaja on kindel, et kodumeeskond ei võida (ka selliseid ennustusi saab kihlveokontorites teha) ja teeb sellele panuse, mis tähendab, et panus toob raha juurde kui tegelik tulemus on H või D. Eeldame, et α ja β ennustused on esitatud erinevate kihlveokontorite poolt, siis kihlveokontor α maksab tegeliku tulemuse korral vähem (toome veelkord välja, et antud kihlveokontor pidas kodumeeskonna kaotuse tõenäosuseks 0.1, kui kihlveokontor β pidas kaotuse tõenäosuseks 0.2).

Tabel 2 näitabki antud statistikute tulemusi iga stsenaariumi jaoks ning lisaks teeb kindlaks, kas antud näite põhjal on hinnangud mudeli headustele õiged. „Linnuke" tähendab, et statistik hindab mudelit α paremaks kui mudelit β , „rist" tähendab, et statistik annab sama väärtuse mõlema mudeli korral ning kaks „risti" tähendab, et statistik hindab valesti mudeli β paremaks. Näitajate binaarne otsus, geomeetriline keskmine ja MLLE korral tähendab suurem tulemus paremat hinnangut, kuid Brieri statistiku ja informatsiooni kao korral on paremal mudelil madalam väärtus.

Tabel 2. Mudeli headuse statistikud iga stsenaariumi korral

Mäng (mudel)	Binaarne otsus	Brieri statistik	Geomeetriline keskmine	Informatsiooni kadu	MLLE statistik
1	✓	✓	✓	✓	✓
α	1	0	1	0	0
β	0	0.02	0.90	0.15	-0.11
2	×	✓	✓	✓	✓
α	1	0.06	0.80	0.32	-0.22
β	1	0.38	0.50	1	-0.69
3	×	✓	×	×	×
α	0	0.74	0.30	1.74	-1.20
β	0	0.86	0.30	1.74	-1.20
4	×	×	×	×	×
α	1	0.25	0.60	0.74	-0.51

β	1	0.25	0.60	0.74	-0.51
5	✗	✗✗	✗✗	✗✗	✗✗
α	1	0.30	0.57	0.81	-0.56
β	1	0.02	0.60	0.74	-0.51

Nagu näha, siis mitteükski mudeli headuse statistik ei anna „õiget“ tulemust iga näite korral. Veelgi enam, stsenaariumite 4 ja 5 ei suuda ükski statistik pidada paremaks mudelit α.

3.2. Mudeli headuse näitaja RPS

Rank Probabily Score (RPS) sai laiemat tuntust 1969 aastal (Epstein, 1969). Antud suurus näitab, kui suure vea me ennustamisel teeme ning ta saab omale väärtuseid vahemikus 0 kuni 1, kusjuures 0 on täiesti täpne ennustus. RPS võtab arvesse tõenäosusjaotuse, st kodu meeskonna võidu, viigi ja võõrsil meeskonna võidu tõenäosused, mitte ainult tegeliku tulemuse. See tähendab ühtlasi, et kui mängu võidab kodumeeskond, siis viigitõenäosust peetakse väiksemaks veaks, kui võõrsil mängiva võistkonna võidu tõenäosust. Antud suurus arvutatakse valemiga:

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r-1}\left(\sum_{j=1}^{i}(p_j - e_j)\right)^2,$$

kus r on võimalike väärtuste arv ning p_j ja e_j on vastavalt ennustused ja mängu tegelik tulemus iga j korral.

Tabelis 3 mängime samad stsenaariumid läbi, mis said eelnevalt kirjeldatud (Tabel 1) ning anname vastavad kumulatiivsed jaotused (prognoositud ja tegelikud). Madalam väärtus näitab, et tegu on parema prognoosiga. Erinevalt eelmistest statistikutest hindab RPS mudeli α parimaks iga stsenaariumi korral.

Antud headuse näitajat kasutame ka oma mudeli kirjeldamiseks.

Tabel 3. RPS statistiku väärtused iga stsenaariumi korral, mis on toodud Tabelis 1

Mäng	Mudel	$\sum_{j=1}^{i=1,2} p_j$	$\sum_{j=1}^{i=1,2} e_j$	RPS
1	α	1,1	1,1	0.000
	β	0.90,1	1,1	0.005
2	α	0.80,0.90	1,1	0.025
	β	0.50,0.75	1,1	0.156
3	α	0.35,0.65	0,1	0.122
	β	0.60,0.90	0,1	0.185
4	α	0.60,0.85	1,1	0.091
	β	0.60,0.75	1,1	0.111
5	α	0.57,0.90	1,1	0.097
	β	0.60,0.80	1,1	0.100

4. Panustamisest

Jalgpalli ennustamist peetakse tänapäeval järjest huvitavamaks valdkonnaks, mille peal rakendada erinevaid mudeleid. Inimeste järjest suurem huvi panustamise (viimasel ajal järjest rohkem interneti vahendusel) vastu tähendab, et huvitutakse aina enam, kuidas kihlveokontorid oma mängudel koefitsiente määravad. Kuna jalgpall on oma olemuselt üllatusterohke, siis selle tulemusena on jalgpallimängu lõppresultaadil kihlveokontorile majanduslikus edus järjest suurem roll. Seetõttu on hakatud järjest rohkem rõhku panema optimaalseimate koefitsientide leidmisele. *Global Betting and Gaming Consultants* (2001) väitis, et Suurbritannia jalgpalli kihlveokontorite käive ulatus 1998. aastal 2 miljardi naelani ning jalgpalli tulemustele panustamine on Suurbritannia kihlveokontorite jaoks kõige kiiremini arenev valdkond. Juba 2008. aastal oli ainuüksi ühe suurima kihlveokontori (bwin Group, 2009) käive umbes 2,92 miljardit naela, kusjuures eelmise aastaga võrreldes oli nende käive kasvanud 31,4%.

4.1. Kihlveokontorite täpsus

Julgustades ja meelitades panustajaid tegema nii palju ennustusi, kui on võimalik, suurendavad kihlveokontorid oma kasumlikkust. Seega on kihlveokontorite edukuse hindamiseks oluline nii puhaskasum, kui ka nende koefitsientide määramise oskus, mis tooks ennustajaid nende juurde panuseid tegema. Antud alapeatükis näitame erinevate kihlveokontorite täpsust koefitsientide määramisel.

Antud uuringu läbiviimiseks kasutame me andmeid veebilehelt www.football-data.co.uk Ennustuste täpsuse hindamiseks kasutame statistikut RPS. Antud suuruse kasumlikkus jalgpalli tulemuste ennustamise mudelis sai meil kirjeldatud peatükis 3.

4.2. Kihlveokontorite omavaheline võrdlus täpsuse hindamisel

Kuna kavalamal panustajal on võimalik valida, millise kihlveokontori juures ta oma panuse teeb, siis on kasulik võrrelda kihlveokontoreid omavahel. Erinevate kihlveokontorite võrdlemiseks vaatleme kolme suuremat kontorit aastatel 2014-2016[2]. Võtame aluseks meie andmestiku, kus on kirjeldatud iga mängu korral erinevate kihlveokontorite koefitsiendid ja lõpptulemus. Nende suuruste pealt arvutame headuse näitaja RPS.

Vaadates antud suuruseid tabelis 4 selgub, et kihlveokontorid on oma täpsuses peaaegu võrdsed. Mõlemal hooajal oli teistest veidi ebatäpsem kihlveokontor William Hill. Antud tulemustele tuginedes võrdleme peatükis 6 meie koostatud mudelit William Hilli kihlveokontoriga.

Tabel 4. Kihlveokontorite täpsuse võrdlus

	$RPS_{KESKMINE}$		$RPS_{MEDIAAN}$		RPS_{SD}	
Kihlveokontor	14/15	15/16	14/15	15/16	14/15	15/16
Bet365	0.1968	0.2084	0.1500	0.1581	0.1416	0.1413
Ladbrokes	0.1977	0.2103	0.1537	0.1589	0.1414	0.1452
William Hill	0.1995	0.2105	0.1551	0.1596	0.1437	0.1457

Suuremad erinevused tulevad välja erinevate hooaegade vahel. Antud tulemust saab interpreteerida nii, et mõningatel hooaegadel käituvad meeskonnad rohkem ootuspärasemalt. Näiteks võib hooaega 2015/2016 pidada küllaltki üllatavaks hooajaks, kui suure üllatuse sepistas Leicester City meeskond, võites tollel hooajal liigatiitli. Lisaks ei läinud sellel hooajal kõige paremini kahel väga tugeval meeskonnal – Chelsea ja Mancheter United, kes lõpetasid hooaja tabeli keskpaigas.

[2] Aastad 2014-2016 sisaldavad endas kahte hooaega: 2014-2015, 2015-2016

5. Jalgpalli mängudele panustamine

Jalgpall on üks maailma populaarsemaid spordialasid harrastajate hulgas, kui ka üks kõige populaarsemaid spordialasid, kuhu teha oma panuseid. Kõige traditsioonilisem panustamiseviis on panna oma raha mingile mängu tulemusele (*H, D, A*). Veel on võimalik panustada mõnele konkreetsele seisule, seda nii lõppresultaadile kui ka poolaja tulemusele, *handicap* panustamine, mängus löödud väravate arvule panustamine või näiteks kogu liiga võitja ennustamine. Tänapäeval on võimalik panustada ka märksa kummalisematele sündmustele, nagu näiteks mängus jagatud kaartide arvule, nurgalöökide arvule jne. Selles peatükis anname ülevaate erinevatele panustele, mida on võimalik teha. Kuna me oma mudeliga analüüsime kahte kõige populaarsemat panustamisviisi, siis saavad suurema tähelepanu just need.

5.1. Mängu võitja ennustamine ehk *fixed odds betting*

Mängu võitja ennustamine on üks kõige populaarsemaid ennustamisviise. See töötab selliselt, et kihlveokontorid pakuvad tõenäosuseid mängu igale võimalikule tulemusele (jalgpallimängus on võimalikke tulemusi kolm: H, D, A) ning panustaja saab otsustada, millisele mängule ta soovib millist panust teha. Näiteks, kui toimub lahing Evertoni ja Leicester City vahel ning kihlveokontor annab panustajatele järgmised koefitsiendid:

Tabel 5. Mängu koefitsiendid. Allikas Triobet 09.04.2017

Kodu meeskond	Võõrsil meeskond	H koefitsient (Evertoni võit)	D koefitsient (viik)	A koefitsient (Leicesteri võit)
Everton	Leicester City	1.74	4.00	4.80

Kui panustaja otsustab, et mängu võidab Everton ja panustab sellele mängule 10€ ning kui Everton võidabki, siis saab panustaja tagasi vastavalt tema tehtud panusele korrutatult koefitsiendiga ehk antud juhul saaks sellise sündmuse toimumise korral raha tagasi 17.4€, sh netovõit 7.4€. Kui aga kohtumine peaks lõppema viigiliselt või Leicester City peaks kohtumise võitma, siis on ennustaja oma panustatud rahast ilma.

Inglismaal näiteks kujutatakse koefitsiente kujul x/y (näiteks 1/2, kus on vaja panustada 2 ühikut, et teenida juurde 1 ühik). Eestis (nagu ka toodud näites) aga kujutatakse antud koefitsiente tõenäosuse pöördarvuna. Seega 1/2 on Eesti mõistes 1.50.

Fikseeritud panustega ennustamisel avaldatakse koefitsiendid enamasti mõned päevad enne sündmust ning inimesed saavad teha panuseid vastavalt avaldatud koefitsientidele. Internetis resideeruvad kihlveokontorid võivad aga muuta oma koefitsiente nii palju kui nad tahavad, vastavalt panustamisaktiivsusele ja mõnele uuele faktorile, mis võivad mängukulgu mõjutada. Kihlveokontorite eesmärk on hoida rahavoog tasakaalus ja seeläbi garanteerida omale fikseeritud kasum. Rahavoo tasukaalus hoidmiseks vähendatakse koefitsiendi suurust (sündmuse toimumise korral on makstav tasu väiksem), kui sinna peale tehakse eeldatavast rohkem panuseid ning suurendatakse, kui tehakse palju vähem panuseid, et panustajale ahvatlevamaks antud tulemuse peale panustamine teha.

5.2. *Over/Under* panustamine

Ka see meetod on üks laialdasemat kasutust leidnud meetod panustamisel. Antud meetod töötab nii, et kihlveokontorid valivad suvalise statistiku mängus (enamasti kahe meeskonna peale kokku löödud väravate arv) ning panustajad saavad pakkuda, kas nende hinnangul võib tegelik number kujuneda sellest suuremaks või väiksemaks. Jalgpalli korral on üks populaarsemaid sedatüüpi panustamisviis, et kas kahe meeskonna peale lüüakse üle või alla 2,5 värava. Näiteks eelnevalt toodud mängu (Everton – Leicester City) põhjal olid kihlveokontori poolt pakutud koefitsiendid järgmised:

Tabel 6. Koefitsiendid Over/Under panustamise korra. Allikas: Triobet 09.04.2017

Üle 2,5	Alla 2,5
1.82	1.91

Seega, kui panustaja otsustab, et antud mängus lüüakse üle 2,5 värava (vähemalt 3 väravat) ja panustab sellele sündmusele 10€ ning lüüaksegi vähemalt kolm väravat, siis saab panustaja raha tagasi vastavalt tema tehtud panus korrutatud pakutud koefitsiendiga ehk antud näite korral 18.2€, sh netovõit on 8.2€. Kui aga lüüakse kahe meeskonna peale kuni kaks väravat, siis kaotab panustaja oma raha.

5.3. Panustamise strateegiad

Kõige edukamad ennustajad on sellised, kes otsivad ideaalseid valikuid, kuhu oma panuseid teha, et seeläbi saada võimalikult suur kasum. Iga mängu korral on tulemuse kujunemisel oluliseks ka õnnefaktor, kus mängu lõppresultaat võib kujuneda üllatavaks kõigile ning seeläbi kaotavad enamik panustajatest. Küll aga võiks teadlik ja distsiplineeritud panustaja pikemas perspektiivis ikkagi lõpuks võita.

Selleks, et lõpuks olla kasumis, tuleb panustajatel välja töötada kõige parem panustamise strateegia. Panustaja eesmärk on maksimiseerida oma võite ja minimiseerida kaotuseid. Kui ennustajal õnnestub mingisuguse mudeli või muul viisil kindlaks teha iga mängu *H, D, A* tõenäosused, siis üheks parimaks panustamise strateegiaks võiks olla Kelly (1956) kriteerium. Antud kriteerium on kirjeldatud järgnevalt:

$$S = \frac{p * w - 1}{w - 1},$$

kus *S* on panuse suurus sündmusele *H, D* või *A* st mõõdetuna osana algkapitali suurusest, *p* on tõenäosus, et see sündmus (*H, D* või *A)* toimub ning *w* on kihlveokontorite poolt väljapakutud koefitsient sama sündmuse jaoks. Antud meetodil on kolm väga olulist omadust, mis on kirjeldatud Hausch, Lo ja Ziemba (1994) poolt:

- See maksimeerib kapitali kasvumäära;
- See minimiseerib asümptootiliselt aega, mis kulub eesmärgi saavutamiseks;

- See töötab pikemas perspektiivis paremini kui iga teine panustamise strateegia peaaegu kindlalt.

Antud kriteerium on laiemalt tuntud majandus- ja finantsvaldkonnas kui optimaalne kasvu strateegia, kapitali kasvu kriteerium jne.

Kelly kriteeriumi üheks omaduseks on see, et see eeldab, et tõenäosusjaotus on teada. Küll aga on ka teisi võimalikke strateegiaid, mida panustajad saavad kasutada. Üheks selliseks on minimiseerida laostumise tõenäosust. Alternatiivselt on võimalik kasutada ka Kelly kriteeriumi erivorme, kus panuse suurus on vaid osa Kelly kriteeriumiga saadud panuse suurusest (näiteks 1/2). Antud juhul on panused väiksemad ja seetõttu ollakse veidi konservatiivsemad. Lisaks on veel võimalik kasutada fikseeritud panuseid, kus iga panuse suurus on fikseeritud ja ei sõltu võidu tõenäosusest. Antud töö raames viime läbi katsetused fikseeritud panuse, Kelly, ½ Kelly ja ¼ Kelly kriteeriumite korral.

6. Tulemustest

Antud analüüsi viisime läbi viimase kahe *English Premier League*'i hooaja andmetele tuginedes, võttes siia juurde ka Inglismaa esiliiga tulemused. Seega on meil 2014-2015 aastate hooaja nii meistri-, kui ka esiliiga andmed ja 2015-2016 hooaja meistriliiga andmed. Andmed saadi veebisaidilt www.football-data.co.uk. Antud andmestikud hõlmasid endas järgmist informatsiooni:

- Mängus osalenud meeskonnad
- Eristatud kodumeeskond
- Mängu kuupäev
- Mängus löödud väravad (nii kodu- kui ka võõrsilmeeskonna poolt)
- Mängus saadud karistused (kollased ja punased hoiatuskaardid)
- Erinevate kihlveokontorite koefitsiendid

Oma mudeli koostamisel kasutati olemasolevaid andmeid ja lisaks mõningaid arvutuslikke suuruseid. Tulemusi analüüsiti statistika tarkvaraga R.

6.1. Valime ennustamiseks parima võimaliku mudeli

Antud alapeatükis vaatleme, milline mudel annaks kõige paremaid tulemusi jalgpalli tulemuste prognoosimisel. Erinevate mudelite võrdlemise aluseks võtame statistiku RPS, millel peatusime lähemalt alapeatükis 3.2.

Tuginedes eelnevale teadmisele võtame aluseks Dixoni ja Colesi poolt loodud mudeli (2). Vaatleme, kas sõltumatuse parameeter ρ teeb mudeli paremaks. Kas tasub juurde lisada ka meeskondade vormi arvestava lisaparameetri (viimastes mängudes saadud punktide aritmeetiline keskmine)? Lisaks vaatame, kas mudel muutub paremaks, kui lisada andmestikku juurde punased hoiatuskaardid.

Vaadeldavaid mudeleid on seega kolm:

- Dixoni ja Colesi mudel (M1)
- Mudel, mis arvestab ka viimaste mängude vormi lisaparameetrit (M2)
- Tunnuse „punaste kaartide arv“ lisamine (M3)

Nagu tabelist 7 näeme, siis parimaks mudeliks osutus mudel M2, kuna selle mudeli korral on RPS statistiku väärtus kõige väiksem.

Tabel 7. Mudelite omavaheline võrdlus

Mudel	RPS
M1	0.2089
M2	0.2088
M3	0.2098

Veidi üllatavalt ei muutnud vormi lisaparameetri lisamine mudelit palju paremaks ning edasiminek oli vaid komakoha täpsuses. Punaste kaartide lisamine ei teinud mudelit paremaks. Seega oma edasises analüüsis läheme edasi mudeliga 2.

Leiame antud mudelile nüüd ka parima kaalufunktsiooni $\phi(t) = \exp(-xt)$. Proovime läbi erinevate x väärtuste korral.

Tabel 8. Erinevate kaalude võrdlemine mudelis M3

x_i	RPS
0	0.2097
0.001	0.2088
0.002	0.2084
0.003	0.2088
0.01	0.2110

Nagu näha, siis natuke erineva mõju mineviku tulemused annavad. See tähendab, et lähiminevikus toimunud mängud annavad prognoosile suurema kaalu, kui kaugema aja tagant toimunud lahing. Sellest tulenevalt võtame oma edasiseks mudeliks selle, mille korral $x_i = 0.002$.

6.2. Meie mudel vs kihlveokontori mudel

Parim viis testimaks, kuidas meie mudel töötab, on võrrelda seda mõne kihlveokontoriga. Kui meie mudel toodab meile kasumit, siis tähendaks see, et meie mudel on kihlveokontori omadest parem. Seega saaksime teha selliseid panuseid, et me konstantselt võidaksime ja seetõttu jäädagi oma kapitali kasvatama.

Lisaks on kogenenumal panustajal võimalus erinevate kihlveokontorite vahel valida ning panus teha selle kihlveokontori juures, kes pakub kõige paremaid koefitsiente soovitud mängule. Antud töös vaatleme, kuidas meie mudel töötab William Hilli kihlveokontori vastu (vaata alapeatükk 4.2)

Selleks, et võrrelda meie mudelit kihlveokontoriga vaatame oma kapitali käitumist ajas. Vaadeldavaid panustamise strateegiaid on kokku neli: Kelly, ½ Kelly, ¼ Kelly ja fikseeritud panusega strateegiad. (Panustamise strateegiaid kirjeldasime lähemalt alapeatükis 5.3.) Lisaks hakkame oma ennustusi tegema hooajal 2015-2016 pärast esimesi mänge, täpsemalt kui hooaeg on käinud ühe kuu. Seda eelkõige kahel põhjusel. Esiteks, et meil oleks informatsiooni värskelt esiliigast Premier League tõusnud meeskondade kohta ning teiseks, kuna hooaegade vahel toimub meeskondades tihti suuri muudatusi, siis laseme oma mudelil esimene kuu nende muutustega kohaneda. Näiteks ostavad paljud meeskonnad hooaegade vahel endale uusi mängijaid ning samuti võivad mõned olulised mängijad meeskonnast lahkuda. Seega on prognoosimise alguseks meil loodetavasti piisavalt informatsiooni, st meil on olemas eelteadmised eelmise hooaja *Premier League* ja esiliiga tulemustest ning käimasoleva hooaja esimese kuu tulemustest.

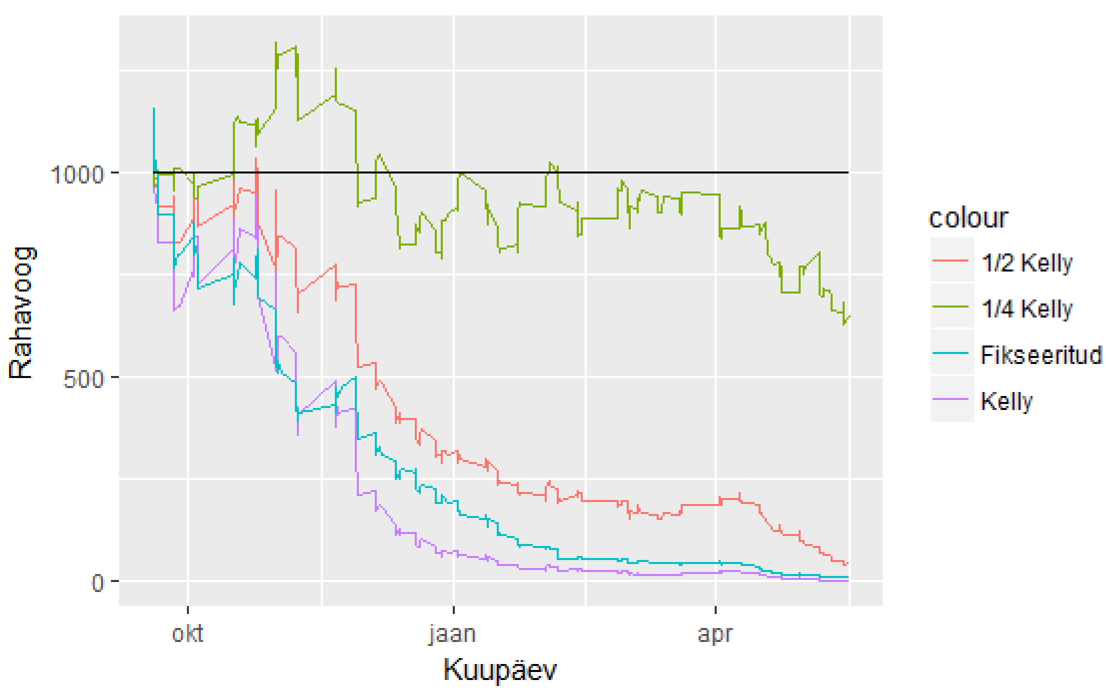
Viime läbi simulatsiooni, kus iga panustamisstrateegia algkapitaliks on 1000 rahaühikut. Kriteeriumiks on, et mängule tehakse panus, kui:

$$p * w - 1 = 0, \tag{7}$$

kus *p* on sündmuse *H*, *D* või *A* toimumise tõenäosus ja *w* on kihlveokontori poolt pandud koefitsient vastavale sündmusele. Panuse suurus sõltub vastavalt panustamise strateegiast. Näiteks fikseeritud panuse korral, tehakse mängule alati panus siis, kui kriteerium (7) on täidetud ning panuse suuruseks on alati 5% olemasolevast kapitalist, st esimest korda tehakse 50 rahaühiku suurune panus.

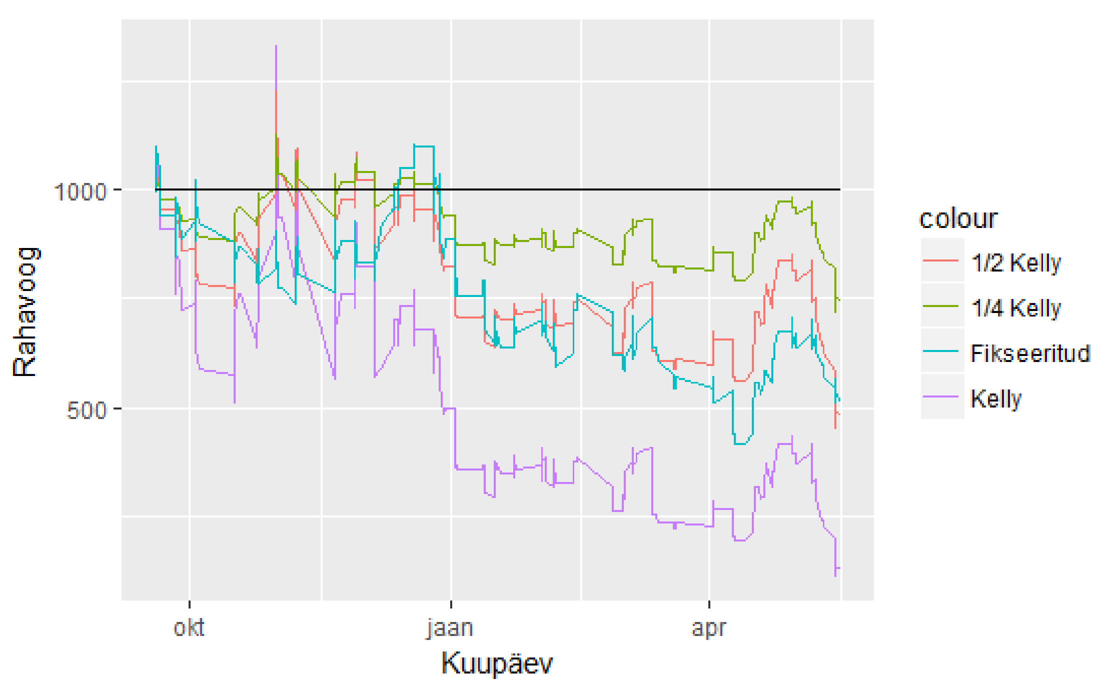
Vaatleme, kuidas kujuneksid tulemused meie parima mudeli korral (M2), kus me eeldame ka lisanduvat vormi parameetrit. Tulemused erinevate panustamisviiside kohta on kujutatud joonistel 2 ja 3, kus on vastavalt *fixed odds* ja *Over/Under* panustamised.

Nagu näha jooniselt 2, siis *fixed odds*i panustamise puhul oleksime hooaja lõppedes iga panustamise strateegiaga negatiivses bilansis. Kõige eeskujulikumalt läheks meil panustamise strateegiat ¼ Kelly kriteerium kasutades. Antud strateegia korral püsiksime esimene pool hooajast positiivses ning ka hiljem naaseksime hetkeks positiivsesse. Tipphetkel (31. oktoober) oleksime selle panustamisstrateegia korral kogunud 1316 rahaühikut. Samal ajal kui teiste strateegiate korral oleksime kiirelt langustrendis ja laostumine oleks saabumas üsna kiiresti.



Joonis 2. Rahavoo käitumine mudeli M2 korral fixed odds panustamises

Nagu näha joniselt 3, siis *Over/Under* panustamise korral oleksime hooaja lõppedes iga panustamise strateegiaga samuti negatiivses bilansis. Samas siin (nagu näha), esineksime selle panustamisviisi korral veidi paremini. Nimelt oleks nii ¼ Kelly kriteeriumit, kui ka fikseeritud panustega panustamise strateegiat kasutades hooaja esimeses pooles positiivses bilansis. Samuti läheks hooaja alguses küllalti edukalt ka ½ Kelly kriteeriumi panustamise strateegiaga. Küll aga kõige riskialtima Kelly kriteeriumi panustamise strateegia korral oleksime hetkeks kõige suurema summaga. Nimelt oleks meil 31. oktoobri seisuga 1332 rahaühikut. Antud kuupäev on ka teiste panustamisstrateegia korral tipphetk, kus rahavoog on maksimaalne. Tegemist oli Manchester City ja Norwichi vahelise kohtumisega, kus oli suur tõenäosus, et lüüakse vähemalt kolm väravat (tuginedes eelteadmisele, et omavahel kohtuvad liigatabeli tippu ja alumisse otsa kuuluvad meeskonnad ja tippmeeskond mängib koduväljakul, siis võiks oodata mängust palju väravaid) . Kuigi kihlveokontor võttis seda ka arvesse, pakkudes antud mängul „üle 2,5 värava löömise“ koefitsiendiks 1.42, siis tuvastas meie mudel, et selline sündmus on märksa tõenäolisem, kui kihlveokontori hinnang ning seetõttu tehti sinna mängule ka suuremad panused. Kuna Manchester City võitis mängu 2:1, mis tähendas, et löödi vähemalt kolm väravat (löödigi täpselt kolm), siis antud panus tasus end ära.



Joonis 3. Rahavoo käitumine mudeli M2 korral Over/Under panustamises

Veel on näha, et iga strateegia korral läheb meil alguses küllalti hästi, kus ollakse peaaegu iga strateegiaga positiivses bilansis. Seda eelkõige hooaja värskuse tõttu, kus erinevate meeskondade vormid tulevad ka kihlveokontoritele üllatusena. Teisalt võib see olla ka kihlveokontorite teadlik valik ja turundusnipp, läbi mille hooaja alguses kaotades saadakse endale juurde uusi potentsiaalseid kliente, et hiljem rohkem tagasi võita.

Vaatame lisaks, kuidas prognoosib meie mudel konkreetset mängu. Võtame vaatluse alla meeskonnad Chelsea ja Southamton. Antud mäng lõppes üllatavalt võõrsil mänginud meeskonna Southampton 3:1 võiduga. Eelmisel hooajal oli aga ühel Inglismaa nimekaimal meeskonnal Chelsea'l suuri probleeme oma potentsiaali kasutamisega. Antud vormid peegelduvad ka seisude prognoosides, kus meeskondasid peaaegu võrdseks võib pidada.

Tabel 9. Chelsea ja Southamptoni mängu prognoosimine.

	Southamton						
Chelsea	*0*	*1*	*2*	*3*	*4*	*...*	*10*
0	0.106	0.119	0.067	0.025	0.007	…	0
1	0.118	0.133	0.075	0.028	0.008	…	0
2	0.066	0.075	0.042	0.016	0.004	…	0
3	0.024	0.027	0.016	0.006	0.002	…	0
4	0.007	0.008	0.004	0.002	0.001	…	0
...	…	…	…	…	…	…	…
10	0	0	0	0	0	…	0

Kokkuvõte ja arutelu

Jalgpalli ja jalgpalli tulemustele panustamise populaarsuse kasvuga on järjest enam proovitud leida statistilist mudelit kihlveokontoritega võistlemiseks. Käesolevas magistritöös on antud ülevaade varasemalt tehtud mudelitele ning omalt poolt proovitud leida statistiline mudel kihlveokontoritega võistlemiseks. Mudeli headuse hindamiseks on kasutatud statistikut RPS. Mudeli eesmärgiks on leida selline panustamise strateegia, mida kasutades jääda konstantselt kasumisse. Antud töös vaadeldakse nelja panustamisstrateegiat – fikseeritud panust, Kelly kriteeriumit, ½ Kelly kriteeriumit ja ¼ Kelly kriteeriumit. Iga strateegia korral tehakse panus ainult sellistele mängudele, kus prognoos ja sellele prognoosile vastav koefitsiendi korrutis on üle lävendi väärtuse. Vaadeldakse kahte erinevat panustamisviisi – *fixed odds* ja *Over/Under*. Erinevate panustamisstrateegiate ja –viiside korral on läbi viidud simulatsioon, kus saadud parimat mudelit on võrreldud kihlveokontori William Hill omaga.

Tulemustest järeldub, et meil ei õnnestu pikas perspektiivis kihlveokontoreid võita. Meie mudel töötab mõningate strateegiate korral kihlveokontoritest paremini, aga seda lühema perioodi jooksul. Seega võiks mõelda, et antud mudelit saaks rakendada hooaja alguses, kui kihlveokontorite pole veel kõige täpsemad ja nagu ka tulemustest näha, siis on võimalik neid lühema perioodi jooksul võita.

Meie loodud mudelis on kasutatud ainult kvantitatiivseid suuruseid. Edasistes uurimustes võiks mudelisse lisada ka erinevaid kvalitatiivseid suuruseid. Näiteks, kui palju sõltub meeskond ühest konkreetsest mängijast. Kui näiteks hüpoteetiliselt Zlatan Ibrahimovic (väga oluline mängija tugevas meeskonnas Manchester United) peaks saama mängus Arsenali (samuti tugev meeskond) vastu mängu lõpus „toorutsemise“ eest punase kaardi, mis tooks kaasa kolmemängulise keelu. Kui nüüd Mancester United võidaks selle mängu, siis hinnatakse meeskonna tugevust varasemast kõrgemini. Oletame, et järgmised kolm mängu ei lähe Unitedil kõige paremini, mis tooks kaasa tugevuse vähenemise. Nüüd on aga järgmiseks mänguks tagasi meeskonna tipuründaja Zlatan Ibrahimovic ning meeskonna moraal peaks olema varasemast veelgi kõrgem, siis meie mudel seda ei suuda arvestada ja tal kulub mõned mängud aega, et

„järele jõuda“. Seega on sellel perioodil Manchester Unitedi tugevus meie mudelis vale ja võib kaasa tuua kehvasid panustamisettepanekuid.

Antud töö raames keskenduti enne mängu panuste tegemisele. Tänapäeval on aga üha enam populaarsust kogumas käimasolevate mängudele panuste tegemine. Edaspidi võiks mõelda mõne sellise mudeli koostamisele, mis aitab meil võistelda kihlveokontoritega käimasolevate mängude korral. Veebisaitidel on käimasolevate mängude panuste hulk märksa suurem võrreldes võimalike enne mängu panuste hulgaga. Seega oleks sellisel mudelil turgu märksa rohkem.

Kasutatud kirjandus

[1] bwin Group. (2009). Annual Report 2008. Vienna, Austria: bwin Interactive Entertainment AG.

[2] Constantinou, A. C., Fenton, N. E. (2012). Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. Journal of Quantitative Analysis in Sports: Vol. Iss. 1, Article 1.

[3] Dixon, M., Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. Applied Statistics, 46, 265-80.

[4] Epstein, E. (1969). A Scoring System for Probability Forecasts of Ranked Categories. Journal of Applied Meteorology, 8, 985-987.

[5] Global Betting and Gaming Consultants. (2001). 1st annual review of the global betting and gaming market, 2001. West Bromwich: Global Betting and Gaming Consultants.

[6] Gneiting, T., Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association, 102(477), 359-378.

[7] Hausch D. B., Lo V. S. Y., Ziemba W. T. (1994), Efficiency of Racetrack Betting Markets, Academic Press.

[8] Jose, V. R., Nau, R. F., Winkler, R. L. (2009). Sensitivity to Distance and Baseline Distribution in Forecast Evaluation. Management Science, 55, 582-590.

[9] Karlis, D., Ntzoufras, I. (2000). On modelling soccer data. Student, 229–244.

[10] Kelly J. L. Jr. (1956), A New Interpretation of Information Rate, The Bell System Technical Journal, 917-926.

[11] Maher, M. J. (1982). Modelling association football scores. Statististica Neerlandica, 36, 109– 118.

[12] Rue, H., Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. The Statistician, 49, Part 3, pp. 339-418.

Lisa

R kood

```
####################################################
########### import andmed ja bindime kokku ########
####################################################


# Seame oma wd
setwd("U:/My Documents/2017/Kool")
dta_A <- read.csv("1516.csv")
dta_B <- read.csv("1415.csv")
dta_C <- read.csv("1415esi.csv")

dta_A$season <- "15-16"
dta_B$season <- "14-15"
dta_C$season <- "14-15"

# Korrastame ja paneme ühtima

dta_B <- dta_B[-c(381),]
dta_B <- dta_B[-c(42:44)]
dta_C <- dta_C[-c(42:44)]
#na.omit(dta_B)

dta <- rbind(dta_A, dta_B, dta_C)
dta <- dta[-c(1313),]
dta$Date = as.Date(as.character(dta$Date), format="%d/%m/%y")

#################################################################
########## Defineerime kasulikud funktsioonid ###################
#################################################################

tau <- Vectorize(function(xx, yy, lambda, mu, rho){
  if (xx == 0 & yy == 0){return(1 - (lambda*mu*rho))
  } else if (xx == 0 & yy == 1){return(1 + (lambda*rho))
  } else if (xx == 1 & yy == 0){return(1 + (mu*rho))
  } else if (xx == 1 & yy == 1){return(1 - rho)
  } else {return(1)}
})

toepara <- function(y1, y2, lambda, mu, rho=0, weights=NULL){
  #rho=0, sõltumatus
  #y1: H väravad
  #y2: A väravad
  sum(weights*log(tau(y1, y2, lambda, mu, rho)) + log(dpois(y1, lambda)) +
log(dpois(y2, mu)))
}
```

```
#######################################################
######## Funktsioonid vormi hindamiseks ##############
#######################################################

pts_hooajal <- function(df, date1, team1, season1){
  wins <- dim(df[df$HomeTeam == team1 & df$season == season1 & df$Date < date1 &
df$FTR == "H",])[1] +
    dim(df[df$AwayTeam == team1 & df$season == season1 & df$Date < date1 & df$FTR ==
"A",])[1]
  draws <- dim(df[df$HomeTeam == team1 & df$season == season1 & df$Date < date1 &
df$FTR == "D",])[1] +
    dim(df[df$AwayTeam == team1 & df$season == season1 & df$Date < date1 & df$FTR ==
"D",])[1]
  points <- wins*3 + draws
  return(points)
}

mange_hooajal <- function(df, date1, team1, season1){
  mange <- dim(df[(df$HomeTeam == team1 | df$AwayTeam == team1) & df$season ==
season1 & df$Date < date1,])[1]
  return(mange)
}

pts_viimased <- function(df, date1, team1, season1, ngames){
  hgm <- df[df$HomeTeam == team1 & df$season == season1 & df$Date < date1,]
  agm <- df[df$AwayTeam == team1 & df$season == season1 & df$Date < date1,]
  hgm <- hgm[order(hgm$Date, decreasing = TRUE),]
  agm <- agm[order(agm$Date, decreasing = TRUE),]
  hgm <- head(hgm, n = ngames)
  agm <- head(agm, n = ngames)
  wins <- dim(hgm[hgm$FTR == "H",])[1] + dim(agm[agm$FTR == "A",])[1]
  draws <- dim(hgm[hgm$FTR == "D",])[1] + dim(agm[agm$FTR == "D",])[1]
  pts <- wins*3 + draws
  return(pts)
}

cnt_viimased <- function(df, date1, team1, season1, ngames){
  hgm <- df[df$HomeTeam == team1 & df$season == season1 & df$Date < date1,]
  agm <- df[df$AwayTeam == team1 & df$season == season1 & df$Date < date1,]
  hgm <- hgm[order(hgm$Date, decreasing = TRUE),]
  agm <- agm[order(agm$Date, decreasing = TRUE),]
  hgm <- head(hgm, n = ngames)
  agm <- head(agm, n = ngames)
  mange <- dim(hgm)[1] + dim(agm)[1]
  return(mange)
}

#############################################
##### Lisame vormi parameetrid tabelisse ####
#############################################

HLP <- NULL
ALP <- NULL
```

```
HLG <- NULL
ALG <- NULL
HFR <- NULL
AFR <- NULL
for(i in 1:dim(dta)[1]){
  hpts <- pts_viimased(dta, dta[i,]$Date, dta[i,]$HomeTeam, dta[i,]$season, 2)
  hcnt <- cnt_viimased(dta, dta[i,]$Date, dta[i,]$HomeTeam, dta[i,]$season, 2)
  apts <- pts_viimased(dta, dta[i,]$Date, dta[i,]$AwayTeam, dta[i,]$season, 2)
  acnt <- cnt_viimased(dta, dta[i,]$Date, dta[i,]$AwayTeam, dta[i,]$season, 2)
  hform <- NA
  aform <- NA
  if(hcnt > 0){hform <- hpts/hcnt}
  if(acnt > 0){aform <- apts/acnt}
  HLP <- c(HLP,hpts)
  ALP <- c(ALP,apts)
  HLG <- c(HLG,hcnt)
  ALG <- c(ALG,acnt)
  HFR <- c(HFR,hform)
  AFR <- c(AFR,aform)
}
dta <- cbind(dta, data.frame(H_Viimased = HLP, H_ViimasedMangud = HLG, H_Vorm = HFR,
A_Viimased = ALP, A_ViimasedMangud = ALG, A_Vorm = AFR))

Vorm_kesk <- mean(c(dta[!is.na(dta$A_Vorm),]$A_Vorm,dta[!is.na(dta$H_Vorm),]$H_Vorm))
dta[is.na(dta$H_Vorm),]$H_Vorm <- Vorm_kesk
dta[is.na(dta$A_Vorm),]$A_Vorm <- Vorm_kesk
##############################################################
####### Defineerime prognoosiks vajaminevad parameetrid ######
##############################################################

## Prognoosi algkpv
date_min = as.Date("2015-09-17")

## Kaalu parameeter
xi = 0.001

kaalufn <- function(dates, currentDate=Sys.Date(), xi=0){
  datediffs <- dates - as.Date(currentDate)
  datediffs <- as.numeric(datediffs *-1)
  w <- exp(-1*xi*datediffs)
  w[datediffs <= 0] <- 0
  return(w)
}

## Kaalude võrdlus
aeg <- sort(unique(dta$Date))
kaalud <- kaalufn(aeg, currentDate=date_min, xi=0.001)
kaalud2 <- kaalufn(aeg, currentDate=date_min, xi=0.002)
kaalud3 <- kaalufn(aeg, currentDate=date_min, xi=0.003)
kaalud0 <- kaalufn(aeg, currentDate=date_min, xi=0)
plot(aeg, kaalud,type="l", col="red", xlab="Kuupäev",ylab = "Kaalud")
lines(aeg, kaalud2, col="green", add=T)

legend("topright",legend=c(0.001,0.002,0.003,0),col=c("red","green","blue","black")
       ,lty=c(1,1))
```

```
##################################################
########## Prognoosimise funktsioon ##############
##################################################

arvuta_prognoos <- function(dta, vorm, slt, bet, date_min, xi){

  ### andmestik, kus hoian tulemusi ja rahavoo infot
  ennustused <- data.frame()

  ### sõltuvuse parameetri hindamiseks
  rhos <- NA

  ### kuupäevad

  kpv <- sort(unique(dta$Date))

  for(curkpv in as.list(kpv)){
    if(curkpv >= date_min){

      ### Võtame arvesse mineviku andmed ja arvutame kaalud
      dta_minevik <- dta[dta$Date < curkpv,]
      dta_minevik$date_diff <- as.numeric(curkpv - dta_minevik$Date)
      dta_minevik$kaal <- exp(-1*xi*dta_minevik$date_diff)
      dta_minevik$kaal[dta_minevik$date_diff <= 0] <- 0 # Tuleviku kaalud = 0

      ### Paneme loetavamale kujule
      andmestik <-
data.frame(Meeskond=as.factor(c(as.character(dta_minevik$HomeTeam),

as.character(dta_minevik$AwayTeam))),
                              Vastane=as.factor(c(as.character(dta_minevik$AwayTeam),

as.character(dta_minevik$HomeTeam))),
                              Varavad=c(dta_minevik$FTHG, dta_minevik$FTAG),
                              Eelis=c(rep(1, dim(dta_minevik)[1]), rep(0,
dim(dta_minevik)[1])),
                              Varavale=c(dta_minevik$HST, dta_minevik$AST),
                              Punaseid=c(dta_minevik$HR, dta_minevik$AR),
                              Vorm=c(dta_minevik$H_Vorm, dta_minevik$A_Vorm),
                              Kaal=c(dta_minevik$kaal,dta_minevik$kaal))

      if(vorm == "jah"){
        mudel <- glm(Varavad ~ Eelis + Meeskond + Vastane + Vorm, data=andmestik,
family=poisson(), weights=Kaal)
      }

      else{
        mudel <- glm(Varavad ~ Eelis + Meeskond + Vastane, data=andmestik,
family=poisson(), weights=Kaal)
      }

      if(slt == "jah"){
        hinnang <- fitted(mudel)
        H.hinnang <- hinnang[1:nrow(dta_minevik)]
        A.hinnang <- hinnang[(nrow(dta_minevik)+1):(nrow(dta_minevik)*2)]
```

```
        RhoFn <- function(par){
          rho <- par[1]
          toepara(dta_minevik$FTHG, dta_minevik$FTAG, H.hinnang, A.hinnang, rho,
dta_minevik$kaal)
        }

        tul <- optim(par=c(0.1), fn=RhoFn, control=list(fnscale=-1), method='BFGS')

        rhos <- c(rhos,tul$par)
      }

      dta_hetkel = dta[dta$Date == curkpv,]

      for(i in 1:dim(dta_hetkel)[1]){
        h_team = as.character(dta_hetkel[i,]$HomeTeam)
        a_team = as.character(dta_hetkel[i,]$AwayTeam)
        h_vorm = dta_hetkel[i,]$H_Vorm
        a_vorm = dta_hetkel[i,]$A_Vorm
#       h_punaseid = dta_hetkel[i,]$HR
#       a_punaseid = dta_hetkel[i,]$AR

        if(vorm == "jah"){
          lambda = predict(mudel, data.frame(Eelis=1, Meeskond=h_team,
Vastane=a_team, Vorm=h_vorm), type="response")
          mu = predict(mudel, data.frame(Eelis=0, Meeskond=a_team, Vastane=h_team,
Vorm=a_vorm), type="response")
        }
        else{
          lambda = predict(mudel, data.frame(Eelis=1, Meeskond=h_team,
Vastane=a_team, Varavale=h_varavale), type="response")
          mu = predict(mudel, data.frame(Eelis=0, Meeskond=a_team, Vastane=h_team,
Varavale=a_varavale), type="response")
        }

        maxgoal <- 10 # kasutan hiljem

        tn_matrix <- dpois(0:maxgoal, lambda) %*% t(dpois(0:maxgoal, mu))

        # Muudavad 0:0 ja 1:1 skoorib rohkem tõenäoliseks
        if(slt == "jah"){
          sc_matrix <- matrix(tau(c(0,1,0,1), c(0,0,1,1), lambda, mu, tul$par),
nrow=2)
          tn_matrix[1:2, 1:2] <- tn_matrix[1:2, 1:2] * sc_matrix
        }

        if(h_team == "Chelsea" & a_team == "Southampton"){
          abimaatriks1 <- tn_matrix
          #             abimaatriks2 <- sc_matrix
          abi1 = tul$par
          abi2 = lambda
          abi3 = mu
          mudel_abi = mudel
        }
```

```
####### 1X2 tõenäosused
if(bet == "1X2"){
  P_H <- sum(tn_matrix[lower.tri(tn_matrix)])
  P_D <- sum(diag(tn_matrix))
  P_A <- sum(tn_matrix[upper.tri(tn_matrix)])

  ####### Betting strategy and outcome

  ## w*p - 1 > 0
  vaartus_H <- dta_hetkel[i,]$WHH*P_H
  vaartus_D <- dta_hetkel[i,]$WHD*P_D
  vaartus_A <- dta_hetkel[i,]$WHA*P_A

  maxbet <- max(vaartus_H, vaartus_D, vaartus_A)

  betKelly <- 0
  betKelly_2 <- 0
  betKelly_4 <- 0
  betFixed <- 0
  ennustus <- NA

  # Paneme vaartus_th ka muutujaks (vaikimisi 1)
  if(maxbet == vaartus_H & maxbet > vaartus_th){
    ennustus <- "H"
    betKelly <- (vaartus_H - 1)/(dta_hetkel[i,]$WHH - 1)*stakeKelly
    betKelly_2 <- (vaartus_H - 1)/(dta_hetkel[i,]$WHH - 1)*stakeKelly_2/2
    betKelly_4 <- (vaartus_H - 1)/(dta_hetkel[i,]$WHH - 1)*stakeKelly_4/4
    betFixed <- stakeFixed*fixedbet_prop
  }
  else if(maxbet == vaartus_D & maxbet > vaartus_th){
    ennustus <- "D"
    betKelly <- (vaartus_D - 1)/(dta_hetkel[i,]$WHD - 1)*stakeKelly
    betKelly_2 <- (vaartus_D - 1)/(dta_hetkel[i,]$WHD - 1)*stakeKelly_2/2
    betKelly_4 <- (vaartus_H - 1)/(dta_hetkel[i,]$WHD - 1)*stakeKelly_4/4
    betFixed <- stakeFixed*fixedbet_prop
  }
  else if(maxbet == vaartus_A & maxbet > vaartus_th){
    ennustus <- "A"
    betKelly <- (vaartus_A - 1)/(dta_hetkel[i,]$WHA - 1)*stakeKelly
    betKelly_2 <- (vaartus_A - 1)/(dta_hetkel[i,]$WHA - 1)*stakeKelly_2/2
    betKelly_4 <- (vaartus_H - 1)/(dta_hetkel[i,]$WHA - 1)*stakeKelly_4/4
    betFixed <- stakeFixed*fixedbet_prop
  }

  voit <- 0
  if(!is.na(ennustus) & ennustus == dta_hetkel[i,]$FTR){
    voit <- 1
    stakeKelly <- stakeKelly + betKelly
    stakeKelly_2 <- stakeKelly_2 + betKelly_2
    stakeKelly_4 <- stakeKelly_4 + betKelly_4
    stakeFixed <- stakeFixed + betFixed
  }
  if(!is.na(ennustus) & ennustus != dta_hetkel[i,]$FTR){
    voit <- -1
    stakeKelly <- stakeKelly - betKelly
```

```
      stakeKelly_2 <- stakeKelly_2 - betKelly_2
      stakeKelly_4 <- stakeKelly_4 + betKelly_4
      stakeFixed <- stakeFixed - betFixed
    }


    ### Paneme kokku
    ennustused <- rbind(ennustused, data.frame(Kpv = as.Date(curkpv), H_Team =
h_team, A_Team = a_team,
                                              OddH = dta_hetkel[i,]$B365H,
OddD = dta_hetkel[i,]$B365D, OddA = dta_hetkel[i,]$B365A,
                                              P_H = P_H, P_D = P_D, P_A = P_A,
                                              vaartus_H = vaartus_H *100,
                                              vaartus_D = vaartus_D *100,
                                              vaartus_A = vaartus_A *100,
                                              Ennustus = ennustus,
                                              betKelly = betKelly,
                                              betKelly_2 = betKelly_2,
                                              betKelly_4 = betKelly_4,
                                              betFixed = betFixed,
                                              stakeKelly = stakeKelly,
                                              stakeKelly_2 = stakeKelly_2,
                                              stakeKelly_4 = stakeKelly_4,
                                              stakeFixed = stakeFixed,
                                              H_Varavaid =
as.numeric(dta_hetkel[i,]$FTHG), A_Varavaid = as.numeric(dta_hetkel[i,]$FTAG),
                                              Tulemus = dta_hetkel[i,]$FTR))
  }

  ####### VäravadKokku tõenäosused
  if(bet == "OU"){
    P_O <- 0
    P_U <- 0
    for(ii in 1:dim(tn_matrix)[1]){
      for(jj in 1:dim(tn_matrix)[2]){
        if(ii+jj-2 > 2.5){
          P_O <- P_O + tn_matrix[ii,jj]
        }
        else{
          P_U <- P_U + tn_matrix[ii,jj]
        }
      }
    }


    ####### Panustamise strateegiad ja tulemused

    vaartus_O <- dta_hetkel[i,]$BbAv.2.5*P_O
    vaartus_U <- dta_hetkel[i,]$BbAv.2.5.1*P_U

    print(dta_hetkel[i,]$BbAv.2.5)

    maxbet <- max(vaartus_O, vaartus_U)

    betKelly <- 0
```

```
          betKelly_2 <- 0
          betKelly_4 <- 0
          betFixed <- 0
          ennustus <- NA

          if(maxbet == vaartus_O & maxbet > vaartus_th){
            ennustus <- "Over"
            betKelly <- (vaartus_O - 1)/(dta_hetkel[i,]$BbAv.2.5 - 1)*stakeKelly
            betKelly_2 <- (vaartus_O - 1)/(dta_hetkel[i,]$BbAv.2.5 -
1)*stakeKelly_2/2
            betKelly_4 <- (vaartus_O - 1)/(dta_hetkel[i,]$BbAv.2.5 -
1)*stakeKelly_4/4
            betFixed <- stakeFixed*fixedbet_prop
          }
          else if(maxbet == vaartus_U & maxbet > vaartus_th){
            ennustus <- "Under"
            betKelly <- (vaartus_U - 1)/(dta_hetkel[i,]$BbAv.2.5.1 - 1)*stakeKelly
            betKelly_2 <- (vaartus_U - 1)/(dta_hetkel[i,]$BbAv.2.5.1 -
1)*stakeKelly_2/2
            betKelly_4 <- (vaartus_U - 1)/(dta_hetkel[i,]$BbAv.2.5.1 -
1)*stakeKelly_4/4
            betFixed <- stakeFixed*fixedbet_prop
          }

          voit <- 0
          if(!is.na(ennustus) & (ennustus == "Over" & dta_hetkel[i,]$FTHG +
dta_hetkel[i,]$FTAG > 2.5
                                 | ennustus == "Under" & dta_hetkel[i,]$FTHG +
dta_hetkel[i,]$FTAG < 2.5)){
            voit <- 1
            stakeKelly <- stakeKelly + betKelly
            stakeKelly_2 <- stakeKelly_2 + betKelly_2
            stakeKelly_4 <- stakeKelly_4 + betKelly_4
            stakeFixed <- stakeFixed + betFixed
          }
          if(!is.na(ennustus) & (ennustus == "Over" & dta_hetkel[i,]$FTHG +
dta_hetkel[i,]$FTAG < 2.5
                                 | ennustus == "Under" & dta_hetkel[i,]$FTHG +
dta_hetkel[i,]$FTAG > 2.5)){
            voit <- -1
            stakeKelly <- stakeKelly - betKelly
            stakeKelly_2 <- stakeKelly_2 - betKelly_2
            stakeKelly_4 <- stakeKelly_4 - betKelly_4
            stakeFixed <- stakeFixed - betFixed
          }


          ### Paneme kokku
          ennustused <- rbind(ennustused, data.frame(Kpv = as.Date(curkpv), H_Team =
h_team, A_Team = a_team,
                                                     OddO = dta_hetkel[i,]$BbAv.2.5,
OddU = dta_hetkel[i,]$BbAv.2.5.1,
                                                     P_O = P_O, P_U = P_U,
                                                     vaartus_O = vaartus_O *100,
                                                     vaartus_U = vaartus_U *100,
```

```
                                                    Ennustus = ennustus,
                                                    betKelly = betKelly,
                                                    betKelly_2 = betKelly_2,
                                                    betKelly_4 = betKelly_4,
                                                    betFixed = betFixed,
                                                    stakeKelly = stakeKelly,
                                                    stakeKelly_2 = stakeKelly_2,
                                                    stakeKelly_4 = stakeKelly_4,
                                                    stakeFixed = stakeFixed,
                                                    Varavaid =
as.numeric(dta_hetkel[i,]$FTHG) + as.numeric(dta_hetkel[i,]$FTAG),
                                                    Tulemus = dta_hetkel[i,]$FTR))
        }

      }
    }
  }

  tulemus <- NA
  tulemus$ennustus <- ennustused
  tulemus$abimaatriks1 <- abimaatriks1
  tulemus$mudel_abi <- mudel_abi

  return (tulemus)

}

#####################################
#### Panustamise strateegia #########
#####################################

vaartus_th = 1.0
fixedbet_prop = 0.05
stakeKelly = 1000
stakeKelly_2 = 1000
stakeKelly_4 = 1000
stakeFixed = 1000

minu_mudel <- arvuta_prognoos(dta = dta, vorm = "jah", slt = "jah", bet = "1X2",
date_min, xi)

plot(minu_mudel$ennustus$stakeKelly, type="l", col="red", xlab = "Mängu nr",ylab =
"Rahavoog")
lines(minu_mudel$ennustus$stakeKelly_2, col="blue", add=T)

library(ggplot2)

ggplot(minu_mudel$ennustus, aes(minu_mudel$ennustus$Kpv)) +
  geom_line(aes(y=stakeKelly, colour="Kelly")) +
  geom_line(aes(y=stakeKelly_2, colour="1/2 Kelly")) +
  geom_line(aes(y=stakeKelly_4, colour="1/4 Kelly")) +
  geom_line(aes(y=stakeFixed, colour="Fikseeritud")) +
  geom_line(aes(y=1000)) +
  labs(x="Kuupäev",y="Rahavoog")
```

```
minu_mudel$ennustus$Tulemus <- factor(minu_mudel$ennustus$Tulemus,
levels=c("A","D","H"))
sum(minu_mudel$ennustus$Ennustus == minu_mudel$ennustus$Tulemus, na.rm = TRUE) # 144

minu_mudel_2 <- arvuta_prognoos(dta = dta, vorm = "jah", slt = "jah", bet = "OU",
date_min, xi)

minu_mudel_2$abimaatriks1

ggplot(minu_mudel_2$ennustus, aes(Kpv)) +
  geom_line(aes(y=stakeKelly, colour="Kelly")) +
  geom_line(aes(y=stakeKelly_2, colour="1/2 Kelly")) +
  geom_line(aes(y=stakeKelly_4, colour="1/4 Kelly")) +
  geom_line(aes(y=stakeFixed, colour="Fikseeritud")) +
  geom_line(aes(y=1000)) +
  labs(x="Kuupäev",y="Rahavoog")

RPS_fn <- function(ennustused, vaadeldud){
  ncat <- ncol(ennustused)
  npred <- nrow(ennustused)

  rps <- numeric(npred)

  for (rr in 1:npred){
    vaadeldud_vec <- rep(0, ncat)
    vaadeldud_vec[vaadeldud[rr]] <- 1
    cumsum <- 0
    for (i in 1:ncat){
      cumsum <- cumsum + (sum(ennustused[rr,1:i]) - sum(vaadeldud_vec[1:i]))^2
    }
    rps[rr] <- (1/(ncat-1))*cumsum
  }
  return(rps)
}

# Võrdleme RPS-e

ennustus <- minu_mudel$ennustus[,7:9]
tegelik <- minu_mudel$ennustus$Tulemus
tegelik <- factor(tegelik, levels=c("H","D","A"))
RPS <- RPS_fn(ennustus,tegelik)
mean(RPS)

####################################
### Võrdleme erinevaid bookie'sid ##
####################################

Bet <- minu_mudel$ennustus[,4:6]**(-1)
mean(RPS_fn(Bet,tegelik))

sd(RPS_fn(Bet,tegelik))

median(RPS_fn(Bet,tegelik))
```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, _______________________Gertis
Aru____, (autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Statistiline mudel kihlveokontoritega konkureerimiseks *English Premier League* näitel"____
____ __,
(lõputöö pealkiri)
mille juhendaja on _____prof. Kalev Pärna____,
(juhendaja nimi)

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **15.05.2017**