

ZÁPADOČESKÁ UNIVERZITA V PLZNI

FAKULTA APLIKOVANÝCH VĚD

KATEDRA MATEMATIKY

## Diplomová práce

Modelování a odhadování výsledků tenisových zápasů

Plzeň, 2018

Bc. Jan Špaček

## **Zadání**

Místo této strany bude vloženo zadání.

## **Prohlášení**

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím literatury a pramenů uvedených v seznamu.

## **Poděkování**

Tímto děkuji vedoucímu mé diplomové práce Ing. Patrice Markovi Ph.D. za cenné rady a čas, který mně věnoval při konzultacích.

V Plzni dne 2. 5. 2018

.....

Jan Špaček

## **Abstrakt**

Tato diplomová práce se zabývá odhadováním výsledků tenisových dvouher v kategorii mužů i žen pomocí modelů založených na Elo rating systému.

Na začátku jsou uvedeny vybrané modely pro odhad výsledků sportovních a tenisových zápasů. Poté se práce věnuje historii Elo rating systému. Je zde popsán způsob, jak vypočítat pravděpodobnost vítězství hráče v zápase na základě hodnot Elo ratingů obou soupeřů. Poté jsou představeny různé metody pro odhad úvodního Elo ratingu hráče i způsoby přepočítávání Elo ratingu hráče po odehraném zápase. V další části jsou optimalizovány parametry Elo modelů. Na závěr dojde k ověření Elo modelů na základě dat vyplývajících ze sázení na modelem vybrané tenisové zápasy.

**Klíčová slova:** Elo rating, Odhad tenisových výsledků, Sázení

## **Abstract**

This diploma thesis focuses on estimating and modelling results of men's and women's singles in tennis. The models are based on the Elo rating system.

After the introduction of selected models for estimating results of sports and tennis matches, the thesis continues with a brief history of the Elo rating system. The theoretical part describes how to calculate a probability of the victory of a player in a match, based on the Elo ratings of both players. It is followed by various methods for estimating the initial Elo rating and methods for recalculation of the Elo rating of the player after the match. The practical part brings optimized parameters of the Elo models and validation of the Elo models done with the use of the data reflecting betting on tennis matches chosen by the models.

**Key words:** Elo rating, Estimating of sport results, Betting

# Obsah

1	Úvod .....	1
2	Přehled vybraných sportovních a tenisových modelů .....	2
3	Popis průběhu tenisových utkání a použitých dat .....	4
3.1	Systém tenisových zápasů .....	4
3.2	Údaje v databázi .....	5
3.3	Přehled databáze .....	6
4	Elo rating systém .....	7
4.1	Stav před zavedením Elo rating systému.....	7
4.2	Elo rating jako normální rozdělení.....	7
4.3	Elo rating jako logistické rozdělení .....	9
5	Odhad Elo ratingu.....	12
5.1	Elo pro nové hráče .....	12
5.1.1	Metoda konstantního Elo ratingu .....	12
5.1.2	Metoda kvantilu .....	12
5.1.3	Metoda prozatímního ratingu.....	14
5.1.4	Zhodnocení metod .....	14
5.2	Přepoččet Elo ratingu .....	14
5.2.1	Stejný koeficient rozvoje .....	15
5.2.2	Více konstantních koeficientů rozvoje .....	16
5.2.3	Koeficient rozvoje jako funkce .....	16
6	Kritéria určující kvalitu modelu .....	18
6.1	Přesnost .....	18
6.2	Logaritmická ztrátová funkce .....	18
6.3	Kalibrace .....	19
7	Elo modely.....	20
7.1	Elo model 1 .....	20
7.1.1	Optimum muži.....	21
7.1.2	Optimum ženy .....	24
7.2	Elo model 2 .....	26

7.2.1	Optimalizace muži .....	27
7.2.2	Optimalizace ženy .....	28
7.3	Elo model 3 .....	28
7.3.1	Optimalizace muži .....	29
7.3.2	Optimalizace ženy .....	31
7.4	Srovnání .....	33
8	Vývoj Elo ratingu v čase .....	37
8.1	Muži .....	37
8.2	Ženy .....	40
9	Ověření modelu pomocí sázení .....	42
9.1	Základní pojmy .....	42
9.2	Výběr zápasů .....	44
9.3	Strategie sázení .....	44
9.3.1	Konstantní vklad .....	44
9.3.2	Procento z banku .....	44
9.4	Elo model 3 pro sázení .....	45
9.5	Program k ověření modelu .....	45
9.6	Sázení muži .....	46
9.6.1	Sázení konstantní vklad .....	46
9.6.2	Sázení procento z banku .....	47
9.7	Sázení ženy .....	48
9.7.1	Sázení konstantní vklad .....	48
9.7.2	Sázení procento z banku .....	50
9.8	Jednoduché modely .....	52
9.8.1	Princip modelů .....	52
9.8.2	Muži .....	53
9.8.3	Ženy .....	55
9.9	Shrnutí .....	57
10	Závěr .....	58
11	Seznam použité literatury a zdroje dat .....	60

## Seznam tabulek

Tabulka 1: Základní přehled databáze .....	6
Tabulka 2: Třídy USCF.....	7
Tabulka 3: Pravděpodobnost výhry hráče A v závislosti na rozdílu Elo ratingu.....	11
Tabulka 4: Koeficient rozvoje $K$ FIDE a USCF .....	16
Tabulka 5: Elo model 2 optimalizace muži .....	27
Tabulka 6: Elo model 2 optimalizace ženy .....	28
Tabulka 7: Kritéria určující kvalitu modelu Elo modely.....	33
Tabulka 8: Kritéria určující kvalitu modelu články .....	35
Tabulka 9: Kritéria určující kvalitu modelu sázková kancelář .....	36
Tabulka 10: Elo rating Top 15 muži .....	39
Tabulka 11: Elo rating Top 15 ženy .....	41
Tabulka 12: Ideální rozložení sázek.....	43
Tabulka 13: Jiné rozložení sázek.....	43
Tabulka 14: Porovnání koeficientu sázení R konstantní vklad muži .....	47
Tabulka 15: Zisk v závislosti na R a procentech muži.....	48
Tabulka 16: Porovnání koeficientu sázení R konstantní vklad ženy .....	50
Tabulka 17: Zisk v závislosti na R a procentech ženy .....	52
Tabulka 18: Porovnání Elo modelu s jednoduchými modely muži .....	53
Tabulka 19: Porovnání Elo modelu s jednoduchými modely ženy.....	55

## Seznam obrázků

Obrázek 1: Ukázka z databáze: sloupce A – K v excelovském souboru .....	5
Obrázek 2: Ukázka z databáze: sloupce J – W v excelovském souboru.....	5
Obrázek 3: Ukázka z databáze: sloupce X – AJ v excelovském souboru .....	5
Obrázek 4: Výkonnost hráče .....	8
Obrázek 5: Pravděpodobnost výhry hráče A.....	9
Obrázek 6: Koeficient rozvoje $K$ funkce.....	17
Obrázek 7: Elo model 1 ztrátová funkce muži.....	22
Obrázek 8: Elo model 1 výřez ztrátové funkce muži .....	22
Obrázek 9: Elo model 1 ztrátová funkce pro $m=50$ muži .....	23
Obrázek 10: Elo model 1 ztrátová funkce pro $K=18$ muži .....	24



Obrázek 11: Elo model 1 ztrátová funkce ženy .....	24
Obrázek 12: Elo model 1 výřez ztrátové funkce ženy .....	25
Obrázek 13: Elo model 1 ztrátová funkce pro $m=60$ ženy .....	26
Obrázek 14: Elo model 1 ztrátová funkce pro $K=23$ ženy .....	26
Obrázek 15: Elo model 3 ztrátová funkce muži.....	29
Obrázek 16: Elo model 3 výřez ztrátové funkce muži .....	30
Obrázek 17: Elo model 3 ztrátová funkce pro $l=23$ muži .....	31
Obrázek 18: Elo model 3 ztrátová funkce ženy .....	31
Obrázek 19: Elo model 3 výřez ztrátové funkce ženy .....	32
Obrázek 20: Elo model 3 ztrátová funkce pro $l=27$ ženy .....	33
Obrázek 21: Elo rating muži světoví hráči .....	37
Obrázek 22: Elo rating muži čeští hráči .....	38
Obrázek 23: Elo rating ženy světové hráčky.....	40
Obrázek 24: Elo rating ženy české hráčky .....	40
Obrázek 25: Velikost zisku v závislosti na R konstantní vklad muži .....	46
Obrázek 26: Velikost banku v závislosti na R procento z banku muži.....	47
Obrázek 27: Velikost zisku v závislosti na R konstantní vklad ženy .....	49
Obrázek 28: Vývoj zisku konstantní vklad ženy.....	50
Obrázek 29: Velikost banku v závislosti na R procento z banku ženy .....	51
Obrázek 30: Vývoj banku při strategii procento z banku ženy.....	51
Obrázek 31: Histogram model Favorit muži.....	54
Obrázek 32: Histogram model Outsider muži .....	54
Obrázek 33: Histogram model Náhodný muži .....	55
Obrázek 34: Histogram model Favorit ženy .....	56
Obrázek 35: Histogram model Outsider ženy .....	56
Obrázek 36: Histogram model Náhodný ženy .....	57

# 1 Úvod

Tato diplomová práce navazuje na bakalářskou práci Špaček (2015), ve které byly odhadovány výsledky fotbalových utkání v české, anglické, italské a španělské lize.

Tato práce se zabývá modely pro odhad výsledků tenisových utkání. Odhadují se výsledky dvouher mužů a žen. Data ze zápasů mužů jsou od roku 2005 do roku 2016 a data ze zápasů žen jsou od roku 2007 do roku 2016.

Inspirací pro diplomovou práci je článek Kovalchik (2016), který se jmenuje *Searching for the GOAT of tennis win prediction*<sup>1</sup>. V tomto článku jsou představeny a mezi sebou porovnány různé modely pro odhad výsledků tenisových utkání. V článku jsou uvedeny i modely založené na Elo rating systému, které v porovnání s ostatními modely odhadovaly výsledky velmi dobře. Proto jsou v této diplomové práci Elo modely zkoumány a poté použity k odhadu výsledků tenisových utkání.

Ve druhé kapitole jsou uvedeny různé modely pro odhadování výsledků sportovních a především tenisových utkání. Jsou zde představeny modely z článku Kovalchik (2016).

Třetí kapitola se zabývá systémem tenisových utkání. Dále je zde popsána databáze tenisových utkání, která je v další části práce využita pro optimalizování parametrů a sázení.

Čtvrtá kapitola se věnuje vývoji Elo rating systému. Je zde popsán Elo rating systém jako normální rozdělení i jako logistické rozdělení. Dále je tu uveden návod, jak odhadnout pravděpodobnosti výsledků utkání na základě Elo ratingu.

V páté kapitole je popsán postup při odhadu úvodního Elo ratingu hráče a dále postup při přepočítávání Elo ratingu v průběhu času.

V šesté kapitole jsou uvedena 3 kritéria určující kvalitu modelu. Konkrétně to jsou logaritmická ztrátová funkce, přesnost a kalibrace.

V sedmé kapitole probíhá optimalizování parametrů tří Elo modelů. Optimalizované Elo modely jsou zde porovnány s modely sázkových kanceláří a s modely z článku Kovalchik (2016).

V osmé kapitole je zobrazen vývoj Elo ratingu vybraných hráčů v čase.

Devátá kapitola se věnuje sázení. Nejdříve jsou zde popsány základní sázkařské pojmy a strategie. Poté dochází k samotnému ověření modelu pomocí sázení proti sázkové kanceláři. Sází se podle Elo modelu na zápasy mužů a žen v letech 2015 a 2016.

Desátá kapitola se věnuje závěrečnému shrnutí výsledků a zhodnocení práce.

---

<sup>1</sup> Hledání nejlepšího modelu pro předpovídání výsledků tenisových utkání

## 2 Přehled vybraných sportovních a tenisových modelů

Ve 20. století došlo k velkému rozvoji sportu. O sport se začalo zajímat velké množství lidí. Majitelé týmů, trenéři i fanoušci chtěli vědět, jaká je šance, že jejich tým vyhraje. Totéž chtěli znát i sázkové kanceláře. Proto matematici a statistici začali modelovat a odhadovat výsledky sportovních zápasů. Z počátku se zabývali především modelováním a odhadováním výsledků fotbalových utkání. Díky rozvoji výpočetní techniky na začátku 21. století bylo možné rychle zpracovávat a analyzovat velké množství dat. V té době se objevilo mnoho modelů pro odhadování výsledků utkání v různých sportech. Tato kapitola představí několik vybraných modelů pro odhadování výsledků sportovních a tenisových utkání.

Maher (1982) ve svém článku *Modelling association football scores* popsal čtyři modely k odhadování fotbalových výsledků. K modelování výsledků utkání je v těchto modelech využíváno Poissonovo rozdělení pravděpodobnosti. K ověření modelů byla použita data z anglické fotbalové ligy. Na tuto práci navázali Dixon a Coles (1997). V jejich modelu bylo také využito Poissonovo rozdělení pravděpodobnosti. Navíc zde byla použita korekce modelu pro nejčastější fotbalové výsledky 0:0, 1:0, 0:1 a 1:1. Tyto modely použil také Špaček (2015) ve své bakalářské práci, kde byly odhadovány výsledky utkání v české, anglické, italské a španělské fotbalové lize. Modely založené na Poissonovo rozdělení pravděpodobnosti lze využít i v jiných sportech, například v hokeji, což popsali Marek, Šedivá a Toupal (2014). V tomto případě byly výsledky modelu ověřeny na české nejvyšší hokejové soutěži. Karlis a Ntzoufras (2003) využili Poissonovo rozdělení k odhadování výsledků vodního póla. Výsledky modelu byly ověřeny na zápasech mistrovství Evropy.

Tato diplomová práce se zabývá odhadem výsledků tenisových utkání. Inspirací pro práci je Kovalchik (2016), která ve svém článku *Searching for the GOAT of tennis win prediction* porovnávala několik tenisových modelů. Modely pro odhadování výsledků tenisových utkání jsou rozděleny do čtyř základních skupin. Do první skupiny patří modely založené na regresi. Druhou skupinu tvoří modely, ve kterých je každý bod v zápase považován za nezávislou a stejně rozdělenou náhodnou veličinu (iid). Do třetí skupiny náleží modely založené na párovém porovnávání. Ve čtvrté skupině je model založený na vypsaných kurzech několika sázkových kanceláří.

Do první skupiny modelů založených na regresi patří například logitový model, který popsali Klaassen a Magnus (2003). Tento model používali pro odhad výsledků zápasů na turnaji Wimbledon. Základem modelu je údaj o aktuálním umístění hráčů na žebříčku ATP či WTA. Pravděpodobnosti vítězství hráčů před začátkem utkání se odhadují pomocí logitové regrese. Ve druhé části tohoto článku je popsán způsob odhadu pravděpodobnosti výsledku zápasu v již rozehraném utkání. Obě části článku autoři použili pro svůj počítačový program Tennisprob, který odhaduje výsledky tenisových zápasů před zápasem i v jeho průběhu.

K odhadu pravděpodobnosti vítězství hráče v zápase lze aplikovat také probitovou regresi, kterou použili ve svém modelu Gilsdorf a Sukhatme (2008). K odhadu výsledků využívá tento model aktuální postavení hráčů na žebříčku a vydělané peníze<sup>2</sup>. Autoři svůj model testovali na zápasech mužů v sezóně 2001 a poté podpořili již předtím vyslovenou hypotézu, že zvyšující se rozdíl v peníze hráčů má statisticky významný vliv na pravděpodobnost výhry v zápase.

Newton a Keller (2005) vytvořili základní model pro odhadování tenisových výsledků založený na nezávislosti a stejném rozdělení (iid) jednotlivých bodů v jednom zápase. V tomto modelu se odhaduje pravděpodobnost vítězství hráče v gamu, setu, zápase a následně i v celém turnaji. Model byl použit na turnaje Wimbledon a US Open v roce 2002.

Další možnost, jak odhadovat výsledky tenisových utkání, je použitím modelu Bradley a Terry (1952) pro párové porovnávání, jež popsali ve svém článku McHale a Morton (2011).

Tenisové výsledky lze odhadovat i na základě modelů využívajících tzv. Elo rating systém. Ten se již dlouhou dobu používá v šachu. Autorem tohoto systému je Arpad Elo (1978), který ho popsal ve své knize *The Rating of Chess Players, Past and Present*. V dnešní době se Elo rating systém a jeho modifikace používají k odhadování výsledků zápasů i v jiných sportech, například Hvattum a Arntzen (2010) ho použili ve fotbale. Modely FiveThirtyEight použité v Kovalchik (2016) jsou též založené na Elo rating systému. V porovnání s modely založenými na regresi či iid jednotlivých bodů si modely založené na Elo rating systému vedly lépe. Proto se tato práce zaměří na odhadování výsledků tenisových zápasů pomocí modelů založených na Elo rating systému.

---

<sup>2</sup> Prize money – peněžní odměna za vyhrané zápasy

### 3 Popis průběhu tenisových utkání a použitých dat

V této diplomové práci se používají data z tenisových utkání. Zdrojem dat je internetová stránka tennis-data.co.uk (2018). Data ze zápasů mužů jsou od roku 2005 do roku 2016 a data ze zápasů žen jsou od roku 2007 do roku 2016. V databázi je celkem 32 510 zápasů mužů a 24 614 zápasů žen.

Databáze mužů a žen se nachází v příložených excelovských souborech *DBMuži.xlsx* a *DBŽeny.xlsx*. V databázi byla provedena korekce některých jmen, která byla napsána různými způsoby. Například se jedná o „Chela J.I.“ a „Chela J.“. Pokud by nebyla provedena tato korekce, vytvořený program by bral tato dvě jména jako dva různé hráče. Dále byla provedena změna u jmen žen, které se vdaly. Například u Ivety Benešové později Melzerové bylo zachováno jméno „Benesova I.“, jinak by opět program bral tato dvě jména jako dvě různé tenistky. Zde však nelze vyloučit to, že u některých vdaných, méně známých tenistek tato změna jmen neproběhla. V tom případě bude vdaná tenistka považována za novou hráčku a bude muset opět odehrát dostatečný počet zápasů, než výsledky jejích zápasů budou odhadovány. Z tohoto důvodu to nebude mít výraznější vliv na výsledky modelu jako celku.

#### 3.1 Systém tenisových zápasů

V zápase hrají proti sobě dva hráči (dvouhra) nebo dvě dvojice hráčů (čtyřhra). Tato práce se zabývá pouze zápasy dvouher.

Každý zápas se skládá z tzv. setů. V utkání žen vyhraje tenistka, která jako první vyhraje 2 sety. V zápasech mužů vyhraje hráč, který jako první vyhraje 2 nebo 3 sety. Počet vyhraných setů nutných k vítězství v zápase závisí na jednotlivých turnajích. Na většině menších turnajů stačí vyhrát 2 sety, ale například na Grand Slamech musí tenista vyhrát 3 sety.

Každý set se skládá z jednotlivých gamů. Soupeři se po každém gamu střídají v podání. Set vyhraje ten hráč, který vyhraje jako první 6 gamů. Musí být však splněna podmínka, že vítězství je nejméně o 2 gamy. V případě, že je stav 6:5, mohou nastat 2 možnosti. Buď další game vyhraje tenista, který vede, a tím vyhraje celý set 7:5, anebo vyhraje jeho soupeř. Stav bude 6:6 a většinou se hraje tzv. tiebreak. Tiebreak se hraje do sedmi bodů a opět musí vítěz vyhrát o alespoň 2 body. Pokud toto není splněno, hraje se do té doby, než někdo zvítězí o 2 body. Vítěz tiebreaku vyhraje celý set 7:6. V některých případech, například v rozhodujících setech Grand Slamů, se hraje bez tiebreaku. Za stavu 6:6 set nekončí, ale pokračuje se až do doby, kdy jeden z tenistů vyhrává o 2 gamy, a tím vyhraje celý set.

Každý game se skládá z jednotlivých výměn. Pokud hráč vyhraje výměnu, získá skóre 15, následně 30, poté 40 a pak už hráč vyhraje celý game. Pokud je však skóre 40:40, nastává tzv. shoda. Pokud hráč vyhraje další výměnu, má tzv. výhodu A:40. Pokud

stejný tenista vyhraje i tu další, tak vyhrál celý game. Jestliže vyhraje soupeř, nastává opět shoda 40:40 a proces se opakuje, dokud jeden z tenistů nevyhraje game.

## 3.2 Údaje v databázi

Na obrázcích 1, 2 a 3 je ukázka z databáze.

WTA	Location	Tourname	Date	Tier	Court	Surface	Round	Best of	Winner	Loser
6	Melbourn	Australian	28.1.2015	Grand Sla	Outdoor	Hard	Quarterfir	3	Keys M.	Williams V.
6	Melbourn	Australian	28.1.2015	Grand Sla	Outdoor	Hard	Quarterfir	3	Williams S.	Cibulkova D.
6	Melbourn	Australian	29.1.2015	Grand Sla	Outdoor	Hard	Semifinals	3	Sharapova M.	Makarova E.
6	Melbourn	Australian	29.1.2015	Grand Sla	Outdoor	Hard	Semifinals	3	Williams S.	Keys M.
6	Melbourn	Australian	31.1.2015	Grand Sla	Outdoor	Hard	The Final	3	Williams S.	Sharapova M.

Obrázek 1: Ukázka z databáze: sloupce A – K v excelovském souboru

Winner	Loser	WRank	LRank	WPts	LPts	W1	L1	W2	L2	W3	L3	Wsets	Lsets
Keys M.	Williams V.	35	18	1390	2370	6	3	4	6	6	4	2	1
Williams S.	Cibulkova D.	1	10	8016	3007	6	2	6	2			2	0
Sharapova M.	Makarova E.	2	11	7335	2970	6	3	6	2			2	0
Williams S.	Keys M.	1	35	8016	1390	7	6	6	2			2	0
Williams S.	Sharapova M.	1	2	8016	7335	6	3	7	6			2	0

Obrázek 2: Ukázka z databáze: sloupce J – W v excelovském souboru

Comment	B365W	B365L	EXW	EXL	LBW	LBL	PSW	PSL	MaxW	MaxL	AvgW	AvgL
Complete	2,25	1,61	2,2	1,65	2,2	1,67	2,45	1,64	2,45	1,67	2,29	1,62
Complete	1,16	5	1,17	4,75	1,2	4,5	1,2	5,29	1,21	5,5	1,18	4,87
Complete	1,18	4,5	1,22	4	1,22	4,33	1,25	4,61	1,26	4,78	1,22	4,28
Complete	1,12	6	1,16	4,8	1,14	5,5	1,16	6,51	1,18	6,51	1,15	5,65
Complete	1,44	2,75	1,4	2,9	1,44	2,75	1,51	2,85	1,53	3,1	1,44	2,84

Obrázek 3: Ukázka z databáze: sloupce X – AJ v excelovském souboru

V databázi jsou u každého zápasu následující údaje: číslo turnaje v sezóně, místo konání turnaje, název turnaje, datum zápasu, druh turnaje (např. Grand Slam), venku nebo vevnitř, povrch turnaje, kolo turnaje, maximální počet setů v zápase, jméno vítěze a poraženého, postavení hráčů v žebříčku, počet bodů hráčů v žebříčku, počet gamů v každém setu, počet setů v zápase, způsob ukončení zápasu (odehrání, vzdání, nenastoupení nebo diskvalifikace), kurzy před zápasem od sázkových kanceláří Bet365, Expekt, Ladbrokes a Pinnacles Sports a dále průměrné i nejvyšší kurzy.

### 3.3 Přehled databáze

V tabulce 1 je základní přehled databáze.

	Muži		Ženy	
Začátek DB	2005		2007	
Konec DB	2016		2016	
Počet sezón	12		10	
Počet zápasů	32 510	100,00%	24 614	100,00%
Odehrané	31 232	96,07%	23 721	96,37%
Vzdané	1 096	3,37%	746	3,03%
Nenastoupené	181	0,56%	147	0,60%
Diskvalifikace	1	0,00%	0	0,00%
Tvrký povrch	17 536	53,94%	14 819	60,21%
Koberec	839	2,58%	166	0,67%
Tráva	3 666	11,28%	2 529	10,27%
Antuka	10 469	32,20%	7 100	28,85%
Vevnitř	5 945	18,29%	2 411	9,80%
Venku	26 565	81,71%	22 203	90,20%
Počet setů	82 294		55 785	
Počet gamů	809 685		518 688	

Tabulka 1: Základní přehled databáze

K dispozici je tenisová databáze od roku 2005 (resp. 2007) do roku 2016. V databázi se nachází 32 510 zápasů mužů a 24 614 zápasů žen.

Více než 96 % zápasů se odehrálo od začátku až do konce. Přes 3 % zápasů hráči vzdali během zápasu. K přibližně 0,6 % zápasů hráči vůbec nenastoupili. Pouze v jediném utkání se stalo, že byl během utkání tenista diskvalifikován. Přihodilo se to v roce 2012 v utkání mezi Davidem Nalbandianem a Marinem Cilicem. Jak uvádí Ornstein (2012), k diskvalifikaci Davida Nalbandiana došlo ihned poté, co nakopl reklamní panel v blízkosti čárového rozhodčího a tím ho zranil.

Více než polovina utkání se hrála na tvrdém povrchu. V necelé třetině případů se tenisté utkali na antuce. Přibližně 10 % zápasů sehráli tenisté na trávě. Přibližně 2,5 % zápasů mužů a 0,7 % zápasů žen se odehrálo na koberci.

Přes 80 % utkání mužů a přes 90 % utkání žen se odehrálo venku. Zbylé zápasy se hrály vevnitř.

## 4 Elo rating systém

Elo rating systém je hodnocení kvality jednotlivých hráčů a již dlouhou dobu se používá v šachu. Autorem je Arpad Elo (1978), který tento systém popsal v knize *The Rating of Chess Players, Past and Present*.

V této kapitole bude popsán vývoj Elo rating systému. Dále zde bude ukázáno, jak lze spočítat v Elo rating systému pravděpodobnost výhry hráče v zápase.

Následující části čerpají základní popisy z Ross (2007) a *TheChessPiece* (n.d.).

### 4.1 Stav před zavedením Elo rating systému

Americká šachová federace USCF (United States Chess Federation) používala před zavedením Elo rating systému Harkness systém, který byl pojmenován po autorovi Kennethu Harknessovi. Při používání tohoto modelu byly pozorovány nepřesnosti. Proto byl Arpad Elo pověřen revizí tohoto systému. Přitom měl dodržet systém výkonnostních tříd a stupnici hodnocení hráčů, které USCF používala. Přehled výkonnostních tříd USCF je v tabulce 2.

Hodnocení	Třída
nad 2400	Senior Master
2200–2399	Master
2000–2199	Expert
1800–1999	Class A
1600–1799	Class B
1400–1599	Class C
1200–1399	Class D
1000–1199	Class E

Tabulka 2: Třídy USCF

### 4.2 Elo rating jako normální rozdělení

Výkonnost hráče v zápase ovlivňuje mnoho náhodných veličin. V šachu to je například zdravotní stav hráče, nálada hráče, únava, doba cestování na zápas atd. V tenise to může být navíc například povrch, diváci, rozhodčí či počasí. Jak uvádí Šachový klub TH JUNIOR Banská Bystrica (n.d.), tak v případě šachu to vedlo Arpada Elo při použití Ljapunovovy centrální limitní věty k tomu, že výkonnost hráče v zápase popsal pomocí normálního rozdělení pravděpodobnosti se střední hodnotou  $\mu$  rovnou Elo ratingovému ohodnocení hráče  $R$  a se směrodatnou odchylkou  $\sigma$  rovnou velikosti jedné třídy (200 bodů). Toto nastavení směrodatné odchylky  $\sigma$  je kvůli tomu, aby bylo zachováno původní nastavení tříd USCF (viz tabulka 2), které byly určeny ještě před zavedením Elo rating systému.

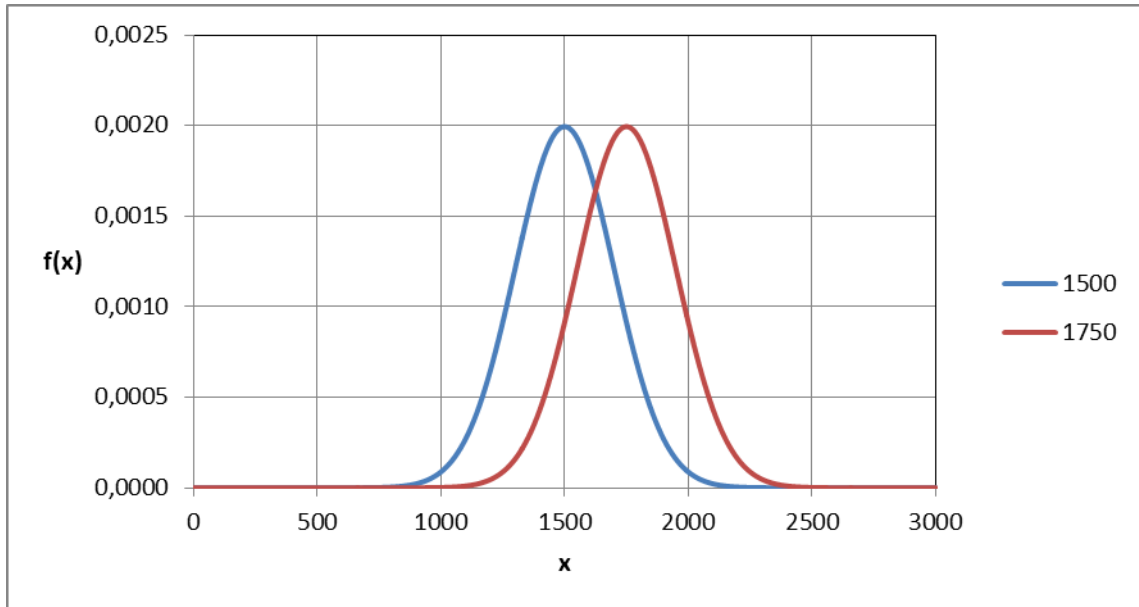
Náhodná veličina  $X$  popisující výkonnost hráče se dle předpokladů řídí normálním rozdělením pravděpodobnosti



$$X \sim N(\mu = R, \sigma = 200). \quad (4.1)$$

Podrobnější popis normálního rozdělení pravděpodobnosti lze nalézt v příloze A.

Na obrázku 4 je ukázka náhodné veličiny popisující výkonnost hráče pro hráče s Elo ratingem 1750 a pro hráče s Elo ratingem 1500.



Obrázek 4: Výkonnost hráče

Pokud jsou známy výkonnosti obou hráčů hrajících zápas, je možné spočítat pravděpodobnost, že hráč  $A$  vyhraje zápas proti hráči  $B$ .

$X$  je označení pro náhodnou veličinu popisující výkonnost hráče  $A$ . Ta se řídí normálním rozdělením

$$X \sim N(\mu = R_A, \sigma = 200). \quad (4.2)$$

$Y$  je označení pro náhodnou veličinu popisující výkonnost hráče  $B$ . Ta se řídí normálním rozdělením

$$Y \sim N(\mu = R_B, \sigma = 200). \quad (4.3)$$

Je předpokládáno, že vyhraje hráč s větší výkonností. Pravděpodobnost  $P_A$ , že hráč  $A$  vyhraje, je tedy rovna pravděpodobnosti, že hráč  $A$  bude mít větší výkonnost než hráč  $B$ . Tedy pravděpodobnost vítězství hráče  $A$  je

$$P_A = P(X > Y) = P(X - Y > 0) = 1 - F_{X-Y}(0). \quad (4.4)$$

Náhodná veličina  $X - Y$  se řídí normálním rozdělením pravděpodobnosti s parametry  $\mu = R_A - R_B$  a  $\sigma = 282,84$

$$(X - Y) \sim N(\mu = R_A - R_B, \sigma = 282,84). \quad (4.5)$$

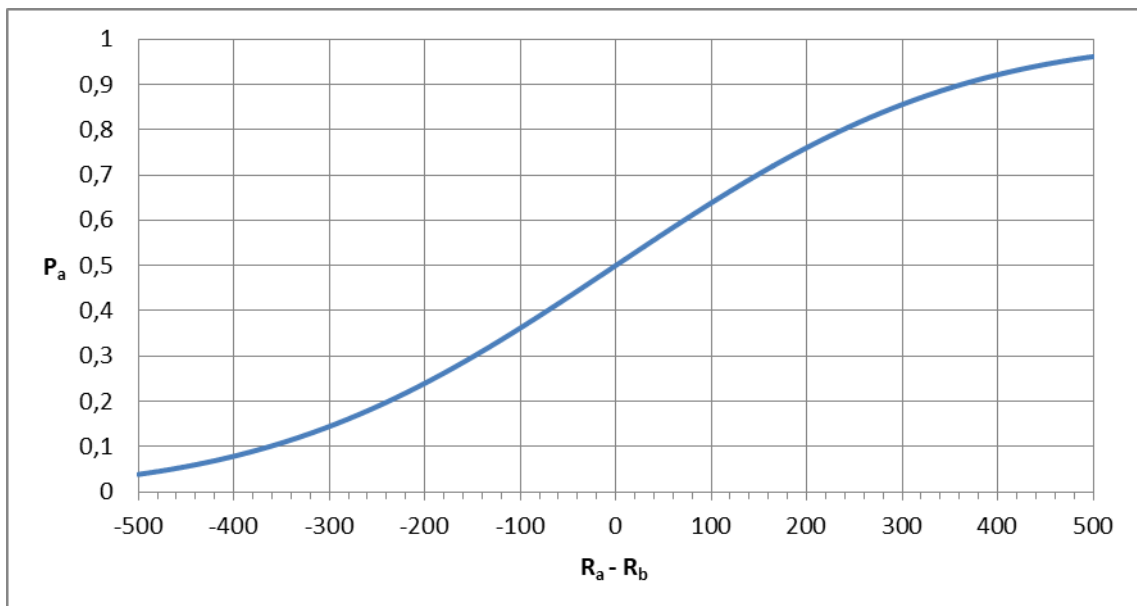
V zápase proti sobě hrají dva hráči s výkonnostmi  $X$  a  $Y$ . Je předpokládáno, že  $X$  a  $Y$  jsou nezávislé náhodné veličiny a obě mají směrodatnou odchylku  $\sigma = 200$ , jak bylo určeno autorem systému Arpadem Elo, kvůli zachování USCF tříd, proto je směrodatná odchylka  $\sigma_{X-Y}$  rovna 282,84

$$\sigma_{X-Y} = \sqrt{\sigma^2 + \sigma^2} = \sqrt{2} \cdot \sigma = \sqrt{2} \cdot 200 \cong 282,84. \quad (4.6)$$

Po přepsání rovnice (4.4) lze pravděpodobnost vítězství hráče  $A$  v zápase určit jako hodnotu distribuční funkce  $F(x)$  normálního rozdělení s parametry  $\mu = 0$  a  $\sigma = 282,84$  v bodě  $x = R_A - R_B$

$$P_A = F_{N(\mu=0, \sigma=282,84)}(R_A - R_B). \quad (4.7)$$

Pravděpodobnost vítězství hráče  $A$  nad hráčem  $B$  nezávisí na absolutních hodnotách Elo ratingů, ale pouze na rozdílu Elo ratingů obou hráčů. Na obrázku 5 je ukázána pravděpodobnost výhry hráče  $A$  v závislosti na rozdílu Elo ratingů obou soupeřů.



Obrázek 5: Pravděpodobnost výhry hráče A

### 4.3 Elo rating jako logistické rozdělení

V Ross (2007) je uvedeno, že popis šachové výkonnosti hráče je přesnější pomocí logistického rozdělení pravděpodobnosti, proto USCF nahradila normální rozdělení pravděpodobnosti logistickým rozdělením pravděpodobnosti. Logistické rozdělení pravděpodobnosti je podrobněji popsáno v příloze A.

Pravděpodobnost vítězství hráče  $A$   $P_A$  se rovná hodnotě distribuční funkce  $F(x)$  logistického rozdělení s parametry  $\mu = 0$  a  $s = \frac{400}{\ln 10}$  v bodě  $x = R_A - R_B$

$$P_A = F_{Log(\mu=0, s=\frac{400}{\ln 10})}(R_A - R_B). \quad (4.8)$$

Parametr  $s$  byl nastaven na hodnotu  $\frac{400}{\ln 10}$  opět kvůli tomu, aby bylo zachováno nastavení tříd USCF. Při změně parametru  $s$  by nedošlo v modelu ke změně pravděpodobnosti výhry hráče  $A$  nad hráčem  $B$ , protože při změně parametru  $s$  by se příslušně změnil i rozdíl hodnot Elo ratingů mezi jednotlivými hráči.

Po rozepsání a úpravě rovnice (4.8) lze pravděpodobnost výhry hráče  $A$  nad hráčem  $B$  vyjádřit ve tvaru, ve kterém se v současné době nejčastěji používá

$$P_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}. \quad (4.9)$$

Analogicky jako pravděpodobnost výhry hráče  $A$  lze vypočítat pravděpodobnost výhry hráče  $B$   $P_B$

$$P_B = \frac{1}{1 + 10^{(R_A - R_B)/400}}. \quad (4.10)$$

Po další úpravě lze pravděpodobnost výhry hráče  $A$  přepsat do tvaru

$$P_A = \frac{Q_A}{Q_A + Q_B}, \quad (4.11)$$

a pravděpodobnost výhry hráče  $B$

$$P_B = \frac{Q_B}{Q_A + Q_B}, \quad (4.12)$$

kde jsou

$$Q_A = 10^{R_A/400}, \quad (4.13)$$

$$Q_B = 10^{R_B/400}. \quad (4.14)$$

Zároveň platí

$$P_B = 1 - P_A. \quad (4.15)$$

*Příklad výpočtu pravděpodobností výhry hráčů v zápase:*

Mějme hráče  $A$  a  $B$ , kteří hrají proti sobě. Hráč  $A$  má Elo rating ve výši 1750 a hráč  $B$  má Elo rating ve výši 1500.

$$P_A = \frac{1}{1 + 10^{(1500 - 1750)/400}} = 80,83 \%, \quad (4.16)$$

$$P_B = 1 - P_A = 19,17 \%. \quad (4.17)$$

Pravděpodobnost výhry hráče  $A$  je v tomto utkání 80,83 % a pravděpodobnost výhry hráče  $B$  je 19,17 %.

V Elo rating systému nezáleží na konkrétní hodnotě Elo ratingu, ale vždy jen na rozdílu hodnot Elo ratingů obou soupeřů. Hráč s větším Elo ratingem je označen jako hráč A. V tabulce 3 je zobrazena pravděpodobnost výhry hráče A při různých rozdílech Elo ratingů.

<b>Rozdíl Elo</b>	<b>Pravděpodobnost</b>
<b>0</b>	50,00 %
<b>25</b>	53,59 %
<b>50</b>	57,15 %
<b>75</b>	60,63 %
<b>100</b>	64,01 %
<b>125</b>	67,25 %
<b>150</b>	70,34 %
<b>175</b>	73,25 %
<b>200</b>	75,97 %
<b>250</b>	80,83 %
<b>300</b>	84,90 %
<b>350</b>	88,23 %
<b>400</b>	90,91 %
<b>500</b>	94,68 %
<b>600</b>	96,93 %
<b>700</b>	98,25 %

**Tabulka 3: Pravděpodobnost výhry hráče A v závislosti na rozdílu Elo ratingů**

## 5 Odhad Elo ratingu

V kapitole 4 bylo ukázáno, jaká je pravděpodobnost vítězství hráče na základě jeho Elo ratingu a Elo ratingu soupeře. Problém však je, že skutečný Elo rating je neznámý. Proto je třeba Elo rating nějakým způsobem odhadnout.

Tato kapitola se věnuje odhadu úvodního Elo ratingu hráče, který zatím ještě žádný Elo rating nemá. Dále se věnuje také přepočtu Elo ratingu hráče, který už nějaký Elo rating má, protože lze předpokládat, že se Elo rating v čase mění.

Následující části čerpají základní popisy z Ross (2007) a TheChessPiece (n.d.).

### 5.1 Elo pro nové hráče

Pokud hráč dosud nemá odhadnutý žádný Elo rating, je třeba mu odhadnout úvodní Elo rating. To lze provést různými způsoby. Jednou z možností je všem hráčům na začátku přiřadit stejný Elo rating. Další možností je vypočítat Elo rating hráči až po odehrání několika zápasů na základě síly jeho soupeřů a úspěšnosti v zápasech. Jednotlivé metody jsou popsány podrobněji v následujících podkapitolách.

#### 5.1.1 Metoda konstantního Elo ratingu

V této metodě je každému novému hráči na začátku přiřazen stejný Elo rating. Například každý nový hráč má na začátku Elo rating 2000.

Ve variantě s konstantním koeficientem rozvoje  $K$ , popsaným v kapitole 5.2.1, jde zároveň o střední hodnotu náhodné veličiny „Elo rating jednotlivých hráčů“, protože každý nový hráč dostane tento Elo rating. Pokud v následném zápase hráč vyhraje, přiroste mu určitý počet Elo bodů (viz následující kapitola 5.2 *Přepočet Elo ratingu*). Ten samý počet bodů však ztratí jeho protivník.

#### 5.1.2 Metoda kvantilu

Tato metoda je založena na výkonnosti hráče v jeho několika prvních zápasech. Po určitém počtu zápasů se vypočte průměrný Elo rating soupeřů. Pokud má hráč úspěšnost  $P$  rovnou 50 %, tak mu je přidělen průměrný Elo rating jeho soupeřů. Pokud má hráč úspěšnost vyšší, přičte se mu několik Elo bodů navíc a naopak. Elo rating se vypočte podle následujícího vzorce

$$R = R_o + D(P), \quad (5.1)$$

kde jsou

- $R$  vypočtený Elo rating,
- $R_o$  průměrný Elo rating soupeřů,
- $D(P)$  korekce.

Korekci Elo ratingu  $D(P)$  lze určit jako kvantil logistického rozdělení  $Q_p$  s parametry  $\mu = 0$  a  $s = \frac{400}{\ln 10} \cong 173,72$ .  $P$  je v tomto případě úspěšnost hráče v utkáních.

Problém u této metody nastává v případě, když je úspěšnost  $P$  rovna 1 (resp. 0). Při této úspěšnosti by byla korekce  $D(P)$  rovna  $+\infty$  (resp.  $-\infty$ ). Proto musí být stanovena maximální (resp. minimální) hodnota  $D(P)$ . Lze použít například hodnotu 800 (resp. -800), což používá šachová federace FIDE (2018). Tyto případy jsou však v tenise velmi vzácné. Například, pokud by se novému hráči odhadoval úvodní Elo rating po 10 zápasech, tak by se takový případ v použité tenisové databázi žen ani mužů nevykytl.

*Odvození:*

V této části bude ukázáno odvození korekce  $D(P)$ . Nejdříve dojde k přepsání rovnice (5.1) do následujícího tvaru

$$D(P) = R - R_o. \quad (5.2)$$

Nyní se vyjde z rovnice (5.2). Myšlenka je taková, že pravděpodobnost úspěšnosti hráče se rovná jeho skutečné úspěšnosti  $P$  a rozdíl Elo ratingů hráčů se rovná korekci  $D(P)$  tedy

$$P = \frac{1}{1 + 10^{-D(P)/400}}. \quad (5.3)$$

V této rovnici je neznámou  $D(P)$ . Tu lze vyjádřit

$$D(P) = 400 \cdot \log\left(\frac{P}{1-P}\right). \quad (5.4)$$

Nyní se dekadický logaritmus převede na přirozený logaritmus

$$D(P) = \frac{400}{\ln 10} \cdot \ln\left(\frac{P}{1-P}\right). \quad (5.5)$$

Korekce Elo ratingu  $D(P)$  se rovná  $p$  kvantilu logistického rozdělení  $Q_p$  s parametry  $\mu = 0$  a  $s = \frac{400}{\ln 10} \cong 173,72$ .

*Příklad výpočtu úvodního Elo ratingu pomocí kvantilu logistického rozdělení:*

Předpokládejme, že má hráč úspěšnost 75 % a průměrný Elo rating jeho soupeřů je 2000. V tomto případě se vypočte úvodní Elo rating hráče následujícím způsobem

$$R = R_o + Q_{75} = 2000 + 190,85 \doteq 2191. \quad (5.6)$$

Úvodní Elo rating hráče je 2191.

### 5.1.3 Metoda prozatímního ratingu

Další metodou, kterou lze určit úvodní Elo rating hráče, je metoda z Glickman (n.d.). Jedná se o úpravu metody kvantilu. Místo kvantilu se používá pro korekci lineární funkce. Elo rating  $R$  se vypočte dle následujícího vzorce

$$R = R_o + 400 \cdot \frac{(W - L)}{N} = R_o + 400 \cdot (2 \cdot P - 1), \quad (5.7)$$

kde jsou

$W$  počet výher,

$L$  počet proher,

$N$  počet zápasů,

$P$  úspěšnost.

### 5.1.4 Zhodnocení metod

Metoda konstantního Elo ratingu se od ostatních metod liší tím, že každý hráč má určen Elo rating ještě před prvním zápasem. Zatímco u zbývajících metod se určí úvodní Elo rating až po několika odehraných zápasech. Na druhé straně jako jediná metoda vůbec nezávisí na výkonu hráče a síle jeho soupeřů.

## 5.2 Přepočítání Elo ratingu

Výkonnost hráče se v čase mění. Tím se samozřejmě mění i Elo rating hráče. Proto je třeba odhad Elo ratingu průběžně přepočítávat. Tento přepočítání může proběhnout po každém odehraném zápase hráče, což je používáno v této práci. Přepočítání může probíhat také po nějakém časovém období. To například používají šachové federace, protože nemají k dispozici výsledky ze všech zápasů a turnajů ihned po odehrání každé partie.

Pro výpočet nového Elo ratingu hráče po odehraném zápase se používá tzv. průběžná metoda. V této metodě se vítězi utkáni k jeho původnímu Elo ratingu přičte určitý počet Elo bodů a poraženému odečte. Jak moc velký bude tento přírůstek či úbytek, záleží na dvou faktorech. Za prvé záleží na koeficientu rozvoje  $K$ . Ten udává maximální možný bodový přírůstek či úbytek Elo ratingu. Více o koeficientu rozvoje bude napsáno v kapitolách 5.2.1–5.2.3. Za druhé záleží na rozdílu Elo ratingů obou hráčů. Pokud porazí favorit slabšího hráče, přičte se mu málo Elo bodů a soupeři se odečte také málo Elo bodů. Naopak pokud slabší hráč porazí silnějšího, přičte se mu hodně Elo bodů a jeho soupeři se hodně Elo bodů odečte. Nový Elo rating  $R_N$  se průběžnou metodou vypočte dle vzorce

$$R_N = R_P + K \cdot (S - S_O), \quad (5.8)$$

kde jsou

- $R_P$  původní Elo rating před zápasem,
- $K$  koeficient rozvoje,
- $S$  výsledek utkání, výhra = 1, prohra = 0,
- $S_O$  pravděpodobnost výhry hráče před zápasem.

*Příklad přepočtu Elo ratingu po odehraném zápase:*

Mějme hráče  $A$  a  $B$ , kteří hrají proti sobě. Hráč  $A$  má Elo rating 1750 a hráč  $B$  má Elo rating 1500. Koeficient rozvoje je pro oba hráče 20 (hodnota koeficientu rozvoje  $K$  bude v kapitole 7 optimalizována). Jak bylo spočteno v kapitole 4.3, pravděpodobnost vítězství hráče  $A$  je 80,83 % a pravděpodobnost vítězství hráče  $B$  je 19,17 %.

V prvním případě vyhrál hráč  $A$ . Pak jeho nový Elo rating  $R_N^A$  je

$$R_N^A = 1750 + 20 \cdot (1 - 0,8083) = 1750 + 3,834 \cong 1754. \quad (5.9)$$

Nový Elo rating  $R_N^B$  hráče  $B$  je

$$R_N^B = 1500 + 20 \cdot (0 - 0,1917) = 1500 - 3,834 \cong 1496. \quad (5.10)$$

V druhém případě vyhrál hráč  $B$ . Nový Elo rating  $R_N^B$  hráče  $B$  je

$$R_N^B = 1500 + 20 \cdot (1 - 0,1917) = 1500 + 16,166 \cong 1516. \quad (5.11)$$

Nový Elo rating  $R_N^A$  hráče  $A$  je

$$R_N^A = 1750 + 20 \cdot (0 - 0,8083) = 1750 - 16,166 \cong 1734. \quad (5.12)$$

Jak již bylo zmíněno výše, koeficient rozvoje  $K$  určuje, jak moc se bude měnit Elo rating. Čím je koeficient rozvoje větší, tím více se bude Elo rating měnit. Elo rating se při jednom přepočtu může zvětšit maximálně o hodnotu  $K$  a to v případě, že hráč vyhraje, ale pravděpodobnost výhry hráče před zápasem se blížila k 0, nebo se může zmenšit maximálně o hodnotu  $K$  a to v případě, že hráč prohraje, ale pravděpodobnost výhry hráče před zápasem se blížila k 1.

Koeficient rozvoje  $K$  lze nastavit různě. První možností je všem hráčům nastavit stejný koeficient rozvoje  $K$ . Případně nastavit více konstantních koeficientů rozvoje  $K$ . Další možností je nastavit koeficient rozvoje  $K$  jako funkci. Jednotlivé možnosti jsou podrobněji popsány v následujících podkapitolách.

### 5.2.1 Stejný koeficient rozvoje

Toto je nejjednodušší metoda nastavení koeficientu rozvoje  $K$ . Po celou dobu a pro všechny hráče je tento koeficient stejný.



### 5.2.2 Více konstantních koeficientů rozvoje

Tuto metodu používají šachové federace. Pro silnější hráče a hráče s více odehranými zápasy je koeficient rozvoje  $K$  nižší. Vychází to z předpokladu, že u silnějších hráčů je výkonnost konstantnější a zároveň po více odehraných zápasech je odhad Elo ratingu přesnější. Naopak pro nové hráče, slabší hráče či hráče mladší 18 let je koeficient rozvoje vyšší, protože u dětí i u nových hráčů se výkonnost mění rychleji.

V tabulce 4 je příklad koeficientů rozvoje  $K$ , které používají šachové federace FIDE (2018) a USCF (TheChessPiece, n.d.).

Federace	K	Hráč
USCF	32	Elo menší než 2100
	24	Elo mezi 2100 a 2400
	16	Elo větší než 2400
FIDE	40	Pro hráče s méně než 30 partiiemi a hráče mladší 18 let, pokud mají Elo menší než 2300
	20	Elo menší než 2400
	10	Elo větší než 2400

Tabulka 4: Koeficient rozvoje  $K$  FIDE a USCF

### 5.2.3 Koeficient rozvoje jako funkce

Koeficientu rozvoje  $K$  lze také určit pomocí funkce v závislosti na počtu odehraných zápasů dle Ingram (2016). Funkce je ve tvaru

$$K = \frac{c}{(M + o)^s}, \quad (5.13)$$

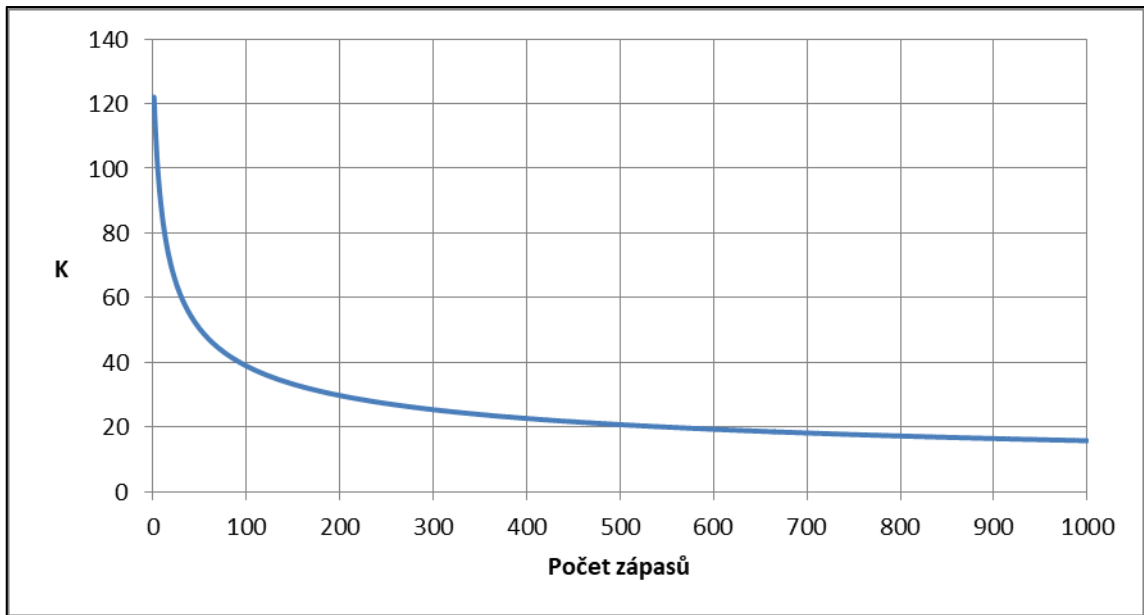
kde jsou

- $c$  konstanta,
- $M$  počet zápasů,
- $o$  parametr korekce pro malý počet zápasů,
- $s$  parametr umožňující flexibilitu tvaru křivky.

Na základě experimentu bylo v Ingram (2016) dospěno k funkci ve tvaru

$$K = \frac{250}{(M + 5)^{0,4}}. \quad (5.14)$$

Na obrázku 6 je zobrazen koeficient rozvoje  $K$  pro různý počet zápasů.



Obrázek 6: Koeficient rozvoje K funkce

## 6 Kritéria určující kvalitu modelu

Každý vytvořený model je třeba nějakým způsobem zhodnotit a určit, jak je dobrý. K tomu jsou použita následující tři kritéria: přesnost, logaritmická ztrátová funkce a kalibrace. Jsou to stejná kritéria, která byla použita v Kovalchik (2016). Proto bude možné později porovnat modely z tohoto článku s modely v této práci.

Tato kritéria určování kvality modelu budou také použita při optimalizaci parametrů v jednotlivých modelech, například při optimalizaci koeficientu rozvoje  $K$  v Elo modelech.

### 6.1 Přesnost

Prvním kritériem je přesnost modelu  $Pr$ . Toto kritérium porovnává počet zápasů, ve kterých vyhrál favorit zápasu dle modelu, se všemi odehranými zápasy. Samozřejmě, čím je přesnost vyšší, tím je model lepší. Vzorec pro vypočtení přesnosti je

$$Pr = \frac{\sum_{i=1}^n (y_i)}{n}, \quad (6.1)$$

kde jsou

$y_i$       indikátor výhry v  $i$ -tém zápase,  
 $n$       celkový počet zápasů.

### 6.2 Logaritmická ztrátová funkce

Dalším kritériem je logaritmická ztrátová funkce  $LogLoss$ . Při výpočtu tohoto kritéria se sčítají logaritmy před zápasem vypočtených pravděpodobností vítězství hráčů v zápasech, kteří později v utkáních vyhráli. Celá suma je pak vydělena celkovým počtem zápasů a vynásobena (-1) proto, aby vyšlo kladné číslo. Cílem je mít toto kritérium co nejmenší.

Myšlenka je taková, že pokud vyhraje hráč, u něhož se pravděpodobnost vítězství blížila 1, tak logaritmus pravděpodobnosti se bude blížit 0 a hodnota kritéria se nebude zvětšovat. Naopak pokud vyhraje hráč, který měl pravděpodobnost vítězství nízkou, tak logaritmus pravděpodobnosti bude velké záporné číslo a hodnota kritéria se po změně znaménka ze záporného na kladné bude zvětšovat. Platí, že čím je hodnota logaritmické ztrátové funkce menší, tím je model lepší.

Vzorec pro vypočtení logaritmické ztrátové funkce je ve tvaru

$$LogLoss = -\frac{1}{n} \cdot \sum_{i=1}^n [y_i \cdot \ln(\hat{p}_i) + (1 - y_i) \cdot \ln(1 - \hat{p}_i)], \quad (6.2)$$

kde jsou

$y_i$       indikátor výhry v  $i$ -tém zápase,  
 $\hat{p}_i$       pravděpodobnost vítězství favorita dle modelu v zápase.

### 6.3 Kalibrace

Posledním kritériem určení kvality modelu je kalibrace  $Kal$ . Ta porovnává součet všech pravděpodobností vítězství favorita dle modelu v zápase, k sumě všech zápasů, ve kterých vyhrál favorit. U dobře kalibrovaného modelu vyjde kalibrace rovna 1.

Myšlenka je taková, že když se například bude hrát 100 zápasů, kdy favorité budou mít vždy pravděpodobnost vítězství 70 %, potom suma všech pravděpodobností vítězství bude 70 a zároveň by měli favorité vyhrát 70 zápasů a tím by byla kalibrace rovna 1.

Vzorec pro vypočítání kalibrace je

$$Kal = \frac{\sum_{i=1}^n \hat{p}_i}{\sum_{i=1}^n y_i}, \quad (6.3)$$

kde jsou

$y_i$             indikátor výhry v  $i$ -tém zápase,

$\hat{p}_i$             pravděpodobnost vítězství favorita dle modelu v zápase.

## 7 Elo modely

V této kapitole jsou představeny a popsány jednotlivé modely, které jsou použity pro odhadování výsledků tenisových utkání v této práci. Je zde provedena optimalizace jednotlivých parametrů na základě kritérií určujících kvalitu modelu, která jsou popsána v kapitole 6. Hlavním optimalizačním kritériem je logaritmická ztrátová funkce, která je popsána v kapitole 6.2.

Jak bylo uvedeno v kapitole 3, k dispozici jsou data ze zápasů mužů od roku 2005 do roku 2016 a data ze zápasů žen od roku 2007 do roku 2016. Na datech z let 2013 a 2014 je provedena optimalizace parametrů modelů (data od roku 2005 u mužů, resp. 2007 u žen, do roku 2012 slouží k učení modelů). Po dobu těchto dvou let jsou výsledky zápasů započítávány do kritérií určujících kvalitu modelu a na základě těchto kritérií jsou poté parametry optimalizovány. Optimalizace parametrů probíhá na datech po dobu dvou sezón, protože i následné ověření modelů pomocí sázení, popsané v kapitole 9, trvá 2 sezóny (2015 a 2016).

Optimalizace probíhá v Matlabu R2014b. K optimalizaci je v Matlabu vytvořen program `Optimalizace_Elo_modely.m`, který je k práci přiložen. Zde si uživatel na začátku zvolí, který ze tří níže popsaných modelů chce optimalizovat a zda chce optimalizovat parametry pro zápasy mužů či žen (způsob nastavení je ukázán v příloze B). Na základě výběru uživatele proběhne optimalizace pro vybraný model. Pro různé kombinace optimalizovaných parametrů jsou spočteny hodnoty kritérií určujících kvalitu modelu. Výsledky jsou zaznamenány do Excelu.

Pro srovnání výsledků optimalizace lze použít článek Kovalchik (2016). V tomto článku je pro dva modely `FiveThirtyEight` založené na Elo rating systému hodnota ztrátové funkce 0,59 (resp. 0,60), přesnost je 0,70 (resp. 0,67) a kalibrace je 1,03 (resp. 0,98). Modely v článku jsou použity jen na zápasy mužů v roce 2014, je zde i kratší historie dat. Proto je srovnání pouze orientační. Kompletní srovnání optimalizovaných modelů z této práce s modely v článku je provedeno v kapitole 7.4.

### 7.1 Elo model 1

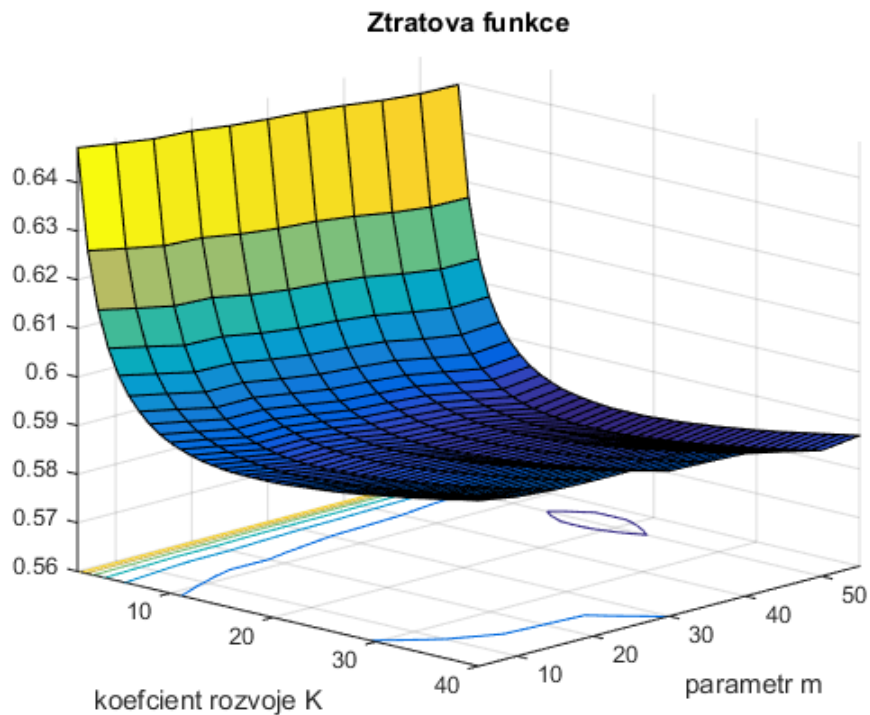
Toto je nejjednodušší Elo model, ve kterém se používá pro odhad úvodního Elo ratingu hráče metoda konstantního Elo ratingu, která byla popsána v kapitole 5.1.1. To znamená, že na počátku dat v roce 2005 (resp. 2007), kdy nejsou u hráčů odhadnuty žádné Elo ratingy, je každému hráči přiřazen Elo rating ve výši 2000. Zároveň každému novému hráči, který vstoupí na okruh ATP či WTA, je také přiřazen Elo rating ve výši 2000. Vzhledem k tomu, že v Elo rating systému nezáleží na konkrétní hodnotě Elo ratingu, jak bylo ukázáno v kapitole 4.3 *Elo rating jako logistické rozdělení*, ale záleží jen na rozdílu Elo ratingů obou hráčů, kteří hrají proti sobě, tak tento počáteční Elo rating není třeba optimalizovat. I v případě, že by byl nastaven jiný úvodní Elo rating, tak by odhady pravděpodobností výsledků tenisových utkání vyšly stejně.

Jelikož se výkonnost hráčů v čase mění, je třeba přepočítávat jejich Elo rating. Ten se přepočítává po každém zápase hráče. Pro přepočítávání Elo ratingu hráče je v tomto případě zvolen konstantní koeficient rozvoje  $K$ , popsáný v kapitole 5.2.1. Tento parametr již bude potřeba optimalizovat. Je to zároveň hlavní parametr, který se v tomto modelu optimalizuje.

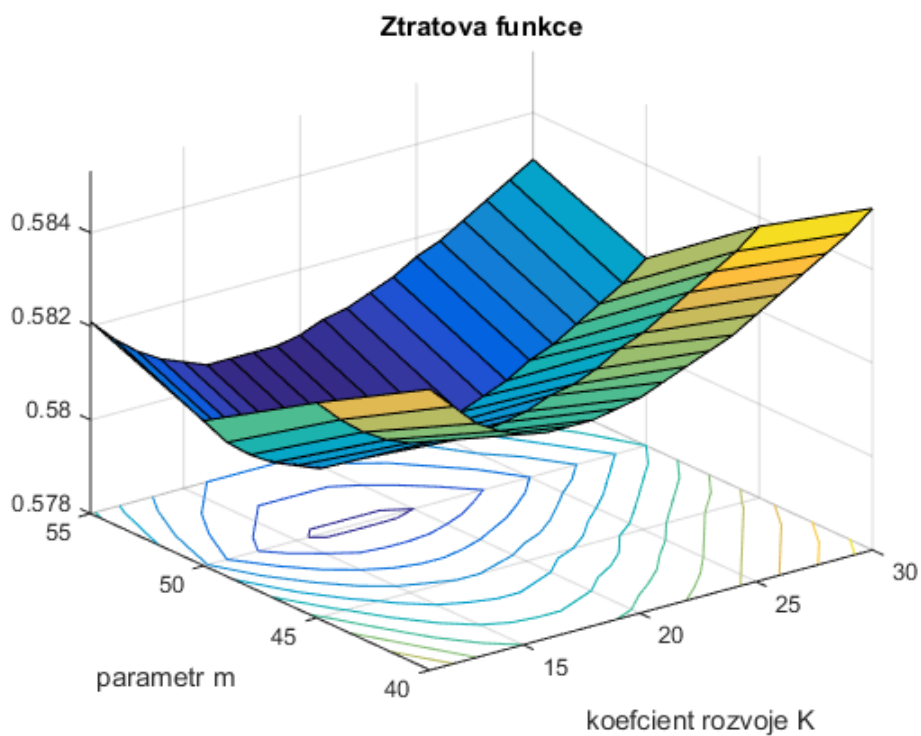
Jak bylo uvedeno výše, všem hráčům je přiřazený na začátku stejný Elo rating ve výši 2000, ale výkonnost jednotlivých hráčů je různá. Odhadnutý Elo rating se hráčům začne stabilizovat až po několika odehraných zápasech. Proto není dobré do kritérií určujících kvalitu modelu i pro sázení započítávat hráče, kteří odehráli jen málo zápasů. K odfiltrování těchto hráčů slouží parametr  $m$ . Tento parametr určuje, kolik musí mít oba soupeři v zápase minimálně odehráno zápasů, aby se jejich vzájemný zápas započítal do kritériálních hodnocení, a zároveň aby bylo možné na daný zápas v pozdější části práce vsadit. Tento parametr  $m$  je třeba také optimalizovat. Zde je třeba volit kompromis mezi minimalizováním logaritmické ztrátové funkce a vybráním velkého počtu zápasů do modelu. Protože při velké hodnotě parametru  $m$  by bylo vybráno jen málo zápasů a málo hráčů. Cílem je však mít kvalitní model pro co nejvíce zápasů a hráčů. Proto je pro optimalizaci tohoto parametru zvolena hodnota prvního lokálního minima, a to i vzhledem k tomu, že velikost ztrátové funkce se i pro vyšší hodnoty parametru  $m$  příliš neliší od hodnoty ztrátové funkce v prvním lokálním minimu. To je ukázáno v přiloženém souboru *EloModel1 optimalizace.xlsx*, ve kterém jsou zaznamenávány výsledky optimalizace tohoto modelu.

### 7.1.1 Optimum muži

Na obrázcích 7 a 8 jsou výsledky optimalizace u mužů. Je zde zobrazena ztrátová funkce v závislosti na parametrech  $m$  a  $K$ .



Obrázek 7: Elo model 1 ztrátová funkce mužů



Obrázek 8: Elo model 1 výřez ztrátové funkce mužů

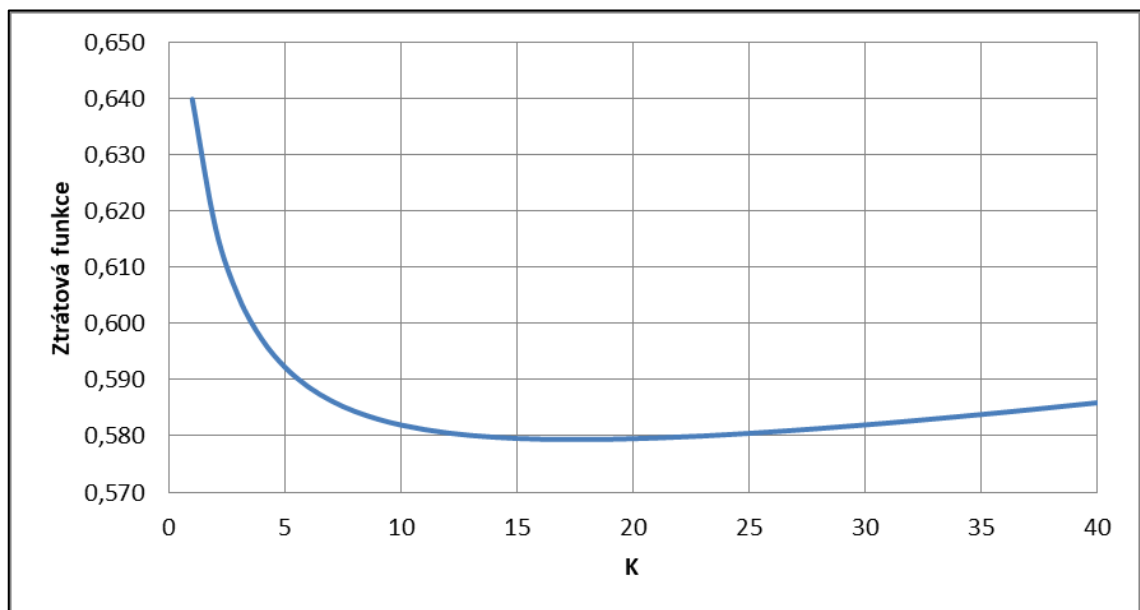
Optimum Elo modelu 1 v kategorii mužů nastává pro hodnoty parametrů  $K = 18$  a  $m = 50$ . V tomto případě je hodnota ztrátové funkce 0,5794, přesnost je 0,6903 a kalibrace je 1,0067. Z grafů lze vypožorovat, že ztrátová funkce v okolí optima je málo

citlivá na změnu hodnot parametrů. Tedy pokud se změní hodnoty parametrů  $K$  a  $m$ , tak se hodnota ztrátové funkce příliš nezmění.

Pro optimalizovanou hodnotu parametru  $m = 50$  je do kvalitativních kritérií započteno 3410 zápasů z celkových 5009 zápasů, které jsou v databázi pro roky 2013 a 2014.

Optimalizovaná hodnota parametru  $K = 18$  je srovnatelná s parametrem  $K = 20$ , který používá šachová federace FIDE pro lepší hráče s více odehranými zápasy, jak bylo napsáno v kapitole 5.2 *Přepočet Elo ratingu*.

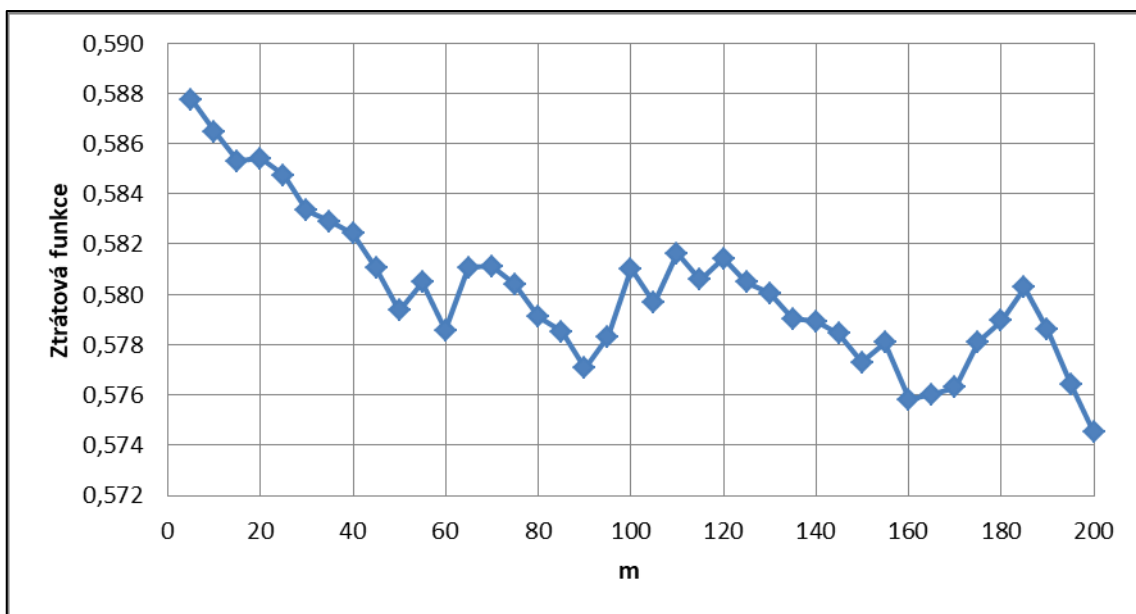
Na obrázku 9 je vykreslena logaritmická ztrátová funkce v závislosti na koeficientu rozvoje  $K$  pro pevně daný parametr  $m = 50$ .



Obrázek 9: Elo model 1 ztrátová funkce pro  $m=50$  muži

Na obrázku 10 je vykreslena logaritmická ztrátová funkce v závislosti na parametru  $m$  pro pevně daný koeficient rozvoje  $K = 18$ .

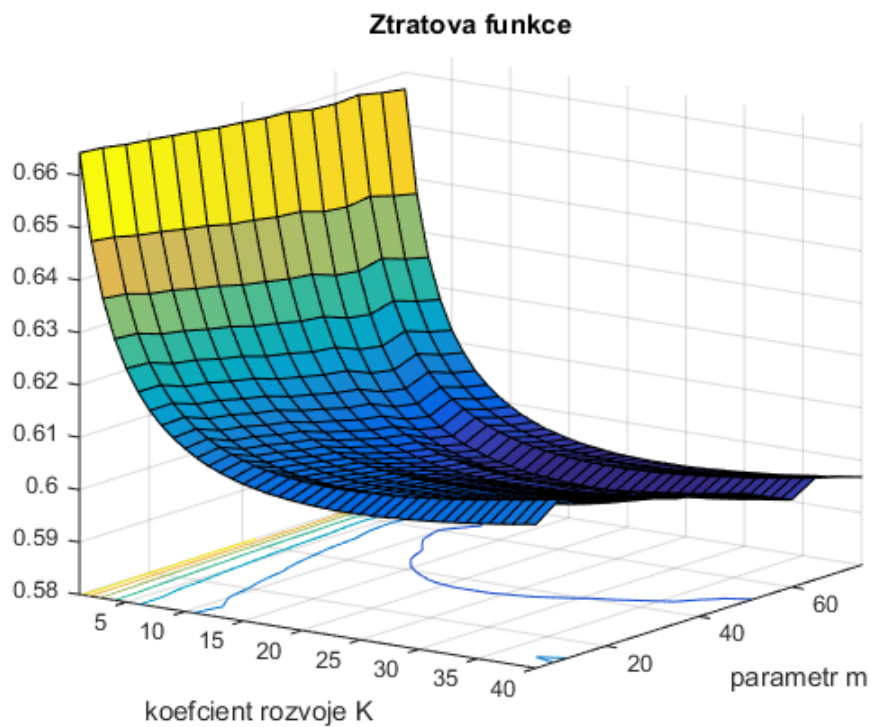




Obrázek 10: Elo model 1 ztrátová funkce pro K=18 muži

### 7.1.2 Optimum ženy

Na obrázcích 11 a 12 jsou výsledky optimalizace u žen. Je zde zobrazena ztrátová funkce v závislosti na parametrech  $m$  a  $K$ .



Obrázek 11: Elo model 1 ztrátová funkce ženy

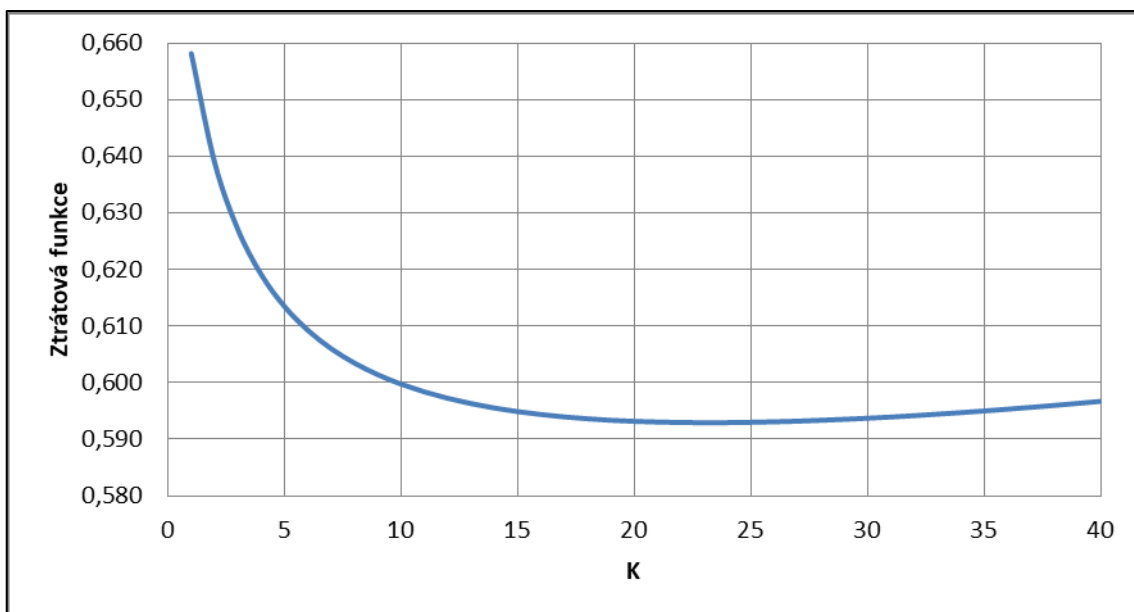


**Obrázek 12: Elo model 1 výřez ztrátové funkce ženy**

Optimum Elo modelu 1 v kategorii žen nastává pro hodnoty parametrů  $K = 23$  a  $m = 60$ . V tomto případě je hodnota ztrátové funkce 0,5929, přesnost je 0,6731 a kalibrace je 1,0154. I v tomto případě je ztrátová funkce v okolí optima málo citlivá na změnu hodnot parametrů  $m$  a  $K$ . Hodnota ztrátové funkce u žen je přibližně o 0,0135 vyšší než v případě mužů.

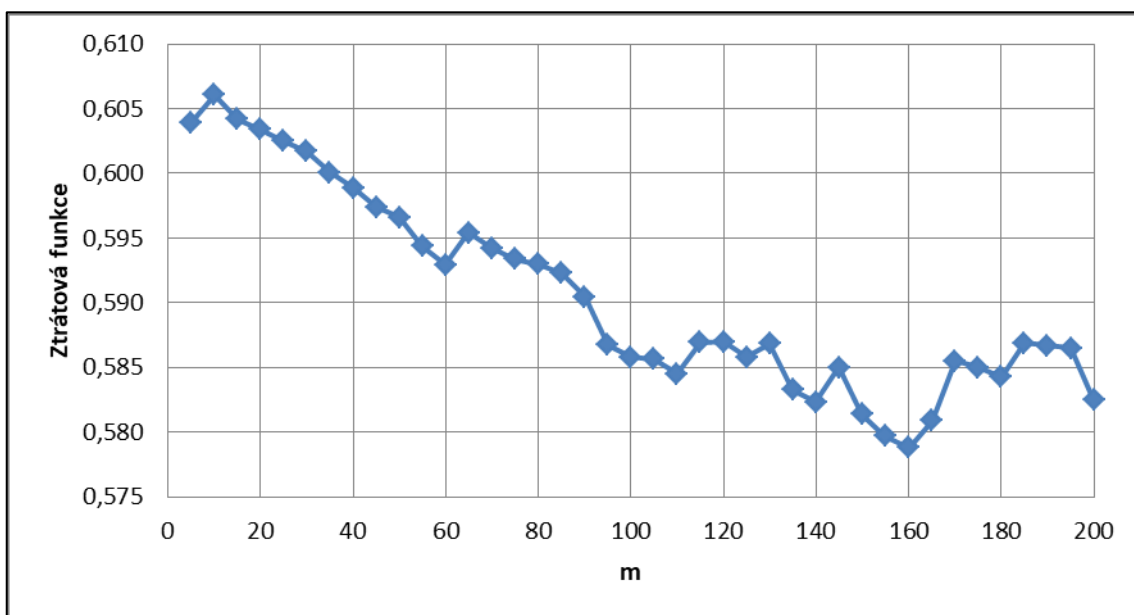
Pro optimalizovanou hodnotu parametru  $m = 60$  je do kvalitativních kritérií započteno 2658 zápasů z celkových 4760 zápasů, které jsou v databázi pro roky 2013 a 2014.

Na obrázku 13 je vykreslena logaritmická ztrátová funkce v závislosti na koeficientu rozvoje  $K$  pro pevně daný parametr  $m = 60$ .



Obrázek 13: Elo model 1 ztrátová funkce pro m=60 ženy

Na obrázku 14 je vykreslena logaritmická ztrátová funkce v závislosti na parametru  $m$  pro pevně daný koeficient rozvoje  $K = 23$ .



Obrázek 14: Elo model 1 ztrátová funkce pro K=23 ženy

## 7.2 Elo model 2

V tomto modelu se opět používá k odhadu úvodního Elo ratingu metoda konstantního Elo ratingu, a tak bude každému novému hráči přiřazen Elo rating ve výši 2000.

Pro přepočítávání Elo ratingu hráče je zde zvolen koeficient rozvoje ve tvaru funkce, popsány v kapitole 5.2.3

$$K = \frac{c}{(M + o)^s}. \quad (7.1)$$

V této funkci jsou optimalizovány parametry  $c$  a  $s$ . Parametr  $s$  udává sklon křivky a  $c$  udává velikost koeficientu rozvoje.

$M$  a  $o$  optimalizovány nejsou, protože  $M$  udává celkový počet zápasů v databázi, který hráč do doby zápasu odehrál. To znamená, že  $M$  je v zápase pevně dáno na základě počtu předchozích odehraných zápasů hráče. Parametr  $o$  slouží jen jako korekce pro hráče, kteří odehráli doposud malý počet zápasů. Proto není tento parametr  $o$  optimalizován, ale je rovnou nastaven na hodnotu  $o = 5$ , což je převzato z článku Ingram (2016).

Stejně jako v Elo modelu 1 se začne odhadnutý Elo rating stabilizovat až po několika zápasech. Proto je třeba opět přidat parametr  $m$ , který určuje, kolik musí mít oba soupeři v zápase minimálně odehráno zápasů, aby se jejich vzájemný zápas započítal do kriteriálních hodnocení.

### 7.2.1 Optimalizace mužů

Výsledky optimalizace mužů jsou v příloženém excelovském souboru *EloModel2 optimalizace muži.xlsx*.

Stejně jako v Elo modelu 1 je první lokální minimum u mužů pro parametr  $m = 50$ . V tabulce 5 jsou vypsaná minima logaritmické ztrátové funkce pro pevně daný parametr  $m = 50$  a různé hodnoty parametru  $s$ .

$s$	$c$	Ztrátová funkce
0,2	50	0,5790
<b>0,3</b>	<b>80</b>	<b>0,5789</b>
0,4	130	0,5790
0,5	215	0,5792
0,6	360	0,5794
0,8	950	0,5804
1,0	2500	0,5821

Tabulka 5: Elo model 2 optimalizace mužů

Minima logaritmická ztrátová funkce nabývá pro  $s = 0,3$  a  $c = 80$ . Pro tyto hodnoty parametrů je hodnota logaritmické ztrátové funkce 0,5789. Přesnost je 0,6891 a kalibrace je 1,0073. Optimální funkce koeficientu rozvoje pro muže je ve tvaru

$$K = \frac{80}{(M + 5)^{0,3}}. \quad (7.2)$$

Hodnota ztrátové funkce u mužů v Elo modelu 2 je téměř stejná jako v Elo modelu 1.

Jelikož optimalizovaná hodnota parametru  $m$  je opět  $m = 50$ , tak je stejně jako v Elo modelu 1 do kvalitativních kritérií započteno 3410 zápasů z celkových 5009 zápasů, které jsou v databázi pro roky 2013 a 2014.

## 7.2.2 Optimalizace ženy

Výsledky optimalizace žen jsou v příloženém excelovském souboru *EloModel2 optimalizace ženy.xlsx*.

Stejně jako v Elo modelu 1 je první lokální minimum u žen pro parametr  $m = 60$ . V tabulce 6 jsou vypsána minima logaritmické ztrátové funkce pro pevně daný parametr  $m = 60$  a různé hodnoty parametru  $s$ .

s	c	Ztrátová funkce
0,1	40	0,5928
<b>0,2</b>	<b>65</b>	<b>0,5928</b>
0,3	100	0,5929
0,4	175	0,5932
0,6	425	0,5942
0,8	1125	0,5958
1,0	2800	0,5984

Tabulka 6: Elo model 2 optimalizace ženy

Minima logaritmická ztrátová funkce nabývá pro  $s = 0,2$  a  $c = 65$ . Pro tyto hodnoty parametrů je hodnota logaritmické ztrátové funkce 0,5928. Přesnost je 0,6754 a kalibrace je 1,0154. Optimální funkce koeficientu rozvoje pro ženy je ve tvaru

$$K = \frac{65}{(M + 5)^{0,2}} \quad (7.3)$$

I v kategorii žen se hodnoty ztrátové funkce v Elo modelu 1 a v Elo modelu 2 téměř neliší.

Jelikož optimalizovaná hodnota parametru  $m$  je opět  $m = 60$ , tak je stejně jako v Elo modelu 1 do kvalitativních kritérií započteno 2658 zápasů z celkových 4760 zápasů, které jsou v databázi pro roky 2013 a 2014.

## 7.3 Elo model 3

V tomto modelu se využívá k odhadu prvního Elo ratingu metoda kvantilů, popsaná v kapitole 5.1.2. Tuto metodu lze použít, pokud hráč hraje proti hráčům, kteří už nějaký Elo rating mají. Tím pádem ji není možné využít hned od začátku dat, protože v té době nemá ještě žádný hráč odhadnutý Elo rating. Proto je pro nastartování modelu použita v prvních sezónách (konkrétně sezóny 2005 a 2006 u mužů a sezóny 2007 a 2008 u žen) metoda konstantního Elo ratingu pro nové hráče, popsaná v kapitole 5.1.1. V dalších sezónách je už určován úvodní Elo rating nových hráčů, kteří vstoupí na okruh, metodou kvantilů.

Pro přepočítání Elo ratingu hráče je v tomto modelu zvolen konstantní koeficient rozvoje  $K$ , popsany v kapitole 5.2.1. Tento koeficient rozvoje  $K$  je třeba optimalizovat.

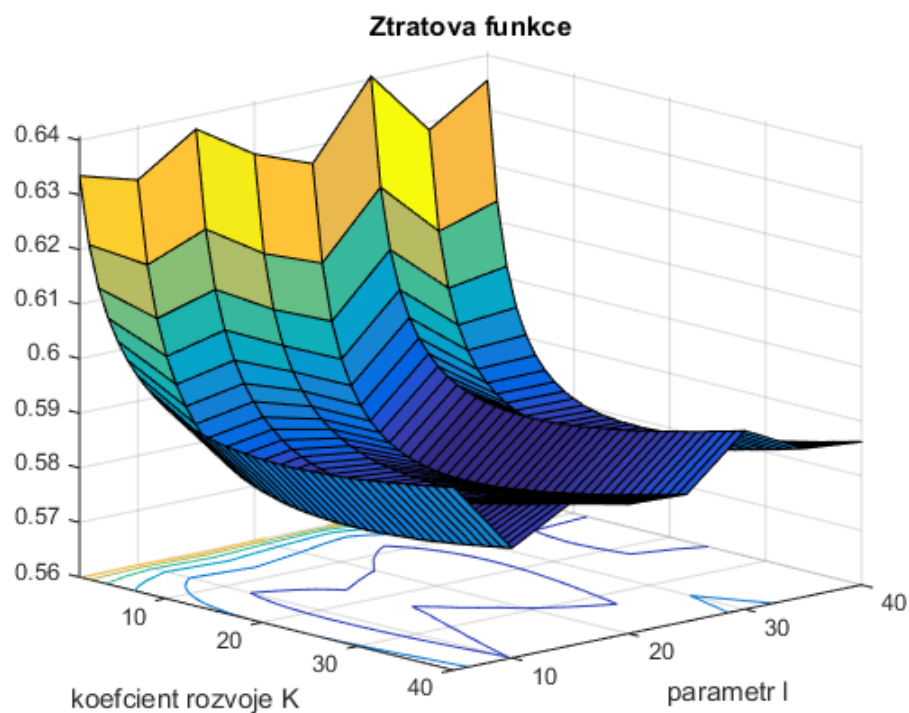
Jak bylo napsáno výše, v tomto modelu se využívá k odhadu prvního Elo ratingu metoda kvantilů. V této metodě je odhadnut úvodní Elo rating hráče až po několika odehraných zápasech na základě jeho výsledků. Kolik musí hráč odehrát zápasů, než mu bude odhadnut Elo rating, bude záviset na parametru  $l$ . Tento parametr udává, po kolika odehraných zápasech proti hráčům s Elo ratingem vyjde hráči úvodní Elo rating. Parametr  $l$  bude třeba také optimalizovat.

Pokud bude hodnota tohoto parametru  $l$  nastavena jako příliš velké číslo, vyvstane problém, protože budou nastupovat proti sobě často hráči bez Elo ratingu. To znamená, že se jim nebude započítávat vzájemný zápas do počtu zápasů potřebných k vytvoření úvodního Elo ratingu.

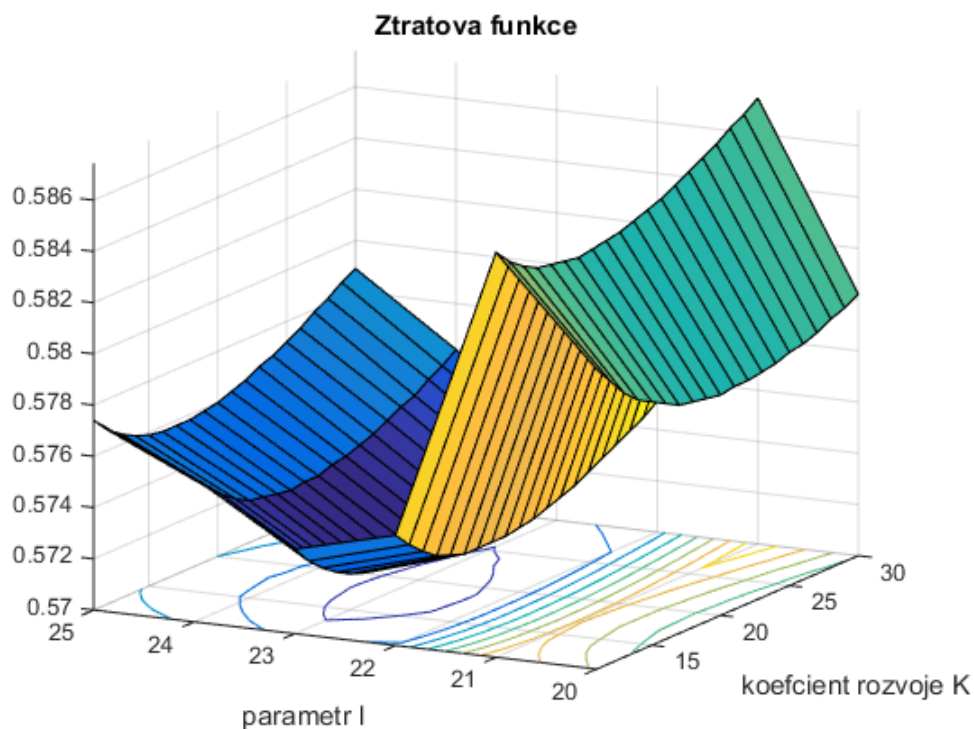
Výsledky optimalizace mužů i žen jsou v příloženém excelovském souboru *EloModel3 optimalizace.xlsx*.

### 7.3.1 Optimalizace muži

Na obrázcích 15 a 16 jsou výsledky optimalizace u mužů. Je zde zobrazena ztrátová funkce v závislosti na parametrech  $l$  a  $K$ .



Obrázek 15: Elo model 3 ztrátová funkce muži



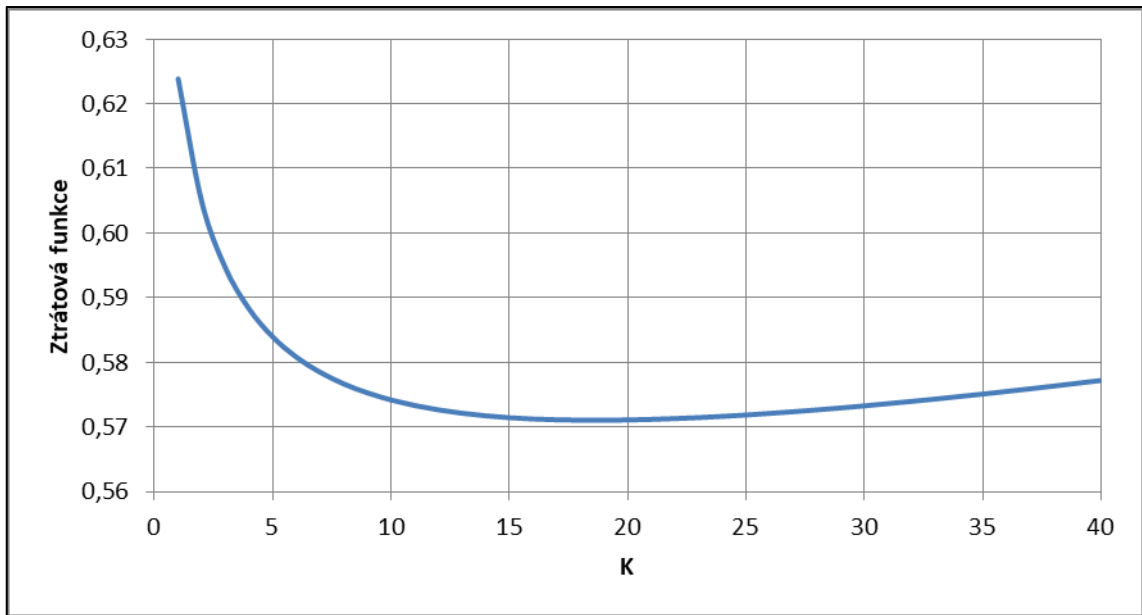
**Obrázek 16: Elo model 3 výřez ztrátové funkce mužů**

Optimum ztrátové funkce pro zápasy mužů nastává pro hodnoty parametrů  $K = 19$  a  $l = 23$ . V tomto případě je hodnota logaritmické ztrátové funkce 0,5710, přesnost je 0,6978 a kalibrace je 1,0056. I v tomto případě platí, že ztrátová funkce je v okolí optima málo citlivá na změnu hodnot parametrů.

Pro optimalizovanou hodnotu parametru  $l = 23$  je do kvalitativních kritérií započteno 2217 zápasů z celkových 5009 zápasů, které jsou v databázi pro roky 2013 a 2014. To je o 1193 zápasů méně než v případě Elo modelů 1 a 2.

V tomto případě je hodnota ztrátové funkce přibližně o 0,008 nižší než u Elo modelů 1 a 2. Tento rozdíl může ovlivňovat rozdílný počet zápasů započítaných do kriteriálních hodnocení.

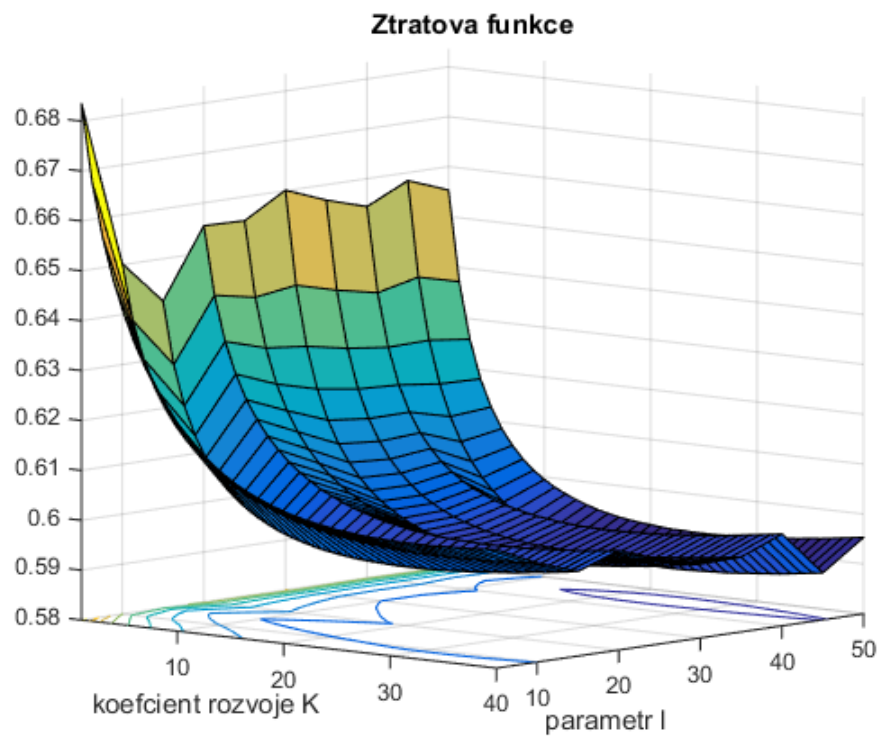
Na obrázku 17 je vykreslena logaritmická ztrátová funkce v závislosti na koeficientu rozvoje  $K$  pro pevně daný parametr  $l = 23$ .



Obrázek 17: Elo model 3 ztrátová funkce pro l=23 muži

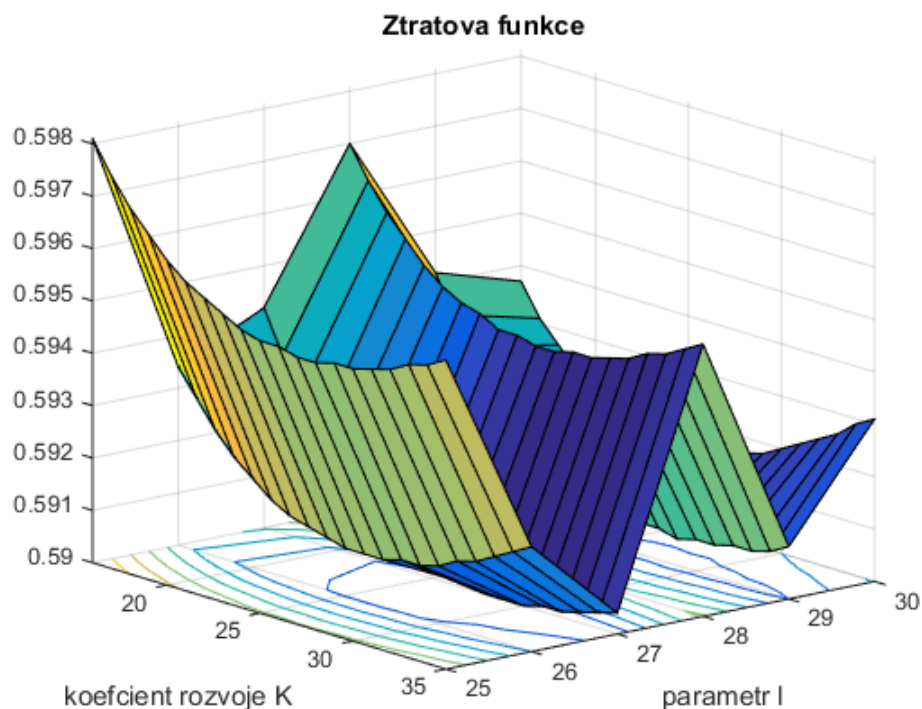
### 7.3.2 Optimalizace ženy

Na obrázcích 18 a 19 jsou výsledky optimalizace u žen. Je zde zobrazena ztrátová funkce v závislosti na parametrech  $l$  a  $K$ .



Obrázek 18: Elo model 3 ztrátová funkce ženy





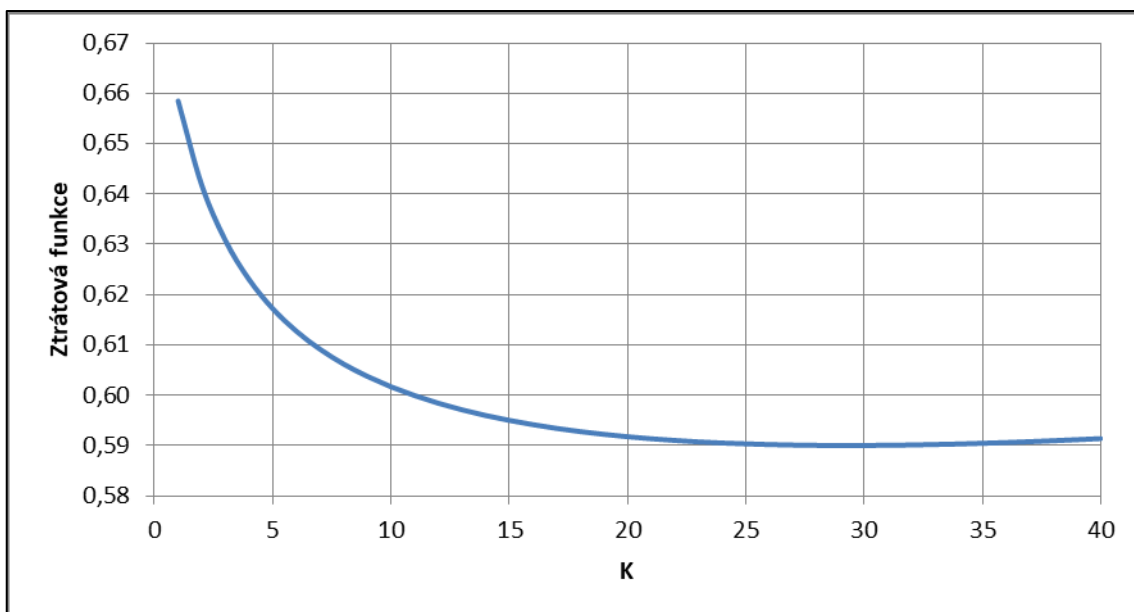
**Obrázek 19: Elo model 3 výřez ztrátové funkce ženy**

Optimum ztrátové funkce pro zápasy žen nastává pro hodnoty parametrů  $K = 29$  a  $l = 27$ . V tomto případě je hodnota logaritmické ztrátové funkce 0,5900, přesnost je 0,6761 a kalibrace je 1,0181.

Pro optimalizovanou hodnotu parametru  $l = 27$  je do kvalitativních kritérií započteno 2584 zápasů z celkových 4760 zápasů, které jsou v databázi pro roky 2013 a 2014. To je o 74 zápasů méně než v případě Elo modelů 1 a 2.

V kategorii žen je hodnota ztrátové funkce u Elo modelu 3 přibližně o 0,003 nižší než u Elo modelů 1 a 2.

Na obrázku 20 je vykreslena logaritmická ztrátová funkce v závislosti na koeficientu rozvoje  $K$  pro pevně daný parametr  $l = 27$ .



Obrázek 20: Elo model 3 ztrátová funkce pro l=27 ženy

## 7.4 Srovnání

V této kapitole budou porovnány výsledky optimalizace všech tří Elo modelů. Dojde zde i ke srovnání výsledků Elo modelů s modely sázkových kanceláří a s modely z článku Kovalchik (2016).

V tabulce 7 jsou hodnoty logaritmické ztrátové funkce, přesnosti a kalibrace v Elo modelech pro optimalizované parametry pro roky 2013 a 2014.

Název	Ztrátová funkce	Přesnost	Kalibrace
<b>Muži</b>			
Elo model 1	0,5794	0,690	1,0067
Elo model 2	0,5789	0,689	1,0073
Elo model 3	0,5710	0,698	1,0056
<b>Ženy</b>			
Elo model 1	0,5929	0,673	1,0154
Elo model 2	0,5928	0,675	1,0154
Elo model 3	0,5900	0,676	1,0181

Tabulka 7: Kritéria určující kvalitu modelu Elo modely

Hodnota ztrátové funkce se u mužů pohybuje mezi 0,57 a 0,58 a u žen mezi 0,59 a 0,60. Přesnost se u mužů pohybuje kolem 0,69 a u žen mezi 0,67 a 0,68. Hodnota kalibrace je vždy větší než 1. U mužů se pohybuje mezi 1,00 a 1,01 a u žen mezi 1,01 a 1,02. U mužů i žen dopadl nejlépe Elo model 3, který měl nejnižší hodnotu ztrátové funkce a zároveň i největší přesnost. Mezi hodnotou ztrátové funkce Elo modelu 3 a ostatních Elo modelů je u mužů rozdíl přibližně 0,008 a u žen je rozdíl přibližně 0,003. Zde je však nutné poznamenat, že na základě parametrů  $m$  a  $l$  bylo do kritériálních hodnocení v kategorii mužů u Elo modelu 3 započteno o 1193 zápasů méně než v případě Elo modelu 1 a 2. Naopak v kategorii žen byl tento rozdíl pouze 74 zápasů.

Ke srovnání výsledků Elo modelů poslouží výsledky z článku Kovalchik (2016). V tomto článku jsou odhadovány výsledky tenisových zápasů za použití různých modelů. Odhadovány jsou pouze výsledky tenisových zápasů mužů v sezóně 2014. V článku se pracuje i s kratší historií dat, proto je nutné brát porovnání výsledků z této práce a z článku jen orientačně.

Jak již bylo zmíněno v kapitole 2, modely v článku jsou rozděleny do čtyř kategorií.

V první kategorii, která je v článku zmíněna, jsou modely založené na regresi (Regression-Based). Jsou zde různé logitové a probitové modely. U těchto modelů je ztrátová funkce o přibližně 0,02 až 0,06 horší než u Elo modelů z této práce.

V další kategorii jsou modely založené na nezávislosti a stejném rozdělení (iid) jednotlivých bodů v zápase (Point-Based). U těchto modelů je ztrátová funkce o přibližně 0,06 až 0,11 horší než u Elo modelů z této práce.

Třetí kategorií jsou modely založené na párovém porovnání (Paired Comparison). Do této kategorie patří modely založené na Bradley-Terry modelu a modely založené na Elo rating systému označené jako FiveThirtyEight. U těchto modelů je ztrátová funkce o přibližně 0,01 až 0,09 horší než u Elo modelů z této práce.

Nejlepším modelem z článku je dle kvalitativních kritérií model založený na kurzech sázkových kanceláří, který je pojmenován BCM. Ztrátová funkce je o přibližně 0,02 lepší než u Elo modelů.

V tabulce 8 jsou shrnuty výsledky z článku Kovalchik (2016). Pouze BCM model měl nižší hodnotu ztrátové funkce než Elo modely z této práce. Všechny ostatní modely z článku včetně modelů FiveThirtyEight založených na Elo rating systému měly hodnotu ztrátové funkce vyšší než Elo modely z této práce. Jak již bylo uvedeno, srovnání je pouze orientační, důležitým závěrem je, že se modely pohybují na podobných hladinách.

Název	Ztrátová funkce	Přesnost	Kalibrace
<b>Regression-Based</b>			
Basic Probit	<b>0,63</b>	0,59	0,92
Prize Probit	<b>0,61</b>	0,68	0,98
Probit plus	<b>0,60</b>	0,67	0,98
Logistic	<b>0,60</b>	0,67	0,97
<b>Point-Based</b>			
Basic IID	<b>0,67</b>	0,63	0,89
Oponent adjusted	<b>0,63</b>	0,67	0,98
Low level	<b>0,68</b>	0,64	0,87
Common oponent	<b>0,66</b>	0,63	0,89
<b>Paired Comparision</b>			
Bradley-Terry 1	<b>0,67</b>	0,62	0,80
Bradley-Terry 2	<b>0,65</b>	0,65	0,80
FiveThirtyEight 1	<b>0,60</b>	0,67	0,98
FiveThirtyEight 2	<b>0,59</b>	0,70	1,03
<b>Other</b>			
BCM	<b>0,55</b>	0,72	0,98

Tabulka 8: Kritéria určující kvalitu modelu článek

Ke srovnání výsledků modelů jsou také spočteny hodnoty ztrátové funkce, přesnosti a kalibrace u mužů i žen pro sezóny 2013 a 2014 pro jednotlivé sázkové kanceláře.

Výpočet probíhá v Matlabu pomocí přiloženého programu KriteriaSK.m.

Při výpočtu pravděpodobností je čerpáno z Cover a Thomas (2006). Pravděpodobnosti vítězství hráčů  $p_i$  dle sázkových kanceláří jsou vypočteny na základě kurzů vypsaných sázkovou kanceláří

$$p_1 = \frac{\frac{1}{o_1}}{\frac{1}{o_1} + \frac{1}{o_2}}, \quad (7.4)$$

kde je

$o_i$  kurz na hráče  $i$ .

Na základě pravděpodobností a skutečných výsledků zápasů jsou vypočtena kritéria určující kvalitu modelu.

V tabulce 9 jsou shrnuty výsledky modelů sázkových kanceláří.

	Ztrátová funkce	Přesnost	Kalibrace
<b>Muži</b>			
<b>Bet365</b>	<b>0,5541</b>	0,697	1,0084
<b>Expekt</b>	<b>0,5556</b>	0,705	0,9808
<b>Ladbrokes</b>	<b>0,5551</b>	0,694	0,9965
<b>Pinnacles Sports</b>	<b>0,5535</b>	0,706	0,9960
<b>Maximální kurz</b>	<b>0,5528</b>	0,707	1,0021
<b>Průměrný kurz</b>	<b>0,5546</b>	0,708	0,9814
<b>Ženy</b>			
<b>Bet365</b>	<b>0,5730</b>	0,685	1,0101
<b>Expekt</b>	<b>0,5748</b>	0,692	0,9828
<b>Ladbrokes</b>	<b>0,5761</b>	0,681	0,9990
<b>Pinnacles Sports</b>	<b>0,5725</b>	0,695	0,9961
<b>Maximální kurz</b>	<b>0,5724</b>	0,695	1,0016
<b>Průměrný kurz</b>	<b>0,5733</b>	0,697	0,9814

Tabulka 9: Kritéria určující kvalitu modelu sázková kancelář

U sázkových kanceláří se pohybuje hodnota ztrátové funkce v kategorii mužů mezi 0,55 a 0,56. U kategorie žen se pohybuje mezi 0,57 a 0,58. V obou kategoriích je to o přibližně 0,02 nižší hodnota než u Elo modelů.

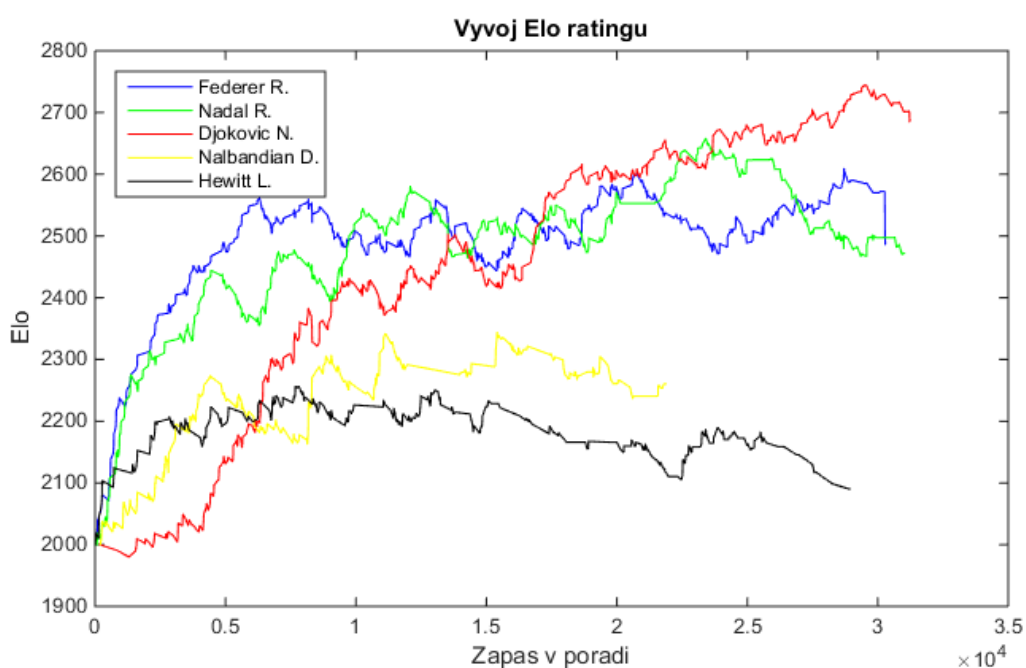
Za zmínku také stojí, že u Elo modelů i u modelů sázkových kanceláří jsou rozdíly hodnot ztrátových funkcí mezi kategoriemi mužů a žen přibližně 0,02.

## 8 Vývoj Elo ratingu v čase

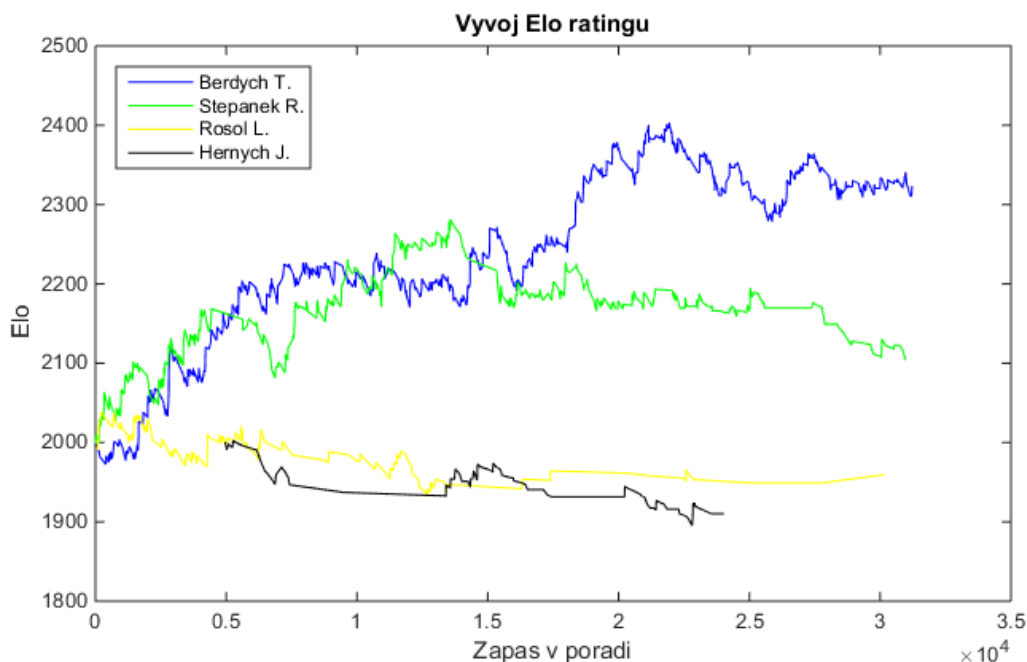
V kapitole 7 byla provedena optimalizace parametrů tří Elo modelů na základě kritérií určujících kvalitu modelu a modely byly poté mezi sebou porovnány. Nejnižší hodnotu ztrátové funkce z těchto tří modelů u mužů i žen měl Elo model 3. V této kapitole bude ukázán vývoj Elo ratingu vybraných světových i českých hráčů v čase právě podle Elo modelu 3.

### 8.1 Muži

Na obrázcích 21 a 22 je zobrazen vývoj Elo ratingu vybraných světových (resp. českých) hráčů od začátku používání Elo modelu 3, tedy od prvního zápasu v roce 2005 do posledního zápasu v roce 2016.



Obrázek 21: Elo rating muži světoví hráči



Obrázek 22: Elo rating muži čeští hráči

Například u hráče Davida Nalbandiana končí křivka znázorňující jeho Elo rating přibližně u 22 000tého zápasu (rok 2013). To znamená, že tento hráč od té doby už nehrál žádný zápas, tudíž již ukončil svoji tenisovou kariéru.

Z obrázků 23 lze vypočítat, že Novak Djokovic měl na začátku modelu (v roce 2005) nejhorší výkonnost z vybraných světových hráčů, ale na konci modelu (v roce 2016) měl výkonnost nejlepší. To ukazuje i tabulka 10, ve které je vypsáno 15 tenistů s největším Elo ratingem na konci roku 2016. Největší Elo rating v tu dobu měl Novak Djokovic s Elo ratingem ve výši 2676, následoval ho Andy Murray s 2602. Větší ztrátu už měl Roger Federer, který byl s Elo ratingem ve výši 2479 třetí v pořadí. V oficiálním žebříčku ATP Rankings (2018) na konci roku 2016 byl na prvním místě Andy Murray, následován Novakem Djokovicem. Roger Federer byl až na 16. místě. To je ale způsobeno tím, že téměř celou druhou polovinu roku vynechal a nehrál žádné zápasy.

Oficiální žebříček ATP zohledňuje pouze výsledky za poslední rok, zatímco Elo rating systém pracuje s delší historií zápasů.

V tabulce tenistů s nejvyšším Elo ratingem se nacházejí i tenisté, kteří už v té době ukončili kariéru, protože se Elo rating hráčů, kteří nehrají zápasy, nemění. Tudíž jim nemůže Elo rating klesnout.

Pořadí	Jméno	Elo rating	Průměrné Elo soupeřů	Počet zápasů s Elo hráči	Počet výher
1	Djokovic N.	<b>2676</b>	2208	705	581
2	Murray A.	<b>2602</b>	2188	631	474
3	Federer R.	<b>2479</b>	2185	762	647
4	Nadal R.	<b>2455</b>	2179	736	612
5	Soderling R.	<b>2401</b>	2133	344	227
6	Del Potro J.M.	<b>2377</b>	2168	369	256
7	Raonic M.	<b>2360</b>	2182	272	182
8	Tsonga J.W.	<b>2322</b>	2183	403	263
9	Cilic M.	<b>2319</b>	2146	433	264
10	Berdych T.	<b>2319</b>	2160	635	398
11	Kyrgios N.	<b>2318</b>	2211	66	42
12	Monfils G.	<b>2311</b>	2128	456	280
13	Gasquet R.	<b>2280</b>	2138	511	317
14	Ferrer D.	<b>2257</b>	2147	686	458
15	Roddick A.	<b>2256</b>	2115	416	307

Tabulka 10: Elo rating Top 15 mužů

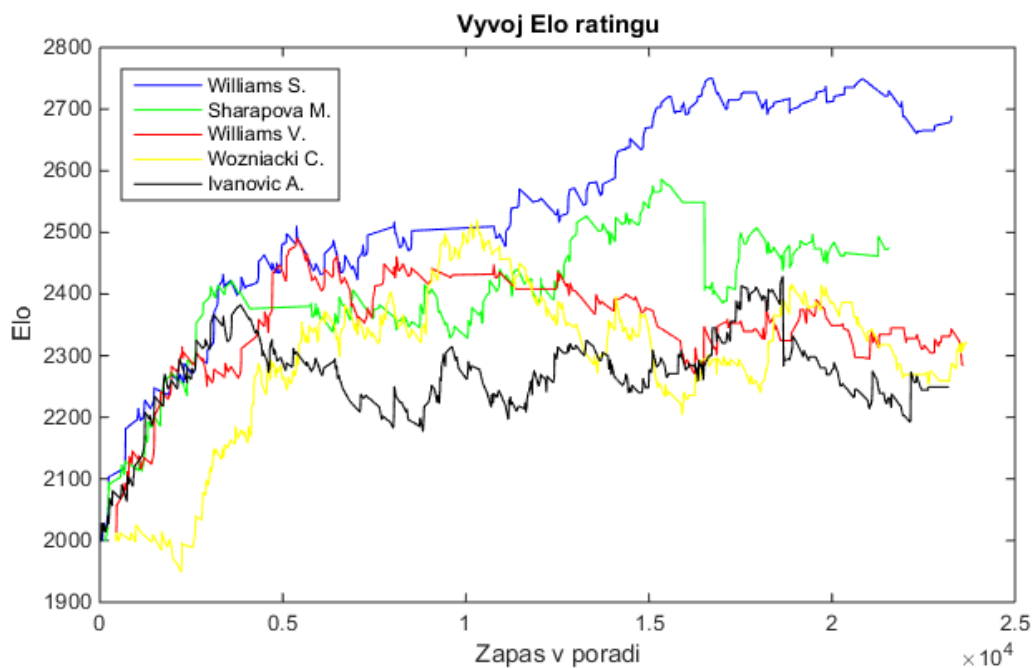
Hráčem s nejnižším Elo ratingem na konci roku 2016 byl na 479. místě K. Kim, který měl Elo rating ve výši 1843. Zbylí hráči žádný Elo rating neměli.

Kompletní žebříček Elo ratingu hráčů z konce roku 2016 lze nalézt v příloženém excelovském souboru *Elo rating žebříček 2016.xlsx*.

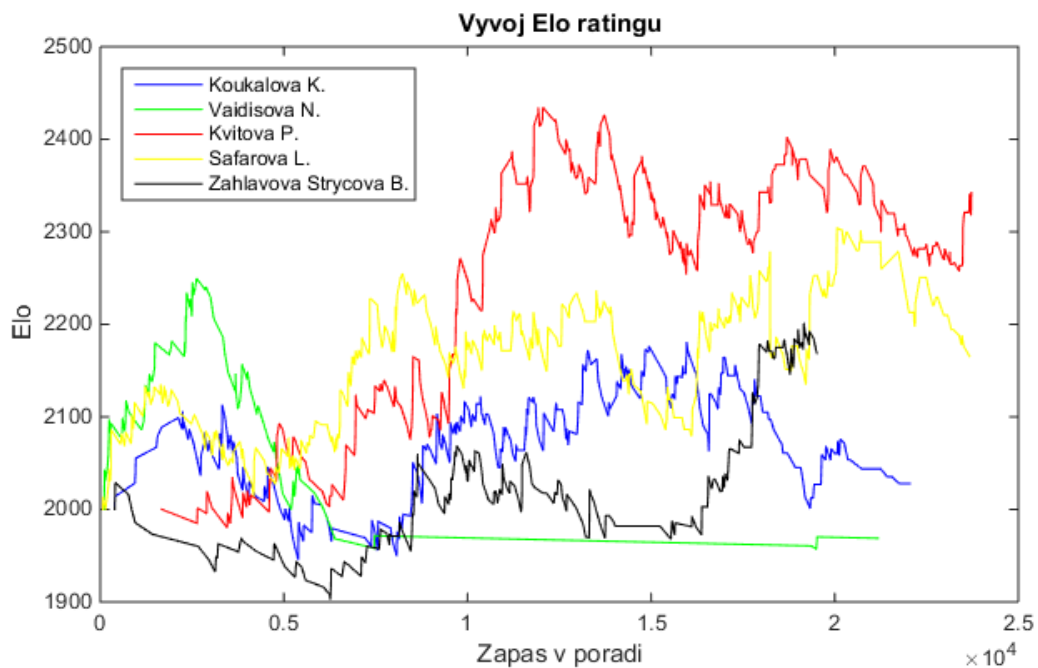


## 8.2 Ženy

Na obrázcích 23 a 24 je zobrazen vývoj Elo ratingu vybraných světových (resp. českých) hráček od začátku používání Elo modelu 3, tedy od prvního zápasu v roce 2007 do posledního zápasu v roce 2016.



Obrázek 23: Elo rating ženy světové hráčky



Obrázek 24: Elo rating ženy české hráčky

Z obrázku 23 lze vyčíst, že na začátku modelu (v roce 2007) byla výkonnost vybraných světových hráček, až na hráčku Caroline Wozniacki, vyrovnaná. V roce 2016 ale jasně dominuje Serena Williams. To dokládá i tabulka 11, ve které je uvedeno 15 tenistek s největším Elo ratingem na konci roku 2016. Serena Williams měla na konci roku 2016 Elo rating ve výši 2661. Téměř 200 Elo rating bodů na ní ztrácela druhá Maria Sharapova s Elo ratingem ve výši 2469. Na třetím místě byla Justin Henin s Elo ratingem ve výši 2444.

V tabulce tenistek s nejvyšším Elo ratingem na konci roku 2016 se nacházejí i tenistky, které už v té době ukončily kariéru, protože se Elo rating hráček, které nehrají zápasy, nemění. Tudíž jim nemůže Elo rating klesnout.

Pořadí	Jméno	Elo rating	Průměrné Elo soupeřů	Počet zápasů s Elo hráčkami	Počet výher
1	Williams S.	2661	2197	427	378
2	Sharapova M.	2469	2189	382	300
3	Henin J.	2444	2138	128	112
4	Azarenka V.	2439	2178	443	336
5	Li N.	2433	2166	349	240
6	Kerber A.	2431	2166	359	216
7	Clijsters K.	2418	2172	118	96
8	Cibulkova D.	2376	2149	421	256
9	Halep S.	2355	2188	261	165
10	Dementieva E.	2348	2142	242	178
11	Kvitova P.	2342	2172	364	232
12	Wozniacki C.	2328	2155	551	391
13	Radwanska A.	2317	2168	544	365
14	Kuznetsova S.	2307	2164	419	267
15	Pennetta F.	2301	2126	429	265

Tabulka 11: Elo rating Top 15 ženy

Tenistkou s nejnižším Elo ratingem na konci roku 2016 byla na 418. místě Cagla Buyukakcay, která měla Elo rating ve výši 1790. Zbylé hráčky žádný Elo rating neměly.

Kompletní žebříček Elo ratingu hráček z konce roku 2016 lze nalézt v příloženém excelovském souboru *Elo rating žebříček 2016.xlsx*.

## 9 Ověření modelu pomocí sázení

Sport je zábavou pro spoustu lidí po celém světě. V poslední době se stal sport velkým byznysem. Jednou částí tohoto byznysu je také kurzové sázení. Kromě klasického sázení na pobočce sázkové kanceláře sází spousta sázkařů i na internetu. Díky rozmachu výpočetní techniky a internetu je nyní sázení jednodušší a rychlejší, než bylo dříve.

Cílem sázejících je výhra, která je vyplácena u sázkových kanceláří v penězích. Proto je důležité vybrat „správné“ hráče nebo týmy, na které vsadit.

Přístup při výběru sázkových příležitostí je různý. Lidé často spoléhají na svoji intuici či sází na svůj oblíbený tým. Dále lidé sází podle získaných informací, což může být například informace o aktuální formě hráče či týmu, zranění hráče atd. Další možností při výběru sázkových příležitostí je použití statistických a matematických modelů, které rovněž využívají sázkové kanceláře pro stanovení úvodních kurzů.

V kapitole 7 byly představeny tři Elo modely, jako nástroje pro odhad výsledků sportovních utkání. Byla zde také provedena optimalizace parametrů tří Elo modelů na základě kritérií určujících kvalitu modelu a modely byly poté mezi sebou porovnány. Nejlépe z těchto tří modelů dopadl u mužů i žen Elo model 3. V této kapitole dojde k ověření Elo modelu 3 pomocí sázení na tenisové zápasy mužů a žen v sezónách 2015 a 2016.

### 9.1 Základní pojmy

Následující část čerpá základní popisy z Cover a Thomas (2006) a Špaček (2015).

V každém tenisovém utkání jsou možné dva výsledky. S pravděpodobností  $p_1$  vyhraje hráč 1 a s pravděpodobností  $p_2$  vyhraje hráč 2. Samozřejmě musí platit, že

$$p_1 + p_2 = 1. \quad (9.1)$$

Skutečné pravděpodobnosti  $p_1$  a  $p_2$  výsledku zápasu jsou však neznámé, jak pro sázkaře, tak pro sázkovou kancelář.

Cílem sázkaře je odhadnout pravděpodobnost  $p_1$  pravděpodobností  $r_1$  a pravděpodobnost  $p_2$  pravděpodobností  $r_2$  co nejpřesněji. Čím lepší odhad sázkař má, tím má větší šanci na zisk při krátkodobém i dlouhodobém sázení. Sázkař může odhadovat tyto pravděpodobnosti například svými „tenisovými znalostmi“ či pomocí matematických a statistických modelů, jako jsou například Elo modely.

Sázková kancelář odhaduje pravděpodobnost  $p_1$  pravděpodobností  $q_1$  a pravděpodobnost  $p_2$  pravděpodobností  $q_2$ . Na základě těchto odhadnutých pravděpodobností  $q_i$  vypisují sázkové kanceláře kurzy  $o_i$  na sportovní utkání. Pro sázkovou kancelář je ideální, pokud sázející sází tak, že při obou výsledcích zápasů

vyplatí sázejícím stejnou částku. V takovém případě je jisté, že sázková kancelář bude v zisku díky marži  $\zeta$ .

Kurz  $o_i$  je roven zisku sázkaře při úspěšné sázce, pokud vsadí 1 jednotku. Kurz, který vypisuje sázková kancelář, je počítán dle následujícího vzorce

$$o_i = \frac{1 - \zeta}{q_i}. \quad (9.2)$$

Bude zde uveden příklad vytváření kurzů a výplat sázejícím. Sázková kancelář odhaduje, že hráč 1 vyhraje s pravděpodobností 60 % a hráč 2 vyhraje s pravděpodobností 40 %. Zároveň chce mít sázková kancelář marži 10 %.

V tomto případě bude vypsán kurz  $o_1$  na vítězství hráče 1

$$o_1 = \frac{1 - 0,1}{0,6} = 1,5. \quad (9.3)$$

V tomto případě bude vypsán kurz  $o_2$  na vítězství hráče 2

$$o_2 = \frac{1 - 0,1}{0,4} = 2,25. \quad (9.4)$$

Sázející chtějí dohromady na tento zápas vsadit 100 jednotek. Pro sázkovou kancelář je ideální, pokud sázející vsadí 60 jednotek na hráče 1 a 40 jednotek na hráče 2.

Výsledek	1	2
Kurz	1,5	2,25
Vsazeno	60	40
Výplata	90	90

Tabulka 12: Ideální rozložení sázek

V tomto případě je sázkové kanceláři jedno, jak zápas dopadne. Vždy sázejícím vyplatí 90 jednotek, a tak zisk sázkové kanceláře bude 10 jednotek.

Sázející mohou ale sázet i v jiném poměru, například 80 jednotek vsadí na hráče 1 a 20 jednotek na hráče 2.

Výsledek	1	2
Kurz	1,5	2,25
Vsazeno	80	20
Výplata	120	45

Tabulka 13: Jiné rozložení sázek

Jestliže vyhraje hráč 1, tak bude sázková kancelář ve ztrátě 20 jednotek, ale pokud vyhraje hráč 2, tak bude sázková kancelář v zisku 55 jednotek. Ve střední hodnotě bude výplata 90 jednotek, což je stejné jako v předcházejícím případě

$$E(X) = 0,6 \cdot 120 + 0,4 \cdot 45 = 90. \quad (9.5)$$

V reálném světě se sice jedno utkání neopakuje, ale každý den se hraje mnoho různých zápasů. Díky tomu sázkové kanceláře dlouhodobě vytvářejí zisk.

## 9.2 Výběr zápasů

Důležitým faktorem při sázení je správný výběr zápasů, na které má sázející vsadit. Pokud sázkař používá k sázení matematické a statistické modely, kterými odhadne pravděpodobnosti vítězství hráčů  $r_i$ , může k výběru zápasů použít například následující kritérium

$$o_i \cdot r_i > R, \quad (9.6)$$

kde jsou

$o_i$  kurz,

$R$  koeficient sázení.

Koeficient sázení  $R$  by udával minimální střední hodnotu výhry za předpokladu, že by odhadnuté pravděpodobnosti vítězství hráčů  $r_i$  byly rovny skutečným, ale neznámým pravděpodobnostem vítězství hráčů  $p_i$ . Z tohoto důvodu se volí koeficient sázení  $R$  alespoň 1,0. Může být však i vyšší. Čím vyšší bude, tím se bude sázet na menší počet zápasů.

## 9.3 Strategie sázení

Při sázení je kromě výběru zápasu a hráče, na kterého vsadit, důležitá také otázka, jakou částku by měl sázející vsadit. Existují různé strategie sázení. Například může vždy hráč vsadit všechny svoje prostředky. Toto však není rozumná strategie, protože dříve nebo později sázkař nějakou sázku prohraje. Tím pádem by přišel o všechny své prostředky a dále už by nemohl sázet.

V této práci budou použity dvě strategie: konstantní vklad a procento z banku.

### 9.3.1 Konstantní vklad

Při této strategii sázkař vsadí na každý vybraný zápas stejnou částku, například 10 jednotek. Výhodou této strategie je, že sázející nemusí počítat, kolik má vsadit, ale vždy sází stejně. Druhou výhodou je, že každý zápas má na celkový zisk stejný vliv, protože je na něj vsazena stejná částka. Nevýhodou při reálném sázení je, že sázející se teoreticky může dostat do hodně velké ztráty, proto je zde nutný předpoklad, že sázkař má k dispozici neomezené prostředky, resp. má takové prostředky, aby mohl vsadit na každý zápas z porovnávané databáze a přitom ani jednou nevyhrát.

### 9.3.2 Procento z banku

Druhou strategií je vklad jako procento z banku. To znamená, že na začátku si sázející určí velikost banku, tedy kolik peněz chce investovat do sázení. To může být například 100 jednotek. Dále si určí, kolik procent z banku bude chtít dávat sázkař na jednotlivé sázky, například 1 %. V tomto případě vsadí sázkař na první zápas 1 jednotku. Pokud

prohraje, zbyde mu ještě 99 jednotek. Na další zápas pak vsadí 0,99 jednotky. Výhodou tohoto systému je, že sázkař může být ve ztrátě maximálně ve velikosti banku a sázkař teoreticky může sázet pořád, protože bude stále mít něco v banku. V realitě to však nepůjde, protože sázkové kanceláře požadují minimální vklad, například sázková kancelář Fortuna požaduje minimální vklad 10 Kč.

## 9.4 Elo model 3 pro sázení

V úvodu této kapitoly bylo řečeno, že se bude sázet podle Elo modelu 3 s optimalizovanými parametry.

Jak bylo popsáno v kapitole 7.3, tento model pro odhad úvodního Elo ratingu hráče používá metodu kvantilů popsanou v kapitole 5.1.2. Parametr  $l$ , udávající kolik zápasů musí tenista odehrát, než mu je vypočten úvodní Elo rating, je optimalizován u mužů na hodnotu  $l = 23$  a u žen na hodnotu  $l = 27$ .

Metodu kvantilů lze použít pouze v případě, že hráč bez odhadnutého Elo ratingu hraje proti hráčům, kteří už nějaký odhadnutý Elo rating mají. Proto je pro odhad úvodního Elo ratingu hráče v prvních dvou sezónách použita metoda konstantního Elo ratingu popsaná v kapitole 5.1.1.

Pro přepočítávání Elo ratingu hráče je v tomto modelu zvolen konstantní koeficient rozvoje  $K$  popsaný v kapitole 5.2.1. Koeficient rozvoje  $K$  je optimalizován u mužů na hodnotu  $K = 19$  a u žen na hodnotu  $K = 29$ .

## 9.5 Program k ověření modelu

Ověření modelu pomocí sázení je provedeno v Matlabu ve verzi R2014b. K ověření modelu je vytvořen a přiložen program *Overeni\_modelu\_sazeni.m*. V první části tohoto programu si uživatel vybere, zda chce sázet na zápasy mužů či žen (způsob nastavení je ukázán v příloze B). Poté se data načtou. V druhé části je použit Elo model 3 k počítání Elo ratingu hráčů a následně se z Elo ratingů vypočtou pravděpodobnosti vítězství hráčů v zápasech, jak bylo popsáno v kapitole 4.3. V třetí části programu je pak samotné sázení. Na základě kurzů sázkových kanceláří, zvoleném koeficientu  $R$  a vypočtených pravděpodobností vítězství hráčů v zápasech se vyberou zápasy, na které se vsadí. Poté je počítán vývoj banku pro 2 strategie sázení: konstantní vklad (kapitola 9.3.1) a procento z banku (kapitola 9.3.2). Na závěr je provedeno sázení pomocí jednoduchých modelů popsaných v kapitole 9.8.

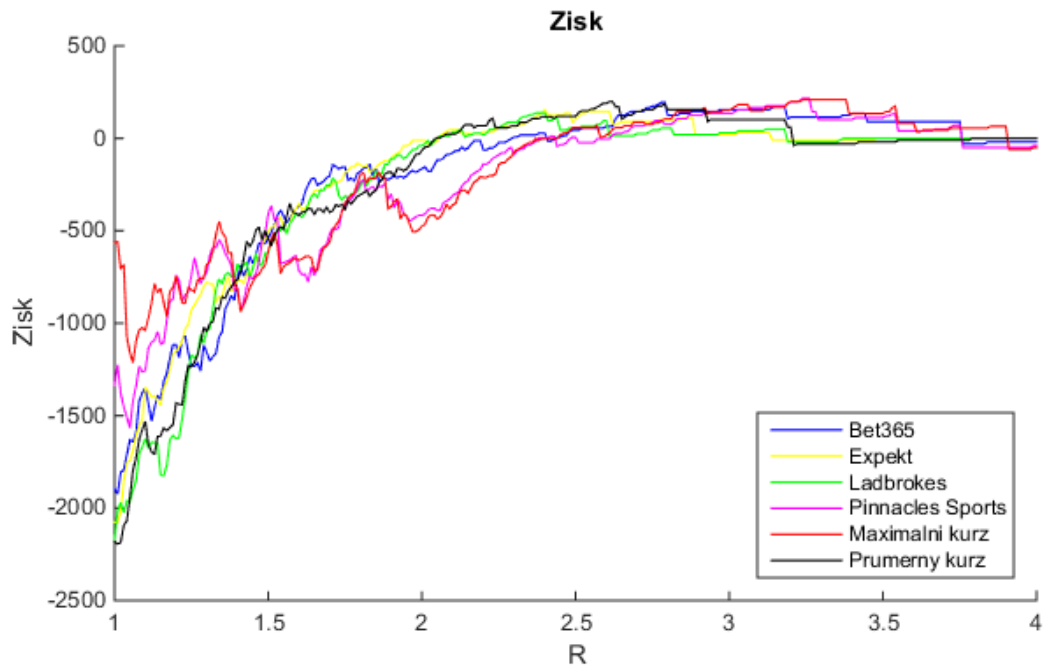
Sázení na zápasy je počítáno také v přiloženém excelovském souboru *OvěřeníModeluSazeniEloModel3.xlsx*, kam jsou vloženy pravděpodobnosti vítězství hráčů v zápase. Následně jsou v Excelu vybrány zápasy, na které se vsadí, a je zde vypočtena velikost zisku.

## 9.6 Sázení muži

### 9.6.1 Sázení konstantní vklad

V této kapitole je provedeno sázení na zápasy mužů v letech 2015 a 2016. Sáží se dle strategie konstantního vkladu, popsané v kapitole 9.3.1. To znamená, že na každý zápas, na který se vsadí, je vsazena stejná částka, a to konkrétně 10 jednotek.

Na obrázku 25 je zobrazen stav banku po dvou letech sázení, tedy na konci sezóny 2016 v závislosti na hodnotě koeficientu sázení  $R$ .



Obrázek 25: Velikost zisku v závislosti na  $R$  konstantní vklad muži

Z grafu lze vyčíst, že pro hodnoty koeficientu sázení  $R$  mezi 1 a 2 je sázející na konci sázení ve větší či menší ztrátě. Pro některé vyšší hodnoty koeficientu sázení  $R$  je však sázkař na konci sázení v mírném zisku.

Maximální hodnota banku na konci sázení je 216 jednotek pro hodnotu koeficientu sázení  $R = 3,25$  u sázkové kanceláře Pinnacles Sports. Tato hodnota koeficientu sázení  $R$  je však extrémní, protože za celé dvě sezóny je vsazeno pouze na 15 utkání. Z toho 4 utkání jsou pro sázkaře výherní. Vzhledem k tomu, že uvedená utkání měla vysoký kurz, bylo dosaženo zisku.

V tabulce 14 je porovnání sázek u sázkové kanceláře Pinnacles Sports při různých hodnotách koeficientu sázení  $R$ .

R	Počet sázek	Vítězných sázek	Vsazená částka	Vyhraná částka	Zisk
1,00	1 665	692	16 650	15 317	-1 334
1,25	651	161	6 510	5 769	-741
1,50	312	59	3 120	2 713	-407
1,75	177	27	1 770	1 402	-369
2,00	104	10	1 040	624	-416
2,25	65	8	650	561	-89
2,50	44	6	440	446	6
2,75	32	5	320	407	87
3,00	23	4	230	366	136
3,25	15	4	150	366	216
3,50	12	2	120	232	112
3,75	6	1	60	130	70
4,00	4	0	40	0	-40

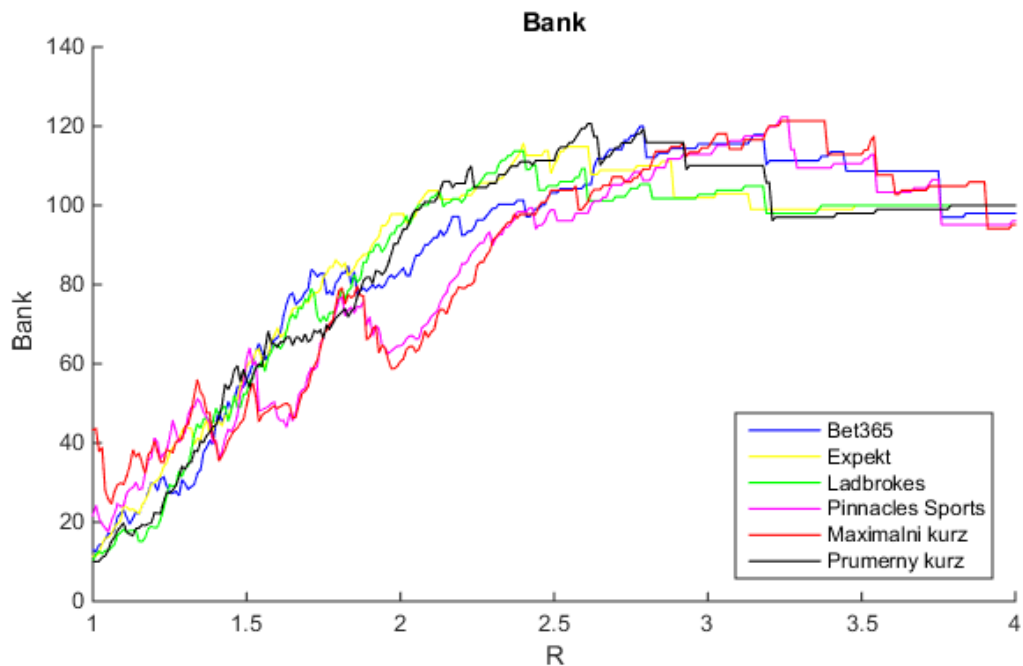
Tabulka 14: Porovnání koeficientu sázení R konstantní vklad muži

Z tabulky je patrné, že zisku se dosahuje při koeficientu sázení  $R$  mezi 2,5 a 3,75.

### 9.6.2 Sázení procento z banku

V této kapitole se bude sázet na zápasy mužů v letech 2015 a 2016 podle strategie procento z banku. To znamená, že na začátku sázení si zvolí sázkař bank, který použije k sázení. V tomto případě je bank o velikosti 100 jednotek. Poté sázkař sází na každý vybraný zápas nějaké předem dané procento z aktuální výše banku.

Na obrázku 26 je zobrazen stav banku po dvou letech sázení, tedy na konci sezóny 2016 v závislosti na hodnotě koeficientu sázení  $R$  a při sázení 1 % banku.



Obrázek 26: Velikost banku v závislosti na R procento z banku muži



Nejvyššího zisku je opět dosaženo pro extrémní hodnotu koeficientu sázení  $R = 3,25$  při sázení u sázkové kanceláře Pinnacles Sports. V tomto případě je na konci sázení bank o velikosti 122,32 jednotek. Zisk je tedy 22,32 jednotek. Vsazeno je opět na 15 utkání, protože výběr utkání je nezávislý na strategii sázení, a 4 utkání jsou výherní. Během doby sázení je vsazeno 17,54 jednotek a vyhráno 39,86 jednotek.

Zisk ze sázení kromě výběru zápasů je také závislý na zvoleném procentu, které udává, kolik procent z banku se bude na vybrané zápasy sázet. To je ukázáno v tabulce 15, kde je vyjádřen zisk při sázení u sázkové kanceláře Pinnacles Sports pro různá procenta a hodnoty koeficientu sázení  $R$ , když na začátku je bank 100 jednotek.

R/Procenta	1 %	2 %	5 %	10 %	20 %	50 %	100 %
1,00	-78	-97	-100	-100	-100	-100	-100
1,25	-58	-86	-100	-100	-100	-100	-100
1,50	-39	-68	-98	-100	-100	-100	-100
1,75	-34	-59	-93	-100	-100	-100	-100
2,00	-36	-60	-92	-100	-100	-100	-100
2,25	-10	-23	-58	-91	-100	-100	-100
2,50	-1	-5	-28	-68	-98	-100	-100
2,75	7	12	11	-22	-83	-100	-100
3,00	13	24	45	39	-37	-100	-100
3,25	22	46	119	224	273	-73	-100
3,50	11	20	40	47	3	-96	-100
3,75	7	12	24	30	11	-78	-100
4,00	-4	-8	-19	-34	-59	-94	-100

Tabulka 15: Zisk v závislosti na  $R$  a procentech mužů

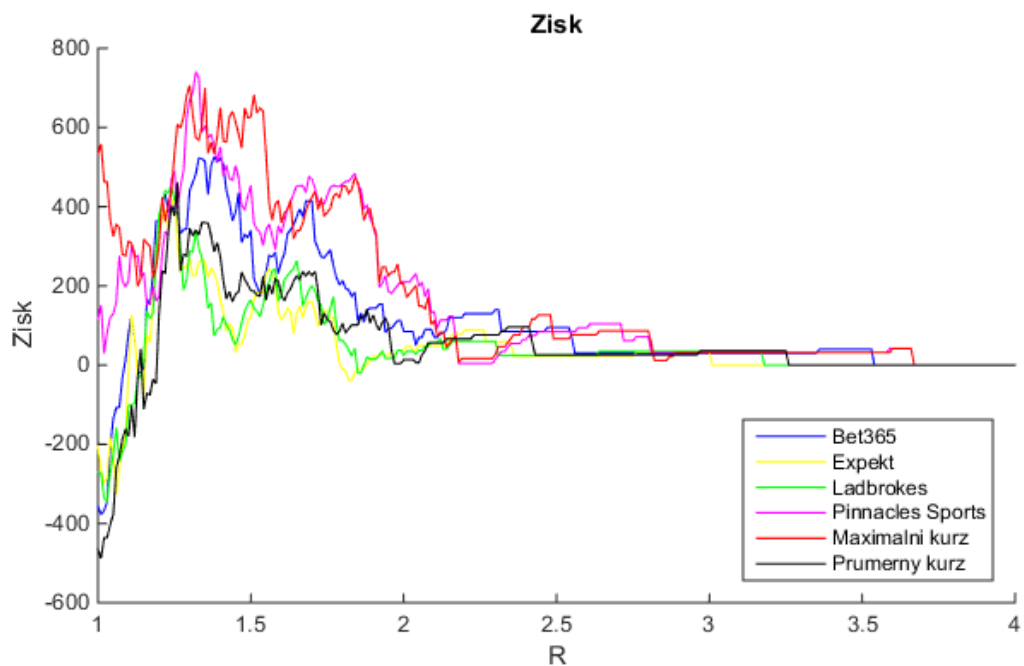
Z tabulky 15 lze vypočítat, že je špatné používat procenta 50 % či dokonce 100 %. Při vyšších procentech z ostatních procent (např. 20 %) se lze dostat k vyšším ziskům. Na druhé straně je tu daleko větší citlivost na hodnotu koeficientu sázení  $R$ .

## 9.7 Sázení ženy

### 9.7.1 Sázení konstantní vklad

V této kapitole se sází dle modelu na zápasy žen v letech 2015 a 2016. Sází se dle strategie konstantního vkladu, popsané v kapitole 9.3.1. To znamená, že na každý zápas, na který se vsadí, bude vsazena stejná částka, a to konkrétně 10 jednotek.

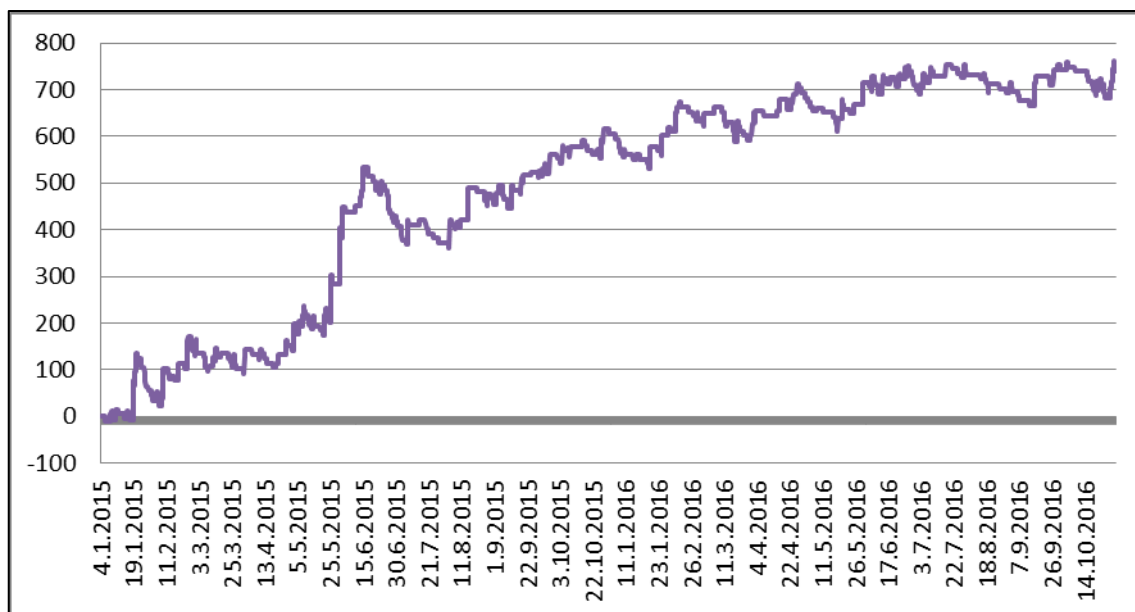
Na obrázku 27 je zobrazen stav banku po dvou letech sázení, tedy na konci sezóny 2016 v závislosti na hodnotě koeficientu sázení  $R$ .



Obrázek 27: Velikost zisku v závislosti na R konstantní vklad ženy

Z grafu lze vypočítat, že sázkář, který sází u sázkové kanceláře Pinnacles Sports či na maximální kurz, dosahuje zisku i pro hodnoty koeficientu sázení  $R$  mezi 1 a 2. Toto je rozdíl oproti sázení v kategorii mužů, kde byl model pro hodnoty koeficientu sázení  $R$  mezi 1 a 2 vždy ve ztrátě. Navíc je v tomto případě zisk i relativně vysoký. Maximálního zisku dosáhne sázkář u sázkové kanceláře Pinnacles Sports při hodnotě koeficientu sázení  $R = 1,32$ . V tomto případě je zisk 740,6 jednotek. Celkově je v tomto případě vsazeno na 382 zápasů, z toho 134 sázek je výherních.

Na obrázku 28 je zobrazen vývoj zisku v čase od začátku sezóny 2015 do konce sezóny 2016 při sázení u sázkové kanceláře Pinnacles Sports při hodnotě koeficientu sázení  $R = 1,32$ .



Obrázek 28: Vývoj zisku konstantní vklad ženy

V tabulce 16 je porovnání sázek u sázkové kanceláře Pinnacles Sports při různých hodnotách koeficientu sázení  $R$ .

$R$	Počet sázek	Vítězných sázek	Vsazená částka	Vyhraná částka	Zisk
1,00	1 420	682	14 200	14 317	117
1,25	504	178	5 040	5 530	490
1,50	193	63	1 930	2 383	453
1,75	88	32	880	1 308	428
2,00	34	11	340	551	211
2,25	15	3	150	155	5
2,50	7	3	70	155	85
2,75	4	2	40	102	62
3,00	2	1	20	52	32

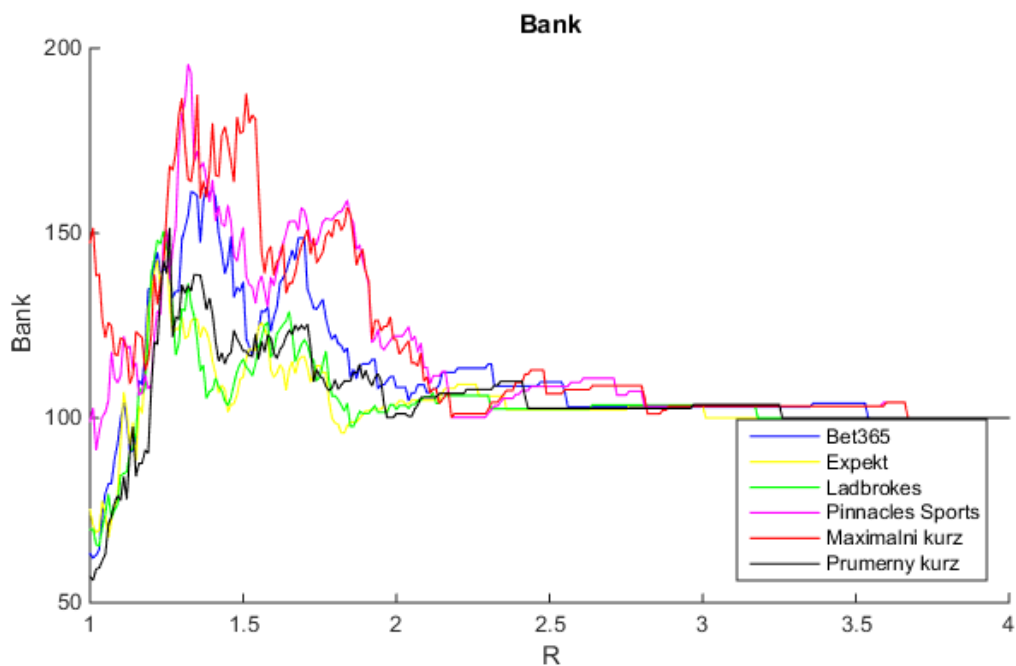
Tabulka 16: Porovnání koeficientu sázení  $R$  konstantní vklad ženy

Pro všechny hodnoty koeficientu sázení  $R$  je sázkař, který sází u sázkové kanceláře Pinnacles Sports, v zisku, což je velký rozdíl oproti sázení na zápasy mužů.

### 9.7.2 Sázení procento z banku

V této kapitole se bude sázet na zápasy žen v sezóně 2015 a 2016 podle strategie procento z banku. To znamená, že na začátku si zvolí sázkař bank, který použije na sázení. V tomto případě je bank o velikosti 100 jednotek. Poté sází na každý vybraný zápas nějaké předem dané procento z aktuální výše banku.

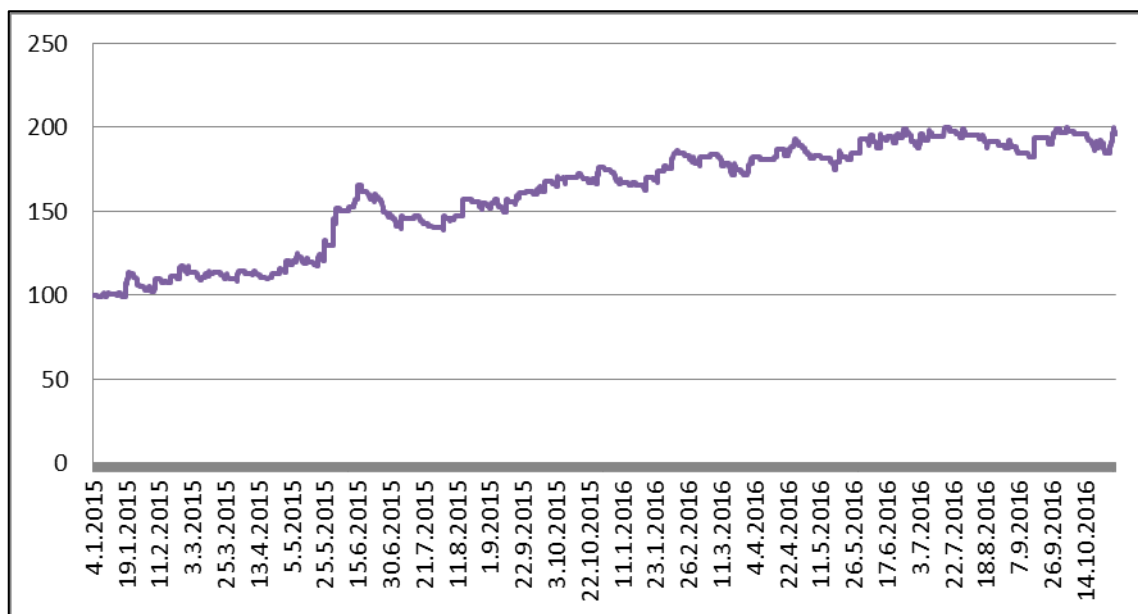
Na obrázku 29 je zobrazen stav banku po dvou letech sázení, tedy na konci sezóny 2016 v závislosti na hodnotě koeficientu sázení  $R$  při sázení 1 % banku.



Obrázek 29: Velikost banku v závislosti na R procento z banku ženy

Nejvyššího zisku je opět dosaženo pro hodnotu koeficientu sázení  $R = 1,32$  při sázení u sázkové kanceláře Pinnacles Sports. V tomto případě je na konci sázení bank o velikosti 195,67 jednotek. Zisk je tedy 95,67 jednotek. Vsazeno je na 382 utkání, z toho 134 utkání je výherních. Celkově bylo prosázeno 596,90 jednotek a vyhrálo se 692,57 jednotek.

Na obrázku 30 je zobrazen vývoj banku v čase od začátku sezóny 2015 do konce sezóny 2016 při sázení u sázkové kanceláře Pinnacles Sports při hodnotě koeficientu sázení  $R = 1,32$ .



Obrázek 30: Vývoj banku při strategii procento z banku ženy

V tabulce 17 je vyjádřen zisk při sázení u sázkové kanceláře Pinnacles Sports pro různá procenta a hodnoty koeficientu sázení  $R$ , když na začátku je bank 100 jednotek.

R/Procenta	1 %	2 %	5 %	10 %	20 %	50 %	100 %
1,00	-1	-23	-91	-100	-100	-100	-100
1,25	51	96	94	-82	-100	-100	-100
1,50	52	114	308	293	-88	-100	-100
1,75	50	116	410	1046	823	-100	-100
2,00	22	46	124	240	229	-98	-100
2,25	0	0	-5	-20	-58	-99	-100
2,50	9	17	43	86	151	82	-100
2,75	6	12	31	60	111	131	-100
3,00	3	6	15	28	47	55	-100

Tabulka 17: Zisk v závislosti na  $R$  a procentech ženy

Z tabulky 17 lze vyzorovat, že opět není rozumné používat 50 % či 100 %. Při použití 1 % je sázkař až na jeden případ, kdy ztratí pouze 1 jednotku, vždy v zisku. Na druhou stranu nedosahuje takového zisku jako v ostatních případech.

## 9.8 Jednoduché modely

Pro demonstrování účinnosti Elo modelu 3 při sázení jsou vytvořeny tři jednoduché modely: Favorit, Outsider a Náhodný.

### 9.8.1 Princip modelů

Princip fungování těchto modelů je jednoduchý. Nejdříve je třeba určit, na které zápasy je možno vsadit. V tomto případě to jsou všechny zápasy mužů (resp. žen) v letech 2015 a 2016.

Poté je třeba u jednotlivých zápasů určit, zda se na zápas vsadí nebo nevsadí. To je generováno náhodně podle alternativního rozdělení. Parametr alternativního rozdělení  $p$  je vypočten dle následujícího vzorce

$$p = \frac{v}{n}, \quad (9.7)$$

kde jsou

$v$  počet zápasů, na které bylo vsazeno Elo modelem v závislosti na parametru  $R$ ,  
 $n$  celkový počet zápasů v letech 2015 a 2016.

Cílem je, aby bylo jednoduchými modely vsazeno na podobný počet zápasů jako v případě Elo modelu 3.

Pokud je rozhodnuto, že se na konkrétní zápas bude sázet, zbývá ještě určit, na kterého hráče se vsadí. V tomto případě záleží na volbě modelu. V modelu Favorit (resp. Outsider) se vsadí na hráče s nižším (resp. vyšším) kurzem. Pokud je kurz vypsaný sázkovou kanceláří na oba hráče stejný, rozhodne se model náhodně. V modelu Náhodný je výběr hráče, na kterého se vsadí, vždy náhodný.

## 9.8.2 Muži

V této části budou porovnány výsledky sázení podle Elo modelu 3 a jednoduchých modelů v kategorii mužů u sázkové kanceláře Pinnacles Sports. Sáží se dle strategie konstantního vkladu.

V závislosti na volbě koeficientu sázení  $R$  je vypočítán zisk při sázení podle Elo modelu 3 a určen počet zápasů, na které se vsadilo. Následně je vygenerováno a provedeno 1000 sázení podle jednoduchých modelů. Poté je vypočten průměrný zisk a výběrová směrodatná odchylka (v tabulce označená Std.) zisku při sázení pomocí jednoduchých modelů. Výsledky jsou zaznamenány v tabulce 18.

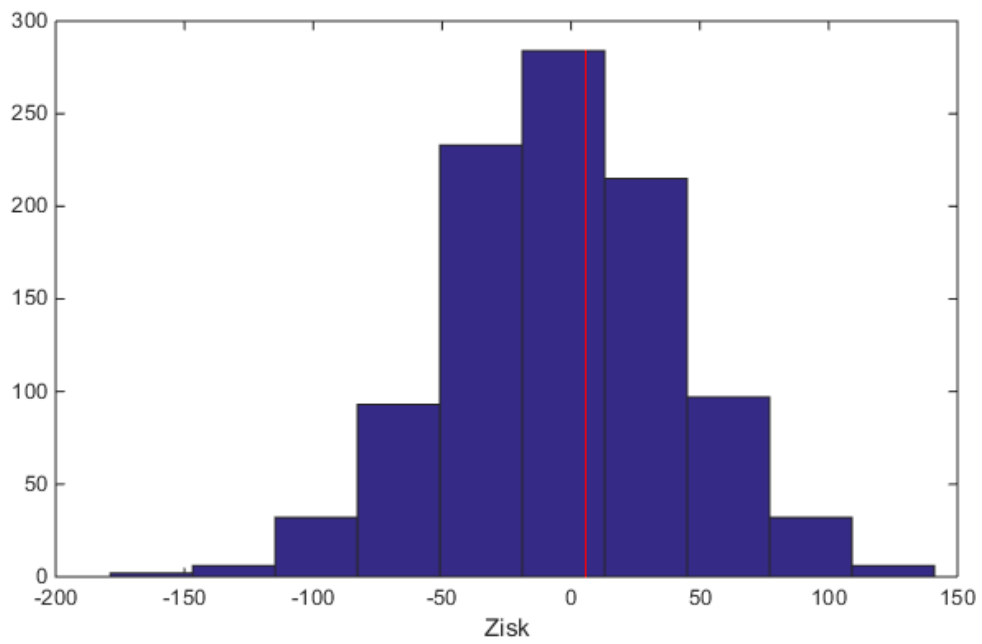
R	Elo zisk	Počet sázek	Favorit Průměr	Favorit Std.	Outsider Průměr	Outsider Std.	Náhodný Průměr	Náhodný Std.
1,00	-1 334	1 665	-115	225	-1 529	613	-835	520
1,25	-741	651	-38	156	-606	430	-333	345
1,50	-407	312	-19	111	-295	299	-146	234
1,75	-369	177	-17	87	-151	242	-82	185
2,00	-416	104	-11	64	-90	175	-53	142
2,25	-89	65	2	53	-67	149	-41	106
2,50	6	44	-3	44	-38	122	-19	96
2,75	87	32	-2	39	-33	101	-16	73
3,00	136	23	-1	30	-23	82	-10	64
3,25	216	15	-2	27	-12	72	-6	53
3,50	112	12	-1	22	-14	61	-6	44
3,75	70	6	0	16	-5	45	-3	36
4,00	-40	4	-1	13	-3	35	-3	25

Tabulka 18: Porovnání Elo modelu s jednoduchými modely mužů

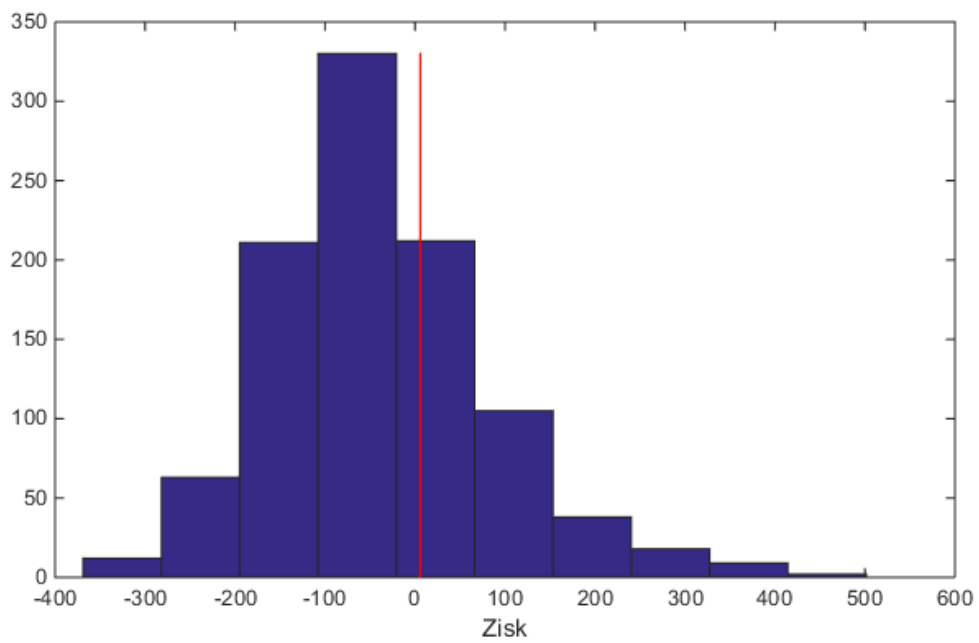
V kategorii mužů se Elo modelu dařilo v porovnání s jednoduchými modely, pokud byl zvolen vyšší koeficient sázení  $R$  (2,5 a vyšší). V tomto případě byl Elo model v zisku, zatímco jednoduché modely byly v mírné ztrátě. Naopak pokud byl volen koeficient sázení  $R$  mezi 1 a 2, tak se Elo modelu příliš nedařilo. Dosahoval vyšších ztrát než jednoduché modely.

Relativně dobře si vedl v souboji se sázkovou kancelářím model Favorit. Ten dosahoval pouze relativně malých ztrát.

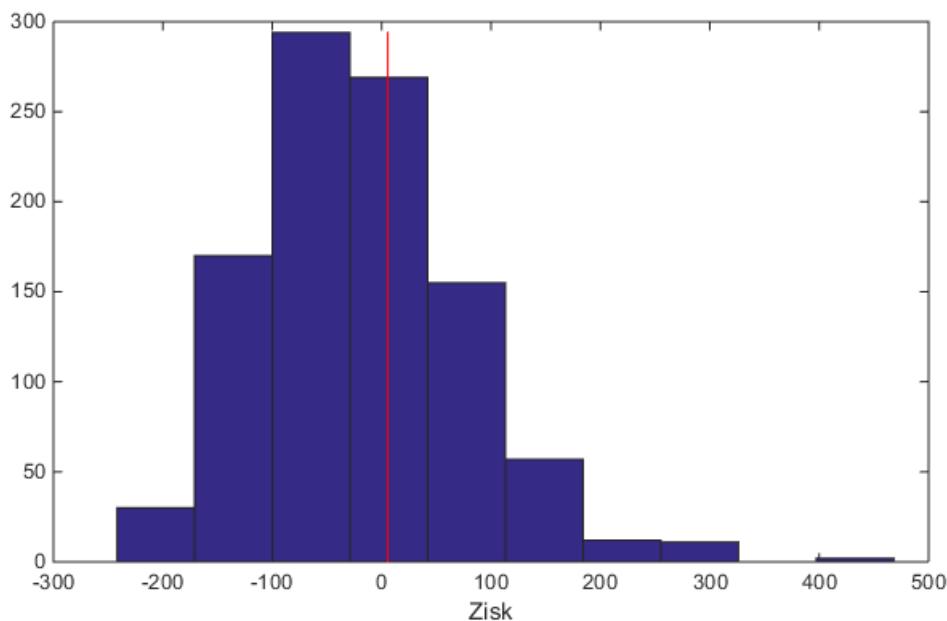
Na obrázcích 31, 32 a 33 jsou ukázány histogramy při sázení pomocí modelů Favorit, Outsider či Náhodný pro koeficient sázení  $R = 2,5$ . Červenou čarou je znázorněn zisk při sázení pomocí Elo modelu.



Obrázek 31: Histogram model Favorit muži



Obrázek 32: Histogram model Outsider muži



Obrázek 33: Histogram model Náhodný muži

### 9.8.3 Ženy

V této části budou porovnány výsledky sázení podle Elo modelu 3 a jednoduchých modelů v kategorii žen u sázkové kanceláře Pinnacles Sports. Sází se dle strategie konstantního vkladu.

V závislosti na volbě koeficientu sázení  $R$  je vypočítán zisk při sázení podle Elo modelu 3 a určen počet zápasů, na které se vsadilo. Následně je vygenerováno a provedeno 1000 sázení podle jednoduchých modelů. Poté je vypočten průměrný zisk a výběrová směrodatná odchylka zisku při sázení pomocí jednoduchých modelů. Výsledky jsou zaznamenány v tabulce 19.

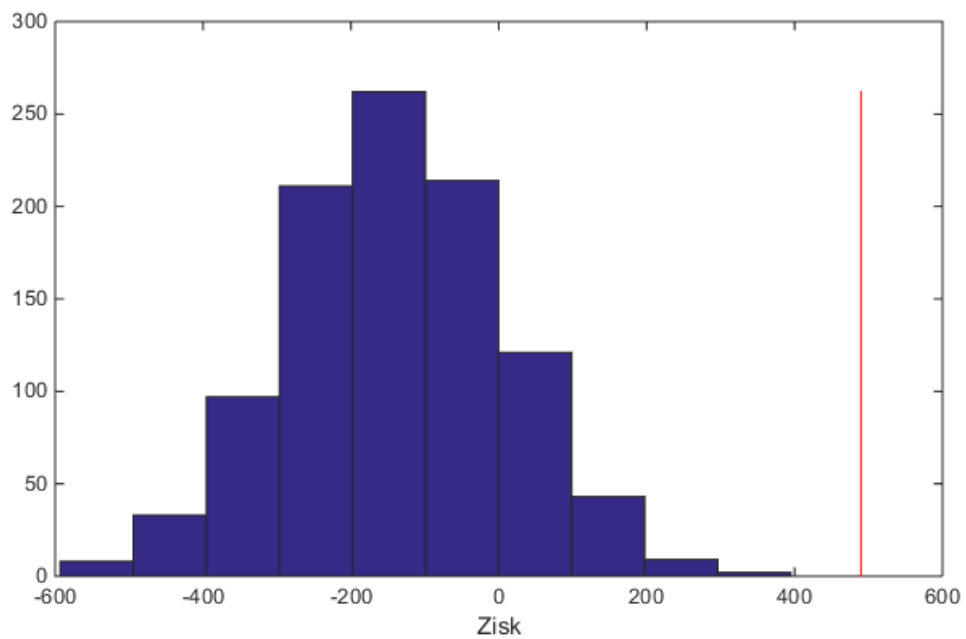
R	Elo zisk	Počet sázek	Favorit Průměr	Favorit Std.	Outsider Průměr	Outsider Std.	Náhodný Průměr	Náhodný Std.
1,00	117	1 420	-378	217	-699	474	-545	458
1,25	490	504	-140	146	-249	326	-198	261
1,50	453	193	-51	99	-92	215	-75	173
1,75	428	88	-23	64	-43	145	-36	113
2,00	211	34	-9	41	-18	89	-11	70
2,25	5	15	-5	27	-5	61	-8	44
2,50	85	7	-2	19	-3	41	-4	31
2,75	62	4	-1	14	-3	29	-1	23
3,00	32	2	-1	10	-1	22	-2	18

Tabulka 19: Porovnání Elo modelu s jednoduchými modely ženy

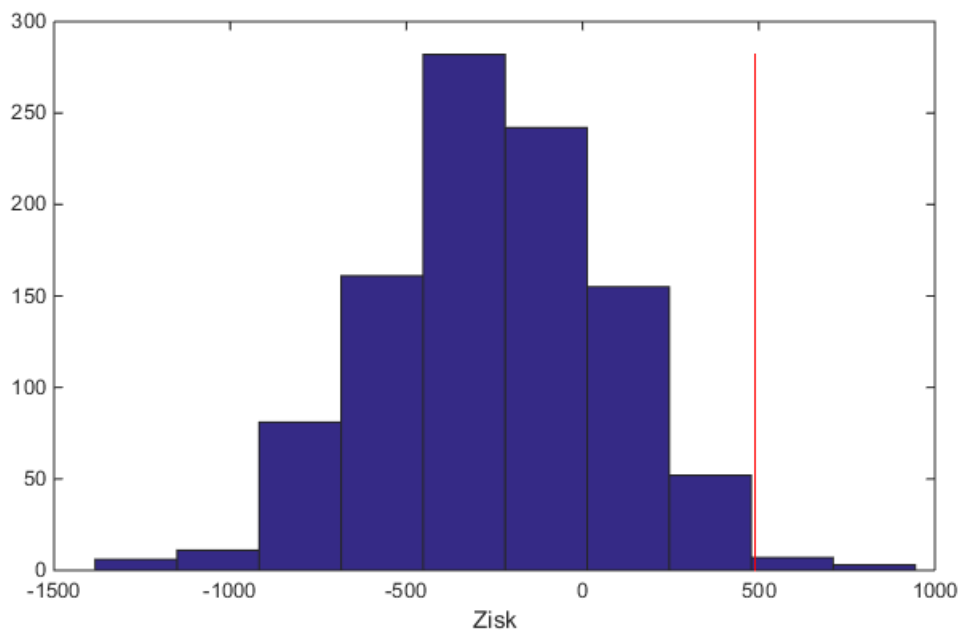
Při sázení na zápasy žen si Elo model 3 v porovnání s jednoduchými modely vedl velmi dobře. Elo model 3 byl vždy v zisku. Naopak jednoduché modely byly vždy ve ztrátě.



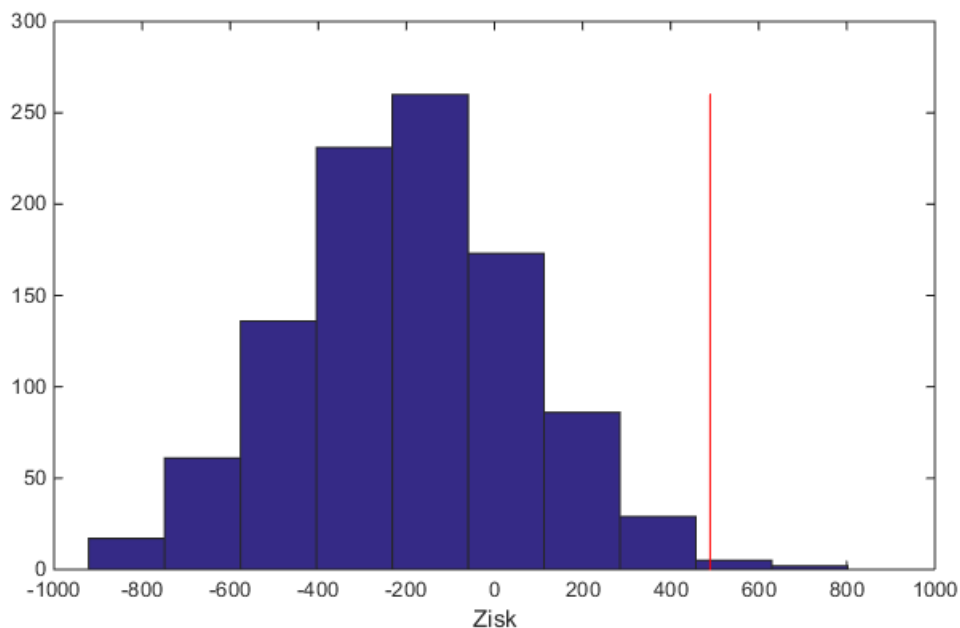
Na obrázcích 34, 35 a 36 jsou ukázány histogramy při sázení pomocí modelů Favorit, Outsider či Náhodný pro koeficient sázení  $R = 1,25$ . Červenou čarou je znázorněn zisk při sázení pomocí Elo modelu.



Obrázek 34: Histogram model Favorit ženy



Obrázek 35: Histogram model Outsider ženy



Obrázek 36: Histogram model Náhodný ženy

## 9.9 Shrnutí

Více úspěšný byl Elo model při sázení na zápasy žen než při sázení na zápasy mužů.

U mužů byl Elo model v mírném zisku pouze pro vyšší hodnoty koeficientu sázení  $R$ . Při nižších hodnotách koeficientu sázení  $R$  byl model vždy ve ztrátě. To může být způsobené například tím, že sázkové kanceláře mají kvalitní modely pro odhadování výsledků utkání v kategorii mužů. Celkově lze říct, že Elo model při sázení v kategorii mužů nebyl příliš efektivní. Při nižších hodnotách koeficientu sázení  $R$  dopadl hůře i než jednoduché modely, proto by bylo třeba pro skutečné sázení Elo model vylepšit.

Naopak při sázení na zápasy žen byl u některých sázkových kanceláří model v zisku vždy nezávisle na výběru koeficientu sázení  $R$ . To znamená, že Elo model byl mírně lepší než modely sázkových kanceláří. Zde je však nutné připomenout, že sázkové kanceláře nevypisují kurzy pouze na základě modelů, ale zohledňují i vsazené prostředky sázkařů. Jejich cílem je mít stejné výplaty sázkařům při jakémkoliv výsledku zápasu, jak bylo popsáno v kapitole 9.1, proto tedy může docházet ke změnám kurzů i v důsledku snahy o vyrovnávání případných výplat sázkařům pro jednotlivé varianty výsledku.

## 10 Závěr

Cílem této práce bylo popsat matematický a statistický model pro odhadování výsledků tenisových utkání.

K odhadování výsledků byl použit Elo rating systém. V první části této práce byly popsány různé metody pro odhad úvodního Elo ratingu hráče a způsoby přepočítávání Elo ratingu hráče po odehraném zápase. Také zde bylo ukázáno, jak vypočítat pravděpodobnost vítězství hráče v zápase na základě hodnot Elo ratingů obou soupeřů.

Následně byly testovány a optimalizovány tři modely založené na Elo rating systému, které se lišily jiným odhadem úvodního Elo ratingu a také jiným koeficientem rozvoje  $K$ , který slouží pro přepočítávání Elo ratingu. K optimalizaci byla použita tři kritéria určující kvalitu modelu: přesnost, logaritmická ztrátová funkce a kalibrace. Poté byly modely mezi sebou porovnány. Nejlépe z nich dopadl Elo model, který odhaduje úvodní Elo rating hráče metodou kvantilů a koeficient rozvoje je volen jako konstanta. Tento model byl později ověřen při sázení proti sázkové kanceláři.

Proběhlo také srovnání těchto modelů s modely z článku Kovalchik (2016), který byl inspirací pro tuto práci. V tomto srovnání si vedly modely z této práce velmi dobře. Pouze model založený na kurzech sázkových kanceláří byl lepší než tyto modely. Ostatní modely z článku včetně modelů založených na Elo rating systému měly hodnotu logaritmické ztrátové funkce větší než modely z této práce. Zde je však nutné dodat, že modely v článku odhadovaly výsledky utkání pouze v roce 2014, zatímco modely v této práci odhadovaly výsledky po dobu dvou let 2013 a 2014. Navíc byla také použita jiná historie dat. Proto je srovnání pouze orientační.

Na závěr probíhalo ověření modelu pomocí sázení. Sázel se po dobu dvou let (2015 a 2016) na zápasy mužů a žen. Při sázení na zápasy žen si vedl Elo model velmi dobře. U některých sázkových kanceláří byl model v zisku bez ohledu na volbu koeficientu sázení  $R$ . V případě, kdy bylo sázeno na každý vybraný zápas 10 jednotek, dosáhl model maximálního zisku při sázení u sázkové kanceláře Pinnacles Sports a hodnotě koeficientu sázení  $R = 1,32$ . V tomto případě byl zisk 741 jednotek. U sázení na zápasy mužů si vedl model hůře než při sázení na zápasy žen. Pro nižší hodnoty koeficientu sázení  $R$  byl model ve ztrátě. Pro vyšší hodnoty koeficientu sázení  $R$  (přibližně od  $R = 2,5$ ) byl model v mírném zisku. V případě, kdy bylo sázeno na každý vybraný zápas 10 jednotek, dosáhl model maximálního zisku při sázení u sázkové kanceláře Pinnacles Sports a hodnotě koeficientu sázení  $R = 3,25$ . V tom případě byl zisk 216 jednotek.

Z výsledků je vidět, že Elo model je v kategorii žen schopný sázkovou kancelář porazit. Proto může Elo model sloužit jako možné vylepšení stávajících modelů, případně jako pomocník sázkařů při sázení na tenisová utkání žen.

Naopak v kategorii mužů si Elo model vedl hůře. Při sázení na kategorii mužů by stálo za uvážení zvolit jiný model nebo lépe optimalizovat částky vsazené na jednotlivé zápasy. To by mohlo být námětem na budoucí rozšíření této diplomové práce.

Dalším možným rozšířením této diplomové práce by mohlo být použití Elo modelu na jiné sporty, například na hokej, fotbal, házenou atd. Zde je však rozdíl oproti tenisu v tom, že v těchto sportech může nastat i remíza. Potom by se musela vyřešit otázka, jaká je pravděpodobnost remízy v zápase, což Elo model v podobě, ve které je představen v této práci, neumožňuje.

# 11 Seznam použité literatury a zdroje dat

## 11.1 Literatura

**BRADLEY, Ralph Allan a Milton E. TERRY.** Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*. 1952, 39(3/4), 324-. DOI: 10.2307/2334029.

**COVER, Thomas a Joy THOMAS.** *Elements of information theory*. 2nd ed. Hoboken, N.J.: Wiley-Interscience, 2006. ISBN 978-0-471-24195-9

**DIXON, Mark a Stuart G. COLES.** Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1997, 46(2), 265-280. DOI: 10.1111/1467-9876.00065.

**ELO, Arpad E.** *The rating of chessplayers, past and present*. New York: Arco Pub., 1978. ISBN 06-680-4721-6.

**FIDE.** *World Chess Federation* [online]. [cit. 2018-01-23]. Dostupné z: <https://www.fide.com/fide/handbook.html?id=197&view=article>

**GILSDORF, Keith F. a Vasant A. SUKHATME.** Testing Rosen's Sequential Elimination Tournament Model. *Journal of Sports Economics*. 2007, 9(3), 287-303. DOI: 10.1177/1527002507306790.

**GLICKMAN, Mark a Albyn JONES.** *Rating the Chess Rating System* [online]. [cit. 2018-01-23]. Dostupné z: <http://glicko.net/research/chance.pdf>

**HVATTUM, Lars Magnus a Halvard ARNTZEN.** Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*. 2010, 26(3), 460-470. DOI: 10.1016/j.ijforecast.2009.10.002.

**INGRAM, Martin.** TENNIS: ELO MODELLING. *Betfair* [online]. [cit. 2018-01-23]. Dostupné z: <https://www.betfair.com.au/hub/betting-better/tennis-modelling-elo/>

**KARLIS, Dimitris a Ioannis NTZOUFRAS.** Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 2003, 52(3), 381-393. DOI: 10.1111/1467-9884.00366.

**KLAASSEN, Franc J.G.M. a Jan R. MAGNUS.** Forecasting the winner of a tennis match. *European Journal of Operational Research*. 2003, 148(2), 257-267. DOI: 10.1016/S0377-2217(02)00682-3.

**KOVALCHIK, Stephanie Ann.** Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*. 2016, 12(3), DOI: 10.1515/jqas-2015-0059.

**MAHER, M. J.** Modelling association football scores. *Statistica Neerlandica*. 1982, 36(3), 109-118. DOI: 10.1111/j.1467-9574.1982.tb00782.x.

**MAREK, Patrice, Blanka ŠEDIVÁ a Tomáš ŤOUPAL.** Modeling and prediction of ice hockey match results. *Journal of Quantitative Analysis in Sports*. 2014, 10(3), DOI: 10.1515/jqas-2013-0129.

**MCHALE, Ian a Alex MORTON.** A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*. 2011, 27(2), 619-630. DOI: 10.1016/j.ijforecast.2010.04.004.

**NEWTON, Paul K. a Joseph B. KELLER.** Probability of Winning at Tennis I. Theory and Data. *Studies in Applied Mathematics*. 2005, 114(3), 241-269. DOI: 10.1111/j.0022-2526.2005.01547.x.

**ORNSTEIN, David.** *David Nalbandian disqualified from Queen's final after kick. 2012* [online]. [cit. 2018-01-23]. Dostupné z: <http://www.bbc.com/sport/tennis/18491229>

**ROSS, Daniel.** *Arpad Elo and the Elo Rating System* [online]. 2007 [cit. 2018-01-23]. Dostupné z: <https://en.chessbase.com/post/arpad-elo-and-the-elo-rating-system>

**Šachový klub TH JUNIOR Banská Bystrica.** *Šachový ELO systém* [online]. [cit. 2018-01-23]. Dostupné z: [http://www.juniorbb.iprofil.cz/ELO\\_CZ.pdf](http://www.juniorbb.iprofil.cz/ELO_CZ.pdf)

**ŠPAČEK, Jan.** *Modelování a odhadování výsledků sportovních utkání*. Plzeň, 2015. Bakalářská práce. Západočeská univerzita, Fakulta aplikovaných věd.

**TheChessPiece.** *What is an Elo rating?* [online]. [cit. 2018-01-23]. Dostupné z: [https://www.thechesspiece.com/what\\_is\\_an\\_elo\\_rating.asp](https://www.thechesspiece.com/what_is_an_elo_rating.asp)

## 11.2 Zdroje dat

**Tennis-data.** *Tennis Betting | Tennis Results | Tennis Odds* [online]. [cit. 2018-01-23]. Dostupné z: <http://www.tennis-data.co.uk/alldata.php>

**ATP Rankings.** *ATP World Tour* [online]. [cit. 2018-01-23]. Dostupné z: <http://www.atpworldtour.com/en/rankings/singles?rankDate=2016-12-26&rankRange=0-100>

## Příloha A

V této příloze jsou popsána dvě rozdělení pravděpodobnosti, která byla použita v této diplomové práci.

### Normální rozdělení

Normální rozdělení je spojité rozdělení pravděpodobnosti náhodné veličiny s dvěma parametry  $\mu$  a  $\sigma$ .

Hustota normálního rozdělení je ve tvaru

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}. \quad (\text{A.1})$$

Střední hodnota a rozptyl jsou

$$E(X) = \mu, \quad (\text{A.2})$$

$$D(X) = \sigma^2. \quad (\text{A.3})$$

### Logistické rozdělení

Logistické rozdělení je spojité rozdělení pravděpodobnosti náhodné veličiny s dvěma parametry  $\mu$  a  $s$ .

Hustota logistického rozdělení je ve tvaru

$$f(x) = \frac{e^{-\frac{x-\mu}{s}}}{s \cdot \left(1 + e^{-\frac{x-\mu}{s}}\right)^2}. \quad (\text{A.4})$$

Střední hodnota a rozptyl jsou

$$E(X) = \mu, \quad (\text{A.5})$$

$$D(X) = \frac{s^2 \cdot \pi^2}{3}. \quad (\text{A.6})$$

Kvantil logistického rozdělení je

$$Q(p) = \mu + s \cdot \ln\left(\frac{p}{1-p}\right). \quad (\text{A.7})$$

## Příloha B

*DP\_Spacek.pdf* – Elektronická verze této diplomové práce

### Matlab

#### Spouštěcí skripty

***Optimalizace\_Elo\_modely.m*** – Spouštěcí skript, který optimalizuje hodnoty parametrů všech 3 modelů

Na následujícím obrázku je ukázka několika prvních řádků skriptu, kde je potřeba provést nastavení dle pokynů popsaných v modrém rámečku.

```
% Nacteni dat
clear all
clc

data = 1; %vyber kategorie, 1=DB muzi,2=DB zeny

model = 3;

NacteniDat
if(data==1)
    ZacatekModelu = ZacatekRoku(2005,Rok);
end
if(data==2)
    ZacatekModelu = ZacatekRoku(2007,Rok);
end
KonecModelu = KonecRoku(2014,Rok);
ZacatekOptimalizace = ZacatekRoku(2013,Rok);
KonecOptimalizace= KonecModelu;
```

Zde se nastaví výběr databáze. Pokud je nastavena 1, tak je vybrána kategorie mužů. Pokud je nastavena 2, tak je vybrána kategorie žen.

Zde se nastavuje, který model se bude optimalizovat. Je možné nastavit 1, 2 a 3. Podle toho se bude optimalizovat buď Elo model 1, 2 nebo 3.

***Overeni\_modelu\_sazeni.m*** – Spouštěcí skript, který ověřuje modely pomocí sázení

Na následujícím obrázku je ukázka několika prvních řádků skriptu, kde je potřeba provést nastavení dle pokynů popsaných v modrém rámečku.



```

%% Nacteni dat
clear all
clc

data = 1; %1=DB muzi,2=DB zeny

NacteniDat

if(data==1)
    ZacatekModelu = ZacatekRoku(2005,Rok);
    k=19
    l=23
    p = ZacatekRoku(2007,Rok);
end
if(data==2)
    ZacatekModelu = ZacatekRoku(2007,Rok);
    k=29
    l=27
    p = ZacatekRoku(2009,Rok);
end
KonecModelu = KonecRoku(2016,Rok);
ZacatekSaz = ZacatekRoku(2015,Rok);
KonecSaz= KonecModelu;

```

Zde se nastaví výběr databáze. Pokud je nastavena 1, tak je vybrána kategorie mužů. Pokud je nastavena 2, tak je vybrána kategorie žen.

**KriteriaSK.m** – Spouštěcí skript, který počítá hodnoty kritérií kvalit určující kvalitu modelu pro sázkové kanceláře

## Funkce

**Calibration.m** – Funkce, která spočítá hodnotu kalibrace

**EloModel1.m** – Počítání Elo ratingu a pravděpodobností vítězství hráčů pomocí Elo modelu 1

**EloModel2.m** – Počítání Elo ratingu a pravděpodobností vítězství hráčů pomocí Elo modelu 2

**EloModel3.m** – Počítání Elo ratingu a pravděpodobností vítězství hráčů pomocí Elo modelu 3

**GrafBank.m** – Funkce, která vytvoří grafy banku v závislosti na koeficientu sázení  $R$

**GrafElo.m** – Funkce, která vytvoří grafy Elo ratingů

**KonecRoku.m** – Určení čísla řádku posledního zápasu ve vybraném roce

**LogLoss.m** – Funkce, která spočítá hodnotu logaritmické ztrátové funkce

**ModelFavorit.m** – Funkce, která vybere favorita utkání při sázení na favorita

**ModelOutsider.m** – Funkce, která vybere outsidera utkání při sázení na outsidera

**ModelNahodneSazeni.m** – Funkce, která vybere jednoho z hráčů při náhodném sázení

**NacteniDat.m** – V tomto programu se načítají data z databáze

**PravdepodobnostSK.m** – Funkce, která počítá pravděpodobnosti vítězství hráčů v zápasech dle sázkových kanceláří

**PredAccur.m** – Funkce, která spočítá hodnotu přesnosti

**SazeniFixni.m** – Funkce, která počítá hodnotu banku po každém zápase při strategii sázení konstantní vklad

**SazeniPomerne.m** – Funkce, která počítá hodnotu banku po každém zápase při strategii sázení procento z banku

**Urcenim.m** – Funkce, která určuje, zda se na zápas může vsadit v závislosti na parametru  $l$

**VsazeniElo.m** – Funkce, která určuje, na které zápasy vsadit

**VyvojEla.m** – Funkce, která vypíše vývoj Elo ratingu vybraného hráče

**ZacatekRoku.m** – Určení čísla řádku prvního zápasu ve vybraném roce

## Excel

**DBMuzi.xlsx** – Databáze zápasů mužů

**DBZeny.xlsx** – Databáze zápasů žen

**DBMuziBezHlavicekOdehrane.xlsx** – Upravená databáze zápasů mužů pro Matlab

**DBZenyBezHlavicekOdehrane.xlsx** – Upravená databáze zápasů žen pro Matlab

**EloModel1 optimalizace.xlsx** – Výsledky optimalizace Elo modelu 1

**EloModel2 optimalizace muži.xlsx** – Výsledky optimalizace Elo modelu 2 v kategorii mužů

**EloModel2 optimalizace ženy.xlsx** – Výsledky optimalizace Elo modelu 2 v kategorii žen

**EloModel3 optimalizace.xlsx** – Výsledky optimalizace Elo modelu 3

**Elo rating žebříček 2016.xlsx** – Žebříček hráčů dle Elo ratingu na konci roku 2016

**Charakteristika DB.xlsx** – Základní charakteristika databáze

**OvěřeníModeluSazeniEloModel3.xlsx** – Výsledky ověření modelů pomocí sázení