Statistical model for chess tournament simulations

By Otto Milvang, 2nd Dec 2024

Introduction

Tournament simulations are powerful tools to explore the properties of pairing systems and tie-breaks. In 2015 I wrote “Probability for the outcome of a chess game based on rating” [1] which was a statistical analysis of 53 305 tournaments between 2010 and 2014. A remarkable observation is that the probability for the outcome of a chess game based on rating did not follow the probability given in the rating regulation in FIDE handbook [2], in tables 8.1.

In 2023 FIDE Qualification Commission did an analysis of the rating system, and built on Jeff Sonas proposal [3], the entire rating system was compressed 1. March 2024. This means that after the compression the expected score for a player meeting another player shall match the tables in FIDE handbook.

The ELO system[4]

Mathematics is from Arpad Elo [4]. If player A has rating Ra and player B rating Rb, then the expected score Ea of player A is:

$$Ea(Ra, Rb) = \frac{1}{1 + 10^{(Rb-Ra)/400}}$$

Similarly, the expected score for player B is

$$Ea(Ra, Rb) = \frac{1}{1 + 10^{(Ra-Rb)/400}}$$

Statistical model

The statistical model is based on the ELO system, but as we know this is only one part of the model. We know from experience that white has a small advantage. More recent sources indicate that White scores approximately 54 to 56 percent. In 2005, Grandmaster (GM) Jonathan Rowson wrote that "the conventional wisdom is that White begins the game with a small advantage and, holding all other factors constant, scores approximately 56% to Black's 44%" [5]. Later in this paper we will look closer into this problem.

Another problem is the draw rate. According to chess analyst Jeff Sonas, the draw rate for the lowest rated players are less than 25%, and for strong grandmasters more than 60% [6].

Our goal is to make a model for $Pw(Rw, Rb)$, $Pd(Rw, Rb)$ and $Pb(Rw, Rb)$ where Pw, Pd and Pb are the probabilities for “White win”, “Draw” and “Black win”, Rw= White rating and Rb = Black rating.

If we have a model for $Pw(Rw, Rb)$ and $Pb(Rw, Rb)$ then

$$Pd(Rw, Rb) = 1 - Pw(Rw, Rb) - Pb(Rw, Rb)$$

and

$$Xa(Ra, Rb) = (2 + Pw(Ra, Rb) + Pb(Rb, Ra) - Pb(Ra, Rb) - Pw(Rb, Ra))/4$$

Now our goal is to have

$$Ea(Ra, Rb) \approx Xa(Ra, Rb)$$

The model proposed in this paper defines

$$Pw(Rw, Rb) = \frac{1}{1 + 10^{(Rb - Rw - W)/S}}$$

And

$$Pb(Rw, Rb) = \frac{1}{1 + 10^{(Rw - Rb + B)/S}}$$

Where W, B and S are parameters that describe the advantages of white, and the draw rate. To make the best models W, B and S will be functions and depend on Rw and Rb

Data

For the analysis, all results from tournaments with standard time control registered for FIDE rating in the period January 2016 to April 2023, with a total of 7 674 837 game results were used to build the model.

The data extracted for each game was "White rating", "Black rating", "result", "date". Since these games were played before March 2024, the rating was adjusted for players in the range 1000 – 2000 such that

$$New\ rating = Old\ rating * f + 2000 * (1.0 - f)$$

In the rating adjustment in March 2024, $f = 0.6$ Deflation has been a huge problem for several years, and has increased month by month. Just to have data that were representative the rating was adjusted with

$$f = -0.00116 * month + 0.78$$

Where $month$ is month number, starting with month number 1 in January 2008.

All games are divided into buckets sorted on "rating mean" $Rm = (Rw + Rb)/2$, and "rating diff" $Rd = Rw - Rb$. The width on the buckets are 50 in both direction, so bucket B(1700, -150), contains all games with rating mean 1675-1724 and rating diff -175 to -126.

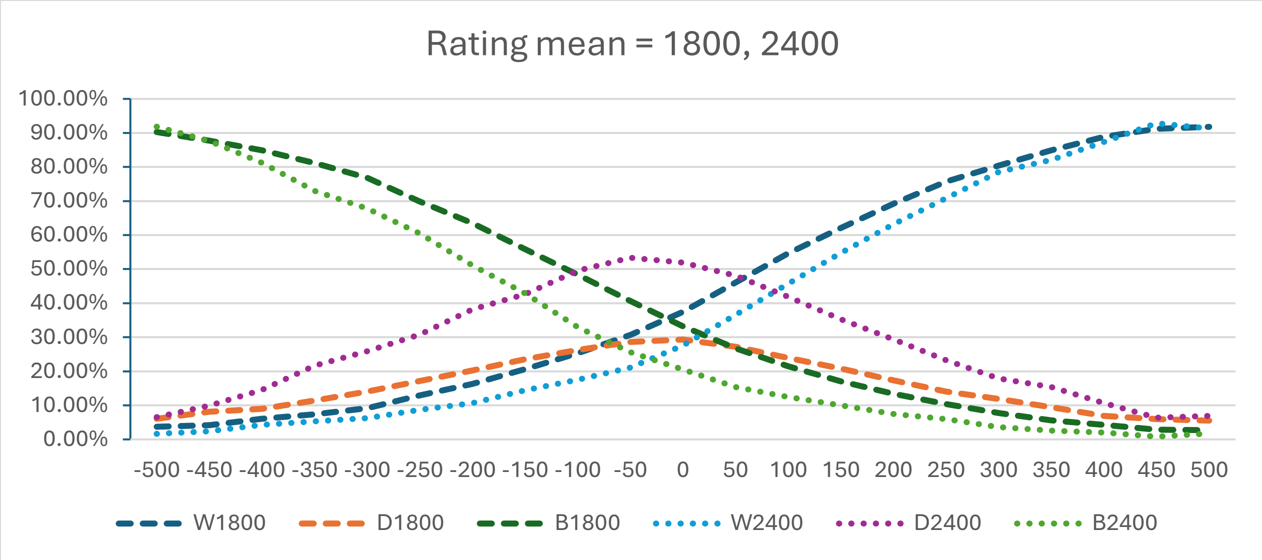


Figure 1. Percentage of White win, Draw, and Black win for buckets with rating mean 1800 (622658 games) and 2400 (97297 games). X-axis shows $Rw - Rb$. A white player with rating 2500 vs black player 2300 has mean rating 2400 and $Rw - Rb$ = 200, then $Pw = 69.3\%$, $Pb = 13.4\%$, $Pd = 100\% - 69.3\% - 13.4\% = 17.3\%$

When we compare games with mean rating 2400 to games with mean rating 1800 the total draw rate increases from 22% to 40%, and the advantages of having white goes from 19 to 36 rating points.

The model

$$\boldsymbol{Pw(Rw, Rb) = \frac{1}{1 + 10^{(Rb - Rw - W(Rm))/S(Rm)}}}$$

$$\boldsymbol{Pb(Rw, Rb) = \frac{1}{1 + 10^{(Rw - Rb + B(Rm))/S(Rm)}}}$$

$$\boldsymbol{Pd(Rw, Rb) = 1.0 - Pw(Rw, Rb) - Pb(Rw, Rb)}$$

$$\boldsymbol{Rm = (Rw + Rb)/2}$$

To find the parameters for $W(Rm)$, $B(Rm)$ and $S(Rm)$ we must analyze the game database, and for each range of Rm estimate the value of $W(Rm)$ such that $Rb - Rw - W(Rm) = 0$ and then estimate which value of S that is the best approximation.

$$Y(Rb - Rw) = (Rb - Rw - W(Rm))/S(Rm) = log10(\frac{1.0}{Pw(Rb,\ Rw)} - 1.0)$$

Example, Rm=2000, calculate $W(Rm)$ and $S(Rm)$

From the database we have

Rw-Rb	#Win	#games	#score	Y(Rw-Rb)	S(Rw-Rb)
-400	4477	5460	0.055495	1.230955	369.788168
-350	6567	8559	0.080266	1.059128	379.83111
-300	9015	12463	0.099174	0.958245	387.749925
-250	12753	19401	0.126849	0.837802	388.034562
-200	16659	28086	0.156768	0.730688	379.063243
-150	19197	37299	0.193088	0.62107	364.078364
-100	18429	42330	0.23142	0.521289	365.490756
-50	14289	40075	0.280324	0.409476	373.801719
0	11590	40763	0.342566	0.283109	393.87463
50	9449	41228	0.423862	0.133301	419.40369
100	7902	42794	0.506964	-0.0121	439.540403
150	5855	38526	0.592275	-0.16216	450.798948
200	3551	29066	0.658467	-0.2851	446.615108
250	1889	20010	0.725137	-0.4213	425.675301
300	1007	13328	0.770483	-0.52595	401.870275
350	495	8823	0.823529	-0.66901	388.187079
400	235	5723	0.860038	-0.78851	361.91351

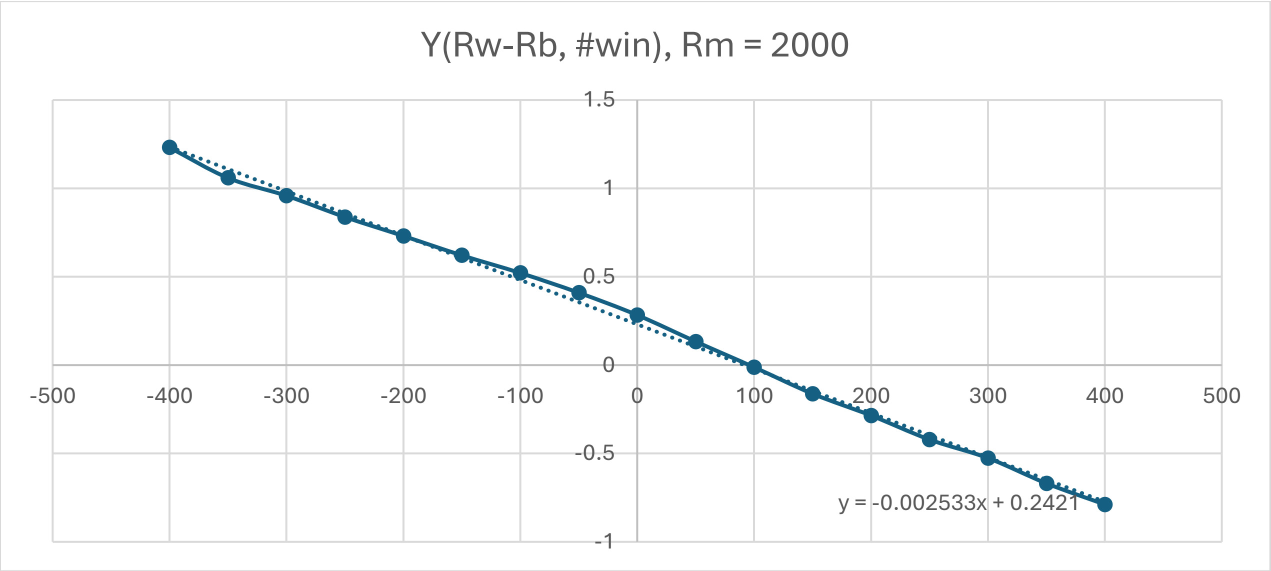


Figure 2. Example, Rm=2000, estimate W(Rm) and S(Rm).

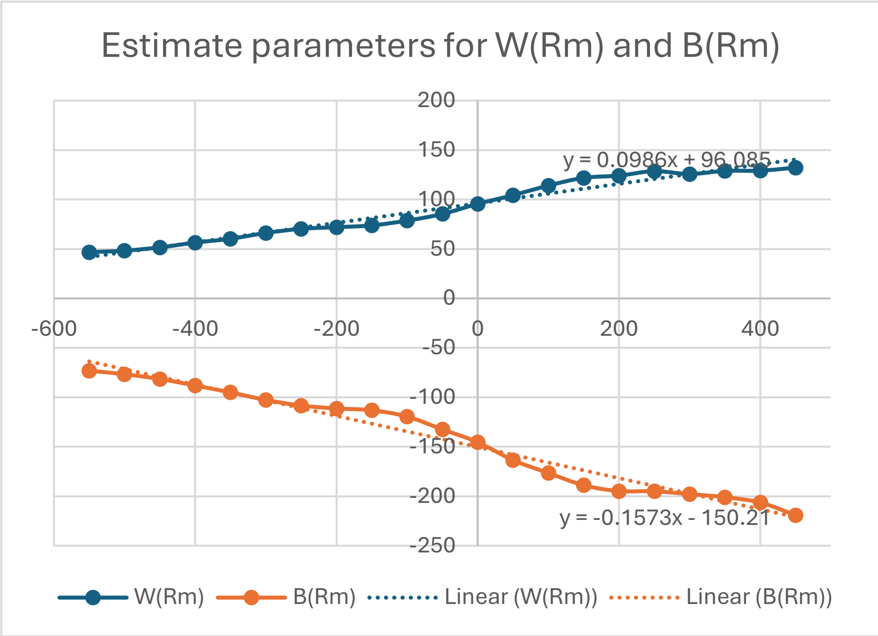
$$W(Rm) = \frac{0.2421}{-0.002533} = 95.56$$

$$S(Rm) = -\left(\frac{1.0}{-0.002533}\right) = 394.7$$

The same is repeated for $W(Rm)$ and $B(Rm)$ for Rm in the range 1450 to 2450

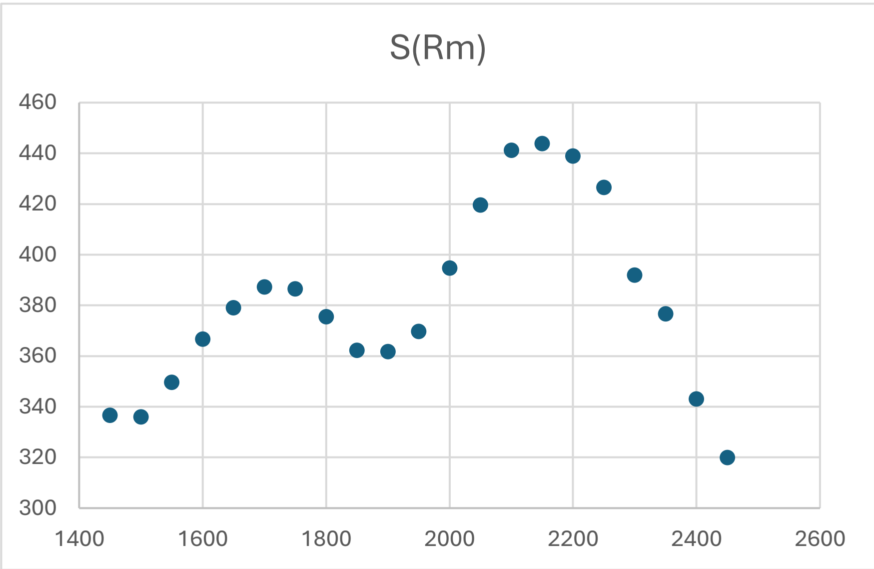
The parameters W(Rm) and B(Rm)

Rm-2000	W(Rm)	B(Rm)
-550	46.76637	-73.1624
-500	48.25023	-76.6159
-450	51.58534	-81.4428
-400	56.41478	-88.0205
-350	60.30304	-95.0001
-300	66.14444	-102.581
-250	70.33684	-108.339
-200	72.01306	-111.205
-150	73.7926	-113.088
-100	78.65722	-119.435
-50	85.33586	-132.401
0	95.55875	-145.337
50	104.3462	-163.472
100	114.0198	-176.427
150	121.89	-188.841
200	124.1146	-194.733
250	128.6276	-194.946
300	125.7489	-197.808
350	128.9348	-200.929
400	129.1103	-206.326
450	132.2747	-219.101



The parameters S(Rm)

R(m)	S(Rm)
1450	336.6554
1500	335.9804
1550	349.617
1600	366.6992
1650	379.0716
1700	387.2078
1750	386.4553
1800	375.441
1850	362.2957
1900	361.7299
1950	369.6249
2000	394.7463
2050	419.6282
2100	441.1824
2150	443.8295
2200	438.839
2250	426.5368
2300	391.8941
2350	376.6906
2400	343.014
2450	319.963



It seems that S(Rm) does not have much impact in the probability distribution. $S(\mathrm{Rm}) = 370$ seems to be a good choice.

The model with parameters

$$\boldsymbol{Pw(Rw, Rb) = \frac{1}{1 + 10^{(Rb-Rw-W(Rm))}/_{S}}}$$

$$\boldsymbol{Pb(Rw, Rb) = \frac{1}{1 + 10^{(Rw-Rb+B(Rm))}/_{S}}}$$

$$\boldsymbol{Pd(Rw, Rb) = 1.0 - Pw(Rw, Rb) - Pb(Rw, Rb)}$$

$$\boldsymbol{Rm = (Rw + Rb)/2}$$

$$\boldsymbol{W(Rm) = (Rm - 2000) * 0.0986 + 96.0}$$

$$\boldsymbol{B(Rm) = -(Rm - 2000) * 0.157 - 150.0}$$

$$\boldsymbol{X(Rm) = (Rm - 1000)/50}$$

$$\boldsymbol{S = 370}$$

Statistical properties

With these equations we calculate $Pw(Rw, Rb)$, $Pd(Rw, Rb)$ and $Pb(Rw, Rb)$, while the rating system defines expected score

$$Ea(Ra, Rb) = \frac{1}{1 + 10^{(Rb-Ra)}/_{400}}$$

This must be equal to the probability of score between two players Ra and Rb

$$Xa(Ra, Rb) = (Pw(Ra, Rb) + 0.5 * Pd(Ra, Rb) + 0.5 * Pd(Rb, Ra) + Pb(Rb, Ra))/2$$

$Ea(Ra, Rb)$ is compared to $Xa(Ra, Rb)$ for $Rm = 1700$ and $Rm = 2300$ in the range $-600 \leq Rb - Ra \leq 600$

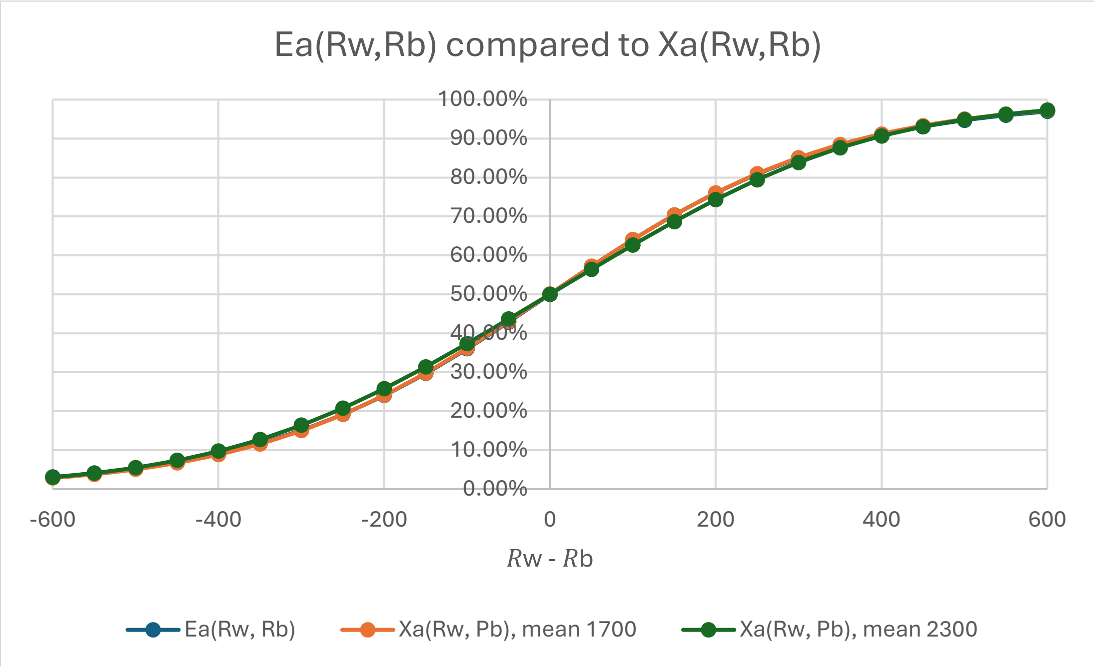


Figure 5. Statistical property Example, Rm=2000, calculate W(Rm) and S(Rm)

It's also possible to minimize $|Ea(Ra, Rb) - Xa(Ra, Rb)|$ by adjusting S, but this will be a trade off between statistical properties, and approximation to the database games.

Verification of the parameters

The results from the game database are compared to the probability model

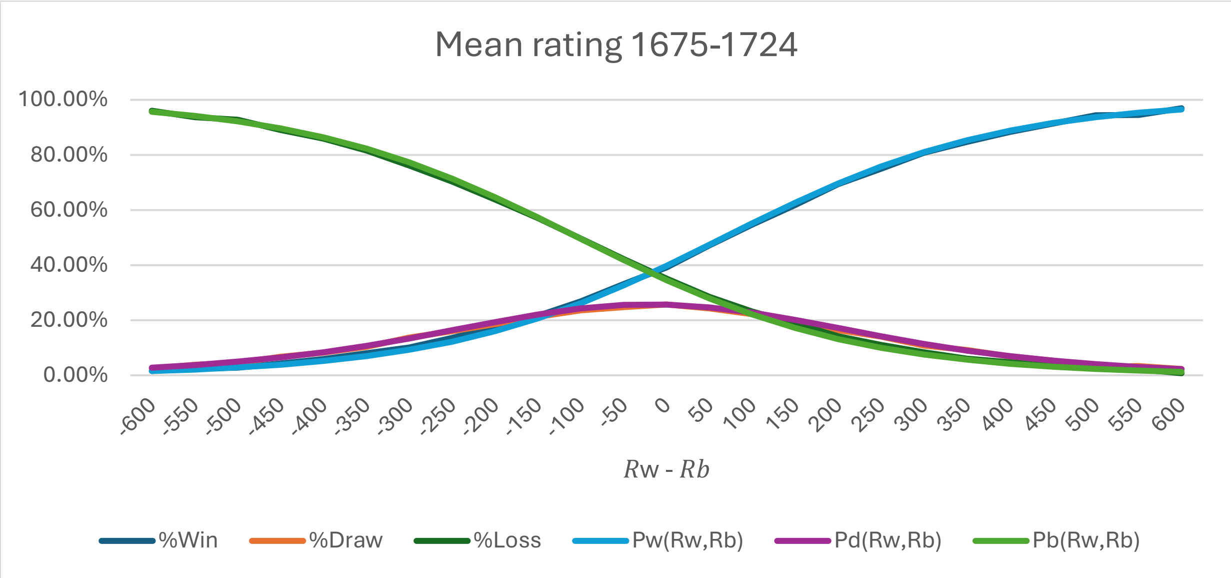


Figure 6. %Win, %Draw and %Loss are based on 542656 games. X-axis is Rw - Rb

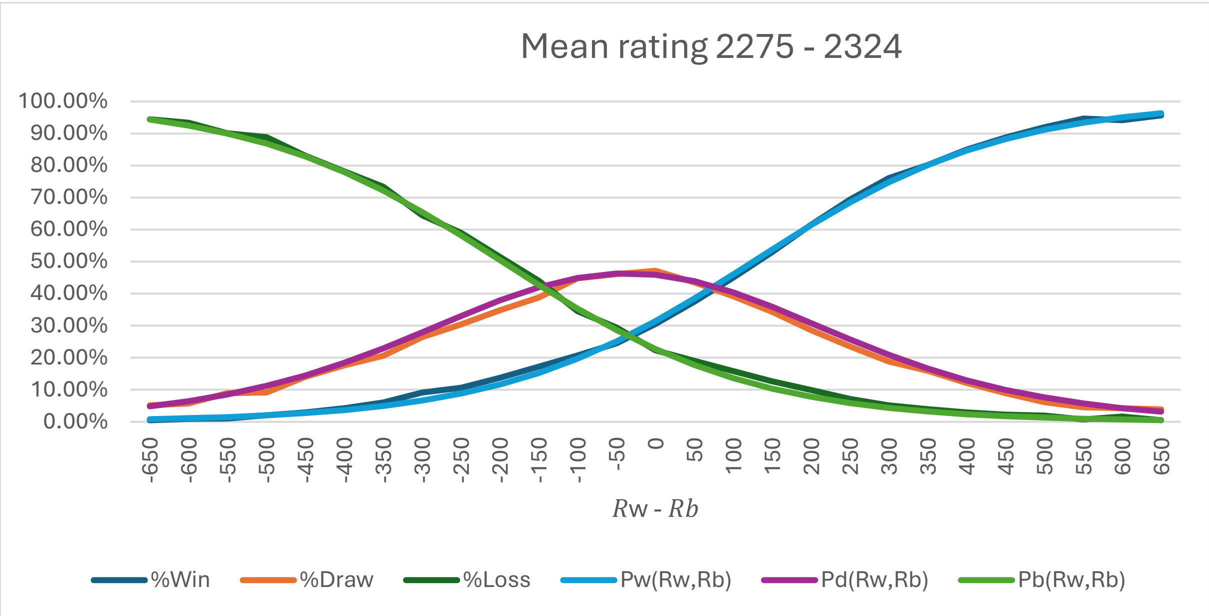


Figure 7. %Win, %Draw and %Loss are based on 146082 games. X-axis is Rw - Rb

These diagrams compare the statistical model to the data from the database. For the first diagram the values for %win, %draw and %loss are based on games extracted from the database where $1675 \le \frac{Rw+Rb}{2} \le 1724$ and the x-axis is $Rw - Rb - 25 \le x < Rw - Rb + 25$ for $x = -600, -550, \ldots, 550, 600$ and $Pw(Rw, Rb)$, $Pd(Rw, Rb)$ and $Pb(Rw, Rb)$ calculated for the same values.
Similar for $\mathrm{Rm} = 2300$

Discussion

This paper has presented a model for estimating tournament results based on a statistical model. The model will work good in the rating range 1600 to 2400, and fairly good in the ranges 1400-1600 and 2400-2600. Fairly good more because of the number of games is low.

The mystery in this work has been to understand the meaning of S, that is the scaling factor. For rating higher than 2300, maybe S should have been adjusted, and for $Rm = 2600$, $S = 320$ may be a better choice than $S = 370$. Anyway, the model will work good for tournament simulations which is the main goal of this paper.

Program in python

```
def prob(rw, rb):
    # rw = Rating white player
    # rb = Rating black player

    # parameters
    wa = 0.0986
    wb = 96.
    ba = -0.157
    bb = -150.
    s = 370

    # rmean
    rm = (rw+rb)/2.0

    # Probability of white win
    w = (rm-2000)*wa + wb
    expw = (rb-rw+w)/s
    pw = 1.0 / (1.0 + pow(10.0, expw))

    # Probability of black win
    b = (rm-2000)*ba + bb
    expb = (rw-rb-b)/s
    pb = 1.0 / (1.0 + pow(10.0, expb))

    # Probability of draw
    pd = 1.0 - pw -pb

    return [pw, pd, pb]
```

Acknowledge

I would like to thank FIDE and Jeff Sonas for providing me with access to the FIDE result database. A large data set is essential for creating a correct statistical model.

I would also like to give special thanks to my daughter Marit for help and useful comments.

References

[1] Probability for the outcome of a chess game based on rating, Otto Milvang 2015. https://www.nordstrandsjakk.no/documents/spp/Probability.pdf

[2] FIDE Handbook, B.02. FIDE Rating Regulations (Qualification Commission) / FIDE Rating Regulations effective from 1 March 2024. https://handbook.fide.com/chapter/B022024

[3] Repairing the FIDE Standard Elo rating system, Jeff Sonas, 20 July 2023, www.fide.com/docs/presentations/Sonas%20Proposal%20-%20Repairing%20the%20FIDE%20Standard%20Elo%20Rating%20System.pdf

[4] Arpad Elo. The rating of chessplayers, past and present, 1978. Ch 8, Rating system theory https://gwern.net/doc/statistics/order/comparison/1978-elo-theratingofchessplayerspastandpresent.pdf

[5] First-move advantage in chess, https://en.wikipedia.org/wiki/First-move_advantage_in_chess

[6] Sonas, Jeff (September 6, 2011). "Sonas: what exactly is the problem?". *en.chesbase.com*. Chessbase. Retrieved October 9, 2015.