Variable k-values in Elo Systems with a Fixed Number of Rounds

NICK GRAYSON, BEN YOUNG

ABSTRACT

In this discourse, we examine some issues with utilizing a traditional Elo system in competitions with a finite, fixed number of rounds. After a process of design, we propose modifications to traditional Elo systems that solve these problems and provide brief calculations and potential examples of implementation.

1. INTRODUCTION

A common pitfall of many tournament formats is that matches are assigned by priorities other than closeness. In elimination formats, for example, the first seed will commonly play the last seed in the first match of the first round of the tournament. This is done because after the first round of a tournament, you wish to have the top half of players win and because the highest seeded player should have the easiest match, while players in the middle of the pack should have a higher chance of making an upset and showing that they are, in fact, a part of the top half of players.

When creating a competitive experience that strives to have even and compelling matches, allowing players to improve their in-game skills, we strive for balanced matches while still being able to rank players. As such we implement a modified Elo-style leaderboard format.

Before we begin, let us define some terms.

A *game* is one instance of two players or teams playing each other. A game has a *score*, which is the resulting outcome of play. That score can be a tie, a win or a loss.

A *match* is a series of games between two players or teams. A match has a format which is usually consistent throughout the tournament such as winning the majority of a predefined number of matches like Best-of-Three (Bo3) or Best-of-Five (Bo5). Match formats can vary greatly by game and involve things such as picks, bans, etc. and are out of scope of this discourse.

A *round* is a group of matches in a tournament that occur at, or are assigned at, a similar time between many pairs of players or teams.

Ranking is the process of ordering teams, players, or other entities in a list by one or more metrics.

Matchmaking is the process of pairing teams or players to play each other in a competitive setting based on one or more metrics.

Rating is a quantitative score assigned to a single player or team based on their performance and/or other metrics. Rating will be the main topic of this discourse.

2. QUESTIONS AND DECISIONS OF DESIGN

In Elo systems, performance is inferred from results of matches against other players and the ratings of both players before any given match. Understanding what the steady state of a given system needs to look like is important before any of the mathematical relationships and detailed are ironed out.

For Elo, some important questions to ask are

1. What rating is an average player or team located at?
2. What rating is a top player or team located at?
3. What rating is a bottom player or team located at?
4. What rating is a new player or team located at?
5. When playing an opponent with the same rating and winning, how much rating is gained?
6. If player A with score S_A plays against player B with score S_B, what chance of victory does he have?

Some of these questions can be decided at the beginning of creating a system (1, 2, 3, 4), others require a lot of thought (5, 6). Some of these are game-agnostic (1, 2, 3, 4) and others are not (6). Some of these questions only matter in relation to each other (2, 3). Some may depend on the scope of a given Elo System as well.

As we will find out, many of these questions have direct analogs in the few formulas used by Elo systems. For now, we will be creating a system that starts players at a rating of 1200.

3. MATHEMATICAL DETAILS AND THEORY

First, we define a metric called the *score*. The score is calculated based on the result of the match and is a representation of how well a given player performed between 0 and 1, with 0 meaning a complete loss and 1 meaning a complete victory. In chess, a score of .5 is a draw.

Score depends a lot on match format, the above detailed layout is what is used for chess by the FIDE and is used mostly for Bo1

Result [Win - Loss - Draw]	Score

matches. In a system for Bo3 of StarCraft II, for example, we may define the expected score as the percentage of games won (and can even account for draws) with the layout detailed to the right.

Table 2.1
Proposed Scores for a Bo3 Match format accounting for ties

2 - 0 - 0	1.00
2 - 1 - 0	0.66
1 - 0 - 1	0.75
2 - 0 - 1	0.83
1 - 2 - 0	0.33
0 - 2 - 0	0.00

Elo is designed such that a rating difference △ means that the stronger player has an expected average score (or *expected score*) of approximately 0.75. In FIDE chess rankings, $\triangle = 400$.

Expected score is the probability of a player winning plus half the probability of drawing. An expected score of 0.75 could represent a 75% chance of winning, a 25% chance of losing, and a 0% of drawing, OR it could represent a 50% chance of winning and a 50% chance of drawing OR anything in between such as a 60% chance of winning, a 30% chance of drawing, and a 10% chance of losing.

To directly cite the Wikipedia page on Elo rating systems[1], If Player A has a rating of $\boldsymbol{R_A}$ and Player B a rating of $\boldsymbol{R_b}$, the exact formula for the expected score of Player A, $\boldsymbol{E_A}$, and that of Player B, $\boldsymbol{E_B}$, are, respectively

$$E_A = \frac{1}{1+10^{(R_B-R_A)/\triangle}} \qquad E_B = \frac{1}{1+10^{(R_A-R_B)/\triangle}}$$

(2.1, 2.2)

And can also be expressed by

$$E_A = \frac{Q_A}{Q_A+Q_B} \qquad E_B = \frac{Q_B}{Q_B+Q_A}$$

(2.3, 2.4)

Where

$$Q_P = 10^{R_P/\triangle}$$

(2.5)

Using equations 2.3 and 2.4, we note that the denominators are the same, meaning that the expected score for player A is Q_A/Q_B times greater, or

$$\frac{E_A}{E_B} = \frac{\frac{Q_A}{Q_A+Q_B}}{\frac{Q_B}{Q_B+Q_A}} = \frac{Q_A}{Q_B}$$

(2.6)

By using equation 2.5 we find

$$\frac{E_A}{E_B} = \frac{Q_A}{Q_B} = \frac{10^{R_A/\triangle}}{10^{R_B/\triangle}} \tag{2.7}$$

This allows us to draw some important conclusions and relationships.

1. If $\boldsymbol{R_A} = \boldsymbol{R_B}$ then $\boldsymbol{E_A} = \boldsymbol{E_B}$
2. For every $\triangle$ points above E_B that E_A is, $\frac{E_A}{E_B}$ increases tenfold

Also of note is that $\boldsymbol{E_A} + \boldsymbol{E_B} = 1$. Thank god I don't know what I would have done if that was not true.

4. UPDATING RATINGS

Supposing Player A was expected to score E_A but actually scored S_A points. The formula for updating their rating from $\boldsymbol{R_A}$ to $\boldsymbol{R`_A}$ is:

$$R`_A = R_A + k_A\,(S_A - E_A) \tag{3.1}$$

Where $\boldsymbol{k_A}$ is the k-value for player A, a multiplier that affects volatility of ratings. A higher k-value results in more rating change resulting from a match. The calculation for k is defined in section 4.

5. CALCULATING THE K VALUE FOR A TEAM

We define a value D as the default k-value. This value must be in scale with the initial rating of teams, the total number of matches, and the expected upper and lower bounds of the distribution of ratings to be. To better understand the effect of k-value, if two teams with even rating play, the team that defeats the other team 1-0 will gain half of their k-value in rating and the other team will lose half of their k-value in rating. For reference, with an initial rating average of 1200 and 20 total games, a k-value of 100 means that a team that wins 20 matches perfectly and loses 0 matches will end with a rating of 2200 provided that they are always placed against a team with the same rating, as they will gain 100/2 = 50 rating upon each match win.

Relative to traditional implementations of Elo, let us give ourselves three specific constraints.

1. We have a small number of participants such that we need to provide structured play instead of a matchmaking queue
2. Our leaderboard is a distinct tournament event with a fixed amount of rounds
3. Matches are played and assigned synchronously but are optional

These constraints, while making the design of a rating system more difficult, if successfully designed around will allow Elo-style leaderboard experiences to be successful in more situations than we are traditionally possible.

We address these three constraints by creating a flexible system for calculating k that we would not be able to do in a traditional setting, relying on our match assignments and rounds, and tracking our confidence in ratings of given teams by their participation.

To find the k-value for a team t in a round r, we first define a value $x_{t,r}$ which is an exact measure of how 'behind' a team is in the sum of their past k-values relative to those of a theoretical team who has played all possible matches available.

$$x_{t,r} = \sum_{n=0}^{r} k_n^0 - \sum_{n=0}^{r} k_n^t \tag{4.1}$$

The k-value for team t in round r of the tournament is then calculated from the value x_t and given by equation (4.2).

$$k_r^t = D \bullet \frac{x_{t,r}}{\left(x_{t,r}+D\right)} + D \tag{4.2}$$

Where team $t = 0$ is a theoretical team that has played every possible match

$$k_r^t = D \bullet \frac{x_{t,r}}{\left(x_{t,r}+D\right)} + D$$

To break this down more, the k-value of team t in round r of the tournament is k_r^t. D is the default k-value. The value x_t is the total amount of k-value that the current team is behind an ideal team, calculated by equation (4.1) where the first sum is the total k-values of an ideal team and the second sum is that of team t.

Number of Games Missed	$x_{t,r}$	k_r^t , $D = 100$
0	0	100
1	100	150
2	200	167
3	300	175
4	400	180
5	500	183
6	600	186
7	700	188
8	800	189
9	900	190
10	1000	191

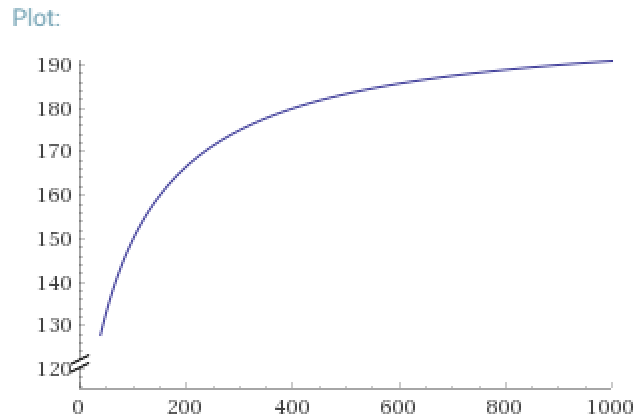

Figure 4.1. An example chart for D = 100, x is how behind a team is in k value. [1]

Figure 4.2. A table of what your first k value is based on how many things you missed

fx =100+100*(((row()-3)*100)-sum(M$3:M21))/(((row()-3)*100)-sum(M$3:M21)+100)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1							K-Value						
2	Week	Game	play	skip w1	skip w1&2	skip w2	skip w1-3	skip w1-4	skip w1-5	skip w1-6	skip w1-7	skip w1,3,5	skip w1,3,5,7
3	1	1	100	0	0	100	0	0	0	0	0	0	0
4	1	2	100	0	0	100	0	0	0	0	0	0	0
5	2	3	100	167	0	0	0	0	0	0	0	167	167
6	2	4	100	157	0	0	0	0	0	0	0	157	157
7	3	5	100	143	180	167	0	0	0	0	0	0	0
8	3	6	100	125	176	157	0	0	0	0	0	0	0
9	4	7	100	108	171	143	186	0	0	0	0	173	173
10	4	8	100	101	163	125	184	0	0	0	0	167	167
11	5	9	100	100	152	108	181	189	0	0	0	0	158
12	5	10	100	100	136	101	178	188	0	0	0	0	144
13	6	11	100	100	117	100	173	186	191	0	0	177	126
14	6	12	100	100	103	100	167	184	190	0	0	172	108
15	7	13	100	100	100	100	157	182	189	192	0	165	0
16	7	14	100	100	100	100	143	179	188	192	0	155	0
17	8	15	100	100	100	100	124	175	187	191	193	140	167
18	8	16	100	100	100	100	107	169	185	190	193	121	157
19	9	17	100	100	100	100	101	160	182	189	192	105	143
20	9	18	100	100	100	100	100	147	180	188	192	100	125
21	10	19	100	100	100	100	100	130	176	187	191	100	108
22	10	20	100	100	100	100	100	111	170	185	190	100	101
23		Total	2000	2000	2000	2000	2000	1999	1837	1515	1152	2000	2000

Figure 4.3. A chart of k-values based on activity

Our system by definition converges upward to a horizontal asymptote at k = 200 as x increases. As x approaches 0, k approaches 100. This is designed as such because unlike traditional elo systems, we don't have an active and accurate system currently in place, and we do not plan on this system being everlasting. This decaying k value causes our system to behave very similarly to the current FIDE chess Elo system with four main differences.

4. We have 5x the magnitude of changes, as k ranges between 200 and 100 instead of 40 and 20/10
5. We are centered at 1200 instead of 1500
6. Our k value curve is more gradual, decreasing from close to 200 down to 100, while they step directly from one value to half of that value whenever a change is made
7. They decrease for players who have reached a certain rating. We do not do this as our system is currently not planned to be ever-living

Sources & Additional Reading

1. [Wikipedia article on Elo Rating System]
2. Rational Polynomial Functions