IB Subject(s): Mathematics

Extended Essay

Title: Reliability of Elo rating and Glicko rating systems

Research Question: To what extent do Elo rating system and Glicko rating system reflect one's performance in chess?

Word Count: 3990

Table of Contents

Introduction

Research question: To what extent do the Elo rating system and Glicko rating system reflect one's performance in chess?

Many chess rating systems, namely the Elo rating system and the Glicko rating system, have been developed by mathematicians over the last century, but they all share ultimately same goal. They strive to provide a rating to chess players that will give them the best possible estimation of their skill. However, different rating systems have different degrees of accuracy which depend on what factors have been considered. It should also be noted that it is simply impossible to produce a value that 'truly' reflects one's skill, due to the complexity of players' performances. Note that some assumptions that are beyond the scope of my understanding were inevitably made to break down the rating systems over the course of the essay.

Elo rating system

Normal distribution of the performance

The first premise of the Elo rating system is that the performance of a player is represented by a normal distribution. By utilizing such a distribution, it accepts the inconsistency of the performance, as well as the considerable magnitude of the deviation in the quality of performance. The mean of the normal distribution is the rating of a chess player and the Elo system arbitrarily set 200 as the standard deviation of such distribution (Pelánek, 2016). To find the area under a normal distribution, which represents the probability that a chess player will play at a certain level, definite integration and the error function were used in the probability density function (PDF) of the normal distribution function. The error function $(\mathrm{erf}\,(x))$ is needed for the calculation, as it is an odd function that is encountered in integrating the normal distribution (Weisstein, 2021).

$PDF\ of\ normal\ distribution$ (Haese, Humphries, Sangwin, & Vo, 2019):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$, where\ \sigma = standard\ deviation, \mu = mean, x = performance\ of\ player$$ **<u>(1)</u>**

$Error\ function$ (Weisstein, 2021):

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int e^{-x^2}\, dx$$

$$, where\ \sigma = standard\ deviation, \mu = mean, x = performance\ of\ player$$ **<u>(2)</u>**

For the sake of simplicity, the sample calculation below used a standard deviation of 1 and mean of 0 to calculate the area within 1 standard deviation of the mean.

$$Area\ under\ the\ PDF\ of\ normal\ distribution\ within\ 1\ \sigma\ of\ \mu\ difference = \int_{-1}^{1} f(x)\, dx,$$

$$Since\ \sigma = 1\ and\ \mu = 0\ for\ this\ sample\ calculation,$$

$$\int_{-1}^{1} f(x)\, dx = \int_{-1}^{1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

$$= \int_{-1}^{1} \frac{1}{1\sqrt{2\pi}} e^{-\frac{(x-0)^2}{2\times 1^2}}\, dx$$

$$= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{x^2}{2}}\, dx$$

From here, a method of integration by substitution was used (Khan Academy, 2021).

$$\frac{1}{\sqrt{2\pi}}\int_{-1}^{1} e^{\frac{-x^2}{2}}\,dx$$

$$Substitue\ u = \frac{x}{\sqrt{2}}, since\ it\ is\ a\ part\ of\ larger\ function\ -\left(\frac{x}{\sqrt{2}}\right)^2$$

Note that the lower and upper bounds of the integral are also affected by integration by substitution.

$$\frac{1}{\sqrt{2\pi}}\int_{-1}^{1} e^{-\frac{x^2}{2}}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-1\left(\frac{1}{\sqrt{2}}\right)}^{1\left(\frac{1}{\sqrt{2}}\right)} e^{-u^2}\,\sqrt{2}du$$

$$= \frac{1}{\sqrt{\pi}}\int_{\frac{-1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}} e^{-u^2}\,du$$

This is where the error function becomes useful as the expression above can be further simplified by substituting the error function.

$$Rearrange\ \operatorname{erf}(u) = \frac{2}{\sqrt{\pi}}\int e^{-u^2}\,du,$$

$$\frac{\sqrt{\pi}}{2}\operatorname{erf}(u) = \int e^{-u^2}\,du,$$

$$substitue\frac{\sqrt{\pi}}{2}\operatorname{erf}(u) = \int e^{-u^2}\,du,$$

$$\frac{1}{\sqrt{\pi}}\int_{\frac{-1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}} e^{-u^2}\,du = \frac{1}{\sqrt{\pi}}\left[\frac{\sqrt{\pi}}{2}\operatorname{erf}\,(u)\right]_{\frac{-1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}}$$

$$Expanding\ the\ equation\ we\ get,$$

$$\frac{1}{\sqrt{\pi}}\left[\left\{\frac{\sqrt{\pi}}{2}\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)\right\} - \left\{\frac{\sqrt{\pi}}{2}\operatorname{erf}\left(-\frac{1}{\sqrt{2}}\right)\right\}\right] = \frac{1}{\sqrt{\pi}} \times \frac{\sqrt{\pi}}{2}\left\{\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) - \left(\operatorname{erf}\left(-\frac{1}{\sqrt{2}}\right)\right)\right\}$$

Note that the error function is an odd function (Weisstein, 2021).

$$\therefore \operatorname{erf}\left(-\frac{1}{\sqrt{2}}\right) = -\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)$$

Using this property of odd function, the equation can be re-expressed.

$$\frac{1}{2}\left\{\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) - \left(\operatorname{erf}\left(-\frac{1}{\sqrt{2}}\right)\right)\right\} = \frac{1}{2}\left\{\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) - -\left(\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)\right)\right\}$$

$$= \frac{1}{2}\left\{2\left(\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)\right)\right\}$$

$$= \operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)$$

$$\therefore \int_{-1}^{1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) \approx 0.6827 = 68.27\%$$

Hence, such conclusion can be made: Given that the performance of a chess player follows normal distribution, the player will perform within a range of 1 standard deviation for approximately 68.27% of the time.

In fact, this sample calculation can be generalized for integration of the PDF of the normal distribution within any standard deviations.

$$\int_{-a}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) \quad \textbf{(3)}$$

Using the equation above, probability of the performance of a player within 2 and 3 standard deviations can be calculated.

Probability of a player to perform within 2σ:

$$\operatorname{erf}\left(\frac{2}{\sqrt{2}}\right) \approx 95.45\ \%$$

Probability of a player to perform within 3σ:

$$\text{erf}\left(\frac{3}{\sqrt{2}}\right) \approx 99.7\ \%$$

Such probability means that the range of a chess player's performance will be within 2 standard deviations from the rating of the player (mean of the normal distribution) for 95.45 % of the time. Hence, it is reasonable to say that the performance of a player can fluctuate within a range of 400 from one's rating, as it is double of 200 which Elo uses as a standard deviation for the distribution (Pelánek, 2016).

Area of intersection between two ratings

Now that we know more about the area under the PDF of the normal distribution, comparing the normal distribution of players will indicate the chance of each player to win, draw or lose. The area under an intersection of two normal distributions is equal to the probability of a draw, since players are performing equally at the area of intersection. As mentioned earlier, different player has different ratings which are the means of their distributions, while all players' standard deviation is fixed to 200 by the Elo system.

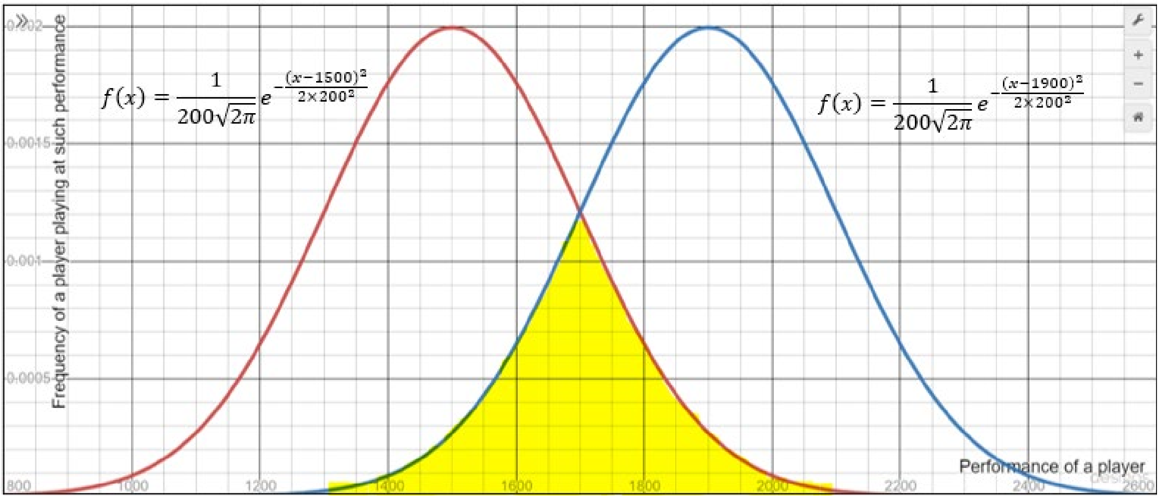
Using the PDF of normal distribution (see equation 1), two normal distributions where μ_1 and μ_2 represent ratings of players 1 and 2 respectively are shown below:

$$normal\ distribution\ of\ player\ 1's\ performance: f(x) = \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\times200^2}}$$

$$normal\ distribution\ of\ player\ 2's\ performance: f(x) = \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\times200^2}}$$

As the equations above show, the only different variable is their rating or mean performance (μ). To help the understanding of how two normal distributions are positioned, player 1 with a rating of 1500 and player 2 with a 1900 are used as an example and were graphed by Desmos (see Graph 1).

Graph 1: Two normal distributions of player 1 (μ_1=1500) and player 2 (μ_2=1900) and highlighted area of intersection



To find the area of intersection, the intersecting point is first required, a point which both x and y coordinates satisfy both functions.

$$x - coordinate\ of\ intersecting\ point:$$

$$\frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\times 200^2}} = \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\times 200^2}}$$

Divide both side by $\frac{1}{200\sqrt{2\pi}}$,

$$e^{\frac{-(x-\mu_1)^2}{2\times 200^2}} = e^{\frac{-(x-\mu_2)^2}{2\times 200^2}}$$

Take natural logarithm on each side,

$$\frac{-(x-\mu_1)^2}{2\times 200^2} = \frac{-(x-\mu_2)^2}{2\times 200^2}$$

Multiply -2×200^2 on each side,

$$(x-\mu_1)^2 = (x-\mu_2)^2$$

$$2x = \frac{{\mu_2}^2 - {\mu_1}^2}{(\mu_2-\mu_1)}$$

$$2x = \frac{(\mu_2+\mu_1)(\mu_2-\mu_1)}{(\mu_2-\mu_1)}$$

$$x = \frac{\mu_1+\mu_2}{2}$$

$$= mean\ of\ two\ ratings$$

$$\therefore x - coordinate\ of\ the\ intersecting\ point\ is\ the\ mean\ of\ two\ ratings\ of\ players$$

Now substitute this x-coordinate to find the y-coordinate of intersecting point.

$$y - coordinate\ of\ theintersecting\ point:$$

$$\frac{1}{200\sqrt{2\pi}} e^{\frac{-(\frac{\mu_1+\mu_2}{2}-\mu_1)^2}{2\times200^2}} = \frac{1}{200\sqrt{2\pi}} e^{\frac{-(\frac{\mu_1+\mu_2}{2}-\mu_2)^2}{2\times200^2}}$$

$$\frac{1}{200\sqrt{2\pi}} e^{-\frac{(\frac{\mu_2-\mu_1}{2})^2}{2\times200^2}} = \frac{1}{200\sqrt{2\pi}} e^{-\frac{(\frac{\mu_1-\mu_2}{2})^2}{2\times200^2}}$$

Note that $(\mu_2-\mu_1)^2 = (\mu_1-\mu_2)^2$.

$$\frac{1}{200\sqrt{2\pi}} e^{-\frac{\frac{(\mu_1-\mu_2)^2}{4}}{80000}} = \frac{1}{200\sqrt{2\pi}} e^{-\frac{(\mu_1-\mu_2)^2}{320000}}$$

$$\therefore The\ intersecting\ point\ of\ two\ normal\ distributions\ of\ chess\ players:$$

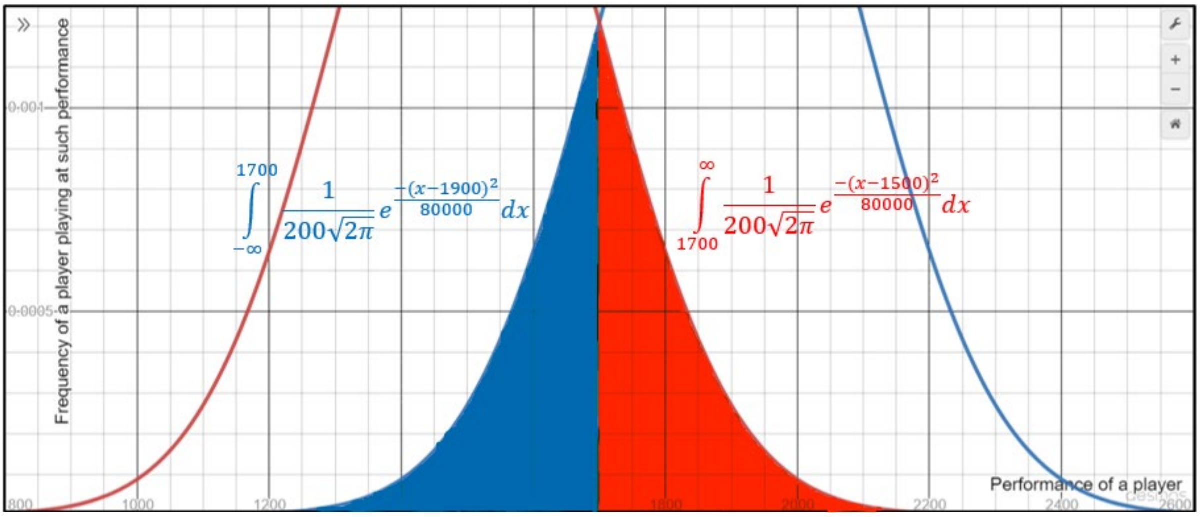
$$\left(\frac{\mu_1+\mu_2}{2}, \frac{1}{200\sqrt{2\pi}} e^{-\frac{(\mu_1-\mu_2)^2}{320000}}\right)$$

Hence, the two normal distributions of the example have an intersecting point at:

$$\left(\frac{1500+1900}{2}, \frac{1}{200\sqrt{2\pi}} e^{-\frac{(1500-1900)^2}{320000}}\right) = (1700{,}0.00121)$$

At this intersecting point, I realized that the area of the intersection is equally divided into two sections (see Graph 2).

Graph 2: Intersecting area divided into two equal parts, where each part is definite integral of each distribution



Thus, the intersected area of two ratings, which are two normal distributions with same standard deviation, can be summed up as following:

$$Intersecting\ area\ = Probability\ of\ drawing = \mathrm{P(D)}$$

$$= \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{80000}} dx + \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{80000}} dx$$

Since the two definite integrals above have identical value, the equation below is simplified by multiplying two of the same definite integral.

$$\mathrm{P(D)} = 2\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \frac{1}{200\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{80000}} dx$$

$$= 2 \times \frac{1}{200\sqrt{2\pi}} \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} e^{\frac{-(x-\mu_1)^2}{80000}} dx$$

$$= \frac{1}{100\sqrt{2\pi}} \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} e^{\frac{-(x-\mu_1)^2}{80000}} dx$$

The method of integration by substitution is used.

$$substitute\ u = \frac{x-\mu_1}{\sqrt{80000}}$$

$$\frac{1}{100\sqrt{2\pi}} \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} e^{\frac{-(x-\mu_1)^2}{80000}} dx = \frac{1}{100\sqrt{2\pi}} \int_{\frac{\left(\frac{\mu_1+\mu_2}{2}\right)-\mu_1}{\sqrt{80000}}}^{\infty} e^{-u^2}\sqrt{80000}\, du$$

$$= \frac{1}{100\sqrt{2\pi}} \times \sqrt{80000} \int_{\frac{\mu_2-\mu_1}{2\sqrt{80000}}}^{\infty} e^{-u^2} du$$

$$= \frac{2}{\sqrt{\pi}} \int_{\frac{\mu_2-\mu_1}{2\sqrt{80000}}}^{\infty} e^{-u^2} du$$

$$Since\ \mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int e^{-u^2}\, du,$$

$$\frac{2}{\sqrt{\pi}} \int_{\frac{\mu_2-\mu_1}{2\sqrt{80000}}}^{\infty} e^{-u^2} du = [\mathrm{erf}(u)]_{\frac{\mu_2-\mu_1}{2\sqrt{80000}}}^{\infty}$$

We found earlier that the error function represents the area of under the curve of the PDF of normal distribution (see equation 3). Hence $\mathrm{erf}(\infty)$ covers all area under the curve of the normal distribution which is 1.

$$Substitute\ \mathrm{erf}(\infty) = 1,$$

$$\left[\mathrm{erf}(\infty) - \mathrm{erf}\left(\frac{\mu_2 - \mu_1}{2\sqrt{80000}}\right)\right] = 1 - \mathrm{erf}\left(\frac{\mu_2 - \mu_1}{400\sqrt{2}}\right)$$

$$\therefore P(D) = 1 - \mathrm{erf}\left(\frac{\mu_2 - \mu_1}{400\sqrt{2}}\right)$$

So, coming back to the example,

$$P(D) = 1 - \mathrm{erf}\left(\frac{1900 - 1500}{400\sqrt{2}}\right)$$

$$= 1 - \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right)$$

$$\approx 31.7\%$$

Hence, a chess game against two players with ratings of 1500 and 1900, or rather rating difference of 400, will have approximately 31.7% as a probability to draw.

Logistic distribution of a comparison between players' performance

However, the Elo's assumption that chess player's performance follows normal distribution was later found not to be true, as it failed to accurately represent the outcomes of games, particularly for the players with lower ratings, since their understanding of the game was not good enough to result in their performance to following a normal distribution (Tenkanen, 2019). Hence, a logistic distribution, which unlike normal distribution, has slightly longer tails and no shape parameter, was introduced and found to be more suitable with the actual results (Tenkanen, 2019). The rating was set so that if a player has a rating that is 400 points more than another player, they are 10 times more likely to win, and this magnitude of 'more likely to win' is increased by a factor of 10 for every 400 more points (Numberphile, 2019). In other words, if a player has a rating 800 points more than another player, they are 100 times more likely to win, and extra 1200 points will result in 1000 times more likeliness to win and so on. With this condition, the expected scores of the players can be calculated as following:

$$Let\ R_A = rating\ of\ player\ A, R_B = rating\ of\ player\ B,$$

$$E_A = expected\ score\ of\ player\ A\ and\ E_B = expected\ score\ of\ player\ B$$

$$E_A = 10^{\frac{R_A - R_B}{400}} \times E_B\ (Numberphile, 2019) \quad \textbf{(4)}$$

This expression represents the expected score of player A in terms of the expected score of player B. As can be seen from the equation above, the difference in ratings between players determines the degree to which a higher rated player has higher expected score than a lower rated player.

For example, if player A has a higher rating than player B by 800, expected score of player A will be 100 times that of the player B's.

$$E_A = 10^{\frac{R_A - R_B}{400}} \times E_B$$

$$= 10^{\frac{800}{400}} \times E_B$$

$$= 100E_B$$

Equation 4 can be expressed as following, since sum of expected scores of players will always be 1, as there will be an outcome at the end of every chess game.

$$E_A + E_B = 1$$

Substitute $E_B = 1 - E_A$ into equation 4,

$$E_A = 10^{\frac{R_A - R_B}{400}} \times E_B$$

$$E_A = 10^{\frac{R_A - R_B}{400}} \times (1 - E_A)$$

$$\frac{E_A}{1-E_A} = 10^{\frac{R_A-R_B}{400}}$$

$$E_A = \frac{10^{\frac{R_A-R_B}{400}}}{1+10^{\frac{R_A-R_B}{400}}}$$

To make this equation look like a logistic distribution, which makes it easier to interpret, the numerator of the fraction above was rearranged as following:

$$E_A = \frac{10^{\frac{R_A-R_B}{400}}}{1+10^{\frac{R_A-R_B}{400}}} \times \frac{\left(\frac{1}{10^{\frac{R_A-R_B}{400}}}\right)}{\left(\frac{1}{10^{\frac{R_A-R_B}{400}}}\right)}$$

$$= \frac{1}{\frac{1}{10^{\frac{R_A-R_B}{400}}}+1}$$

$$\therefore E_A = \frac{1}{1+10^{\frac{R_B-R_A}{400}}} \quad \textbf{(5)}$$

Now the equation for expected score of player A (Veisdal, 2019) is in a form of logistic distribution (see equation 5), since it has similar components of the cumulative density function of logistic distribution as shown below.

$$\textit{Cumulative density function of logistic distribution } \text{(Khan, 2017):}$$

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}}$$

$$,\textit{where } L = \textit{the curve's maximum value},$$

$$x_0 = \textit{the } x \textit{ value of the sigmoid's } (s-\textit{shaped curve}) \textit{ midpoint},$$

$$k = \textit{the steepness of the curve}$$

As can be seen (see equation 5), 1 is the highest expected score of player A, which would be the case when the rating of player A is significantly larger than the rating of player B.

Continuing with the example from the probability of a draw, expected score of player A (rating = 1500) against player B (rating = 1900) will be as follows:

$$E_A = \frac{1}{1 + 10^{\frac{R_B - R_A}{400}}}$$

$$= \frac{1}{1 + 10^{\frac{1900-1500}{400}}}$$

$$= \frac{1}{11}$$

$$\approx 0.0909$$

This leads player B to have an expected score of $1 - 0.0909 = 0.909$, which corroborates the assumption that player with 400 more rating points will have 10 times higher expected score than the opponent's expected score.

Interpretation of expected score

The expected score is what the player should theoretically score based on the significance of their rating to their opponent's rating. In chess, win, draw and loss equates to an actual score of 1, 0.5 and 0, respectively. Therefore, the expected score is a sum of three outcomes with their respective scores.

$$Let\ E = expected\ score\ \times 100\ (\%),$$

$$P(W) = probability\ of\ winning,$$

$$P(D) = probability\ of\ drawing,$$

$$P(L) = probability\ of\ losing$$

$$E = 1 \times P(W) + 0.5 \times P(D) + 0 \times P(L) \qquad \textbf{(6)}$$

For example, the expected score of 40 % does not only mean that a player has 40 % chance of winning and 60 % chance of losing, but can also represent 30 % chance of winning, 20 % chance of drawing and 50 % chance of losing as explained below.

$$If\ E = 40\%,$$

$$The\ following\ set\ of\ probabilities\ works: P(W) = 40\%, P(D) = 0\%, P(L) = 60\%$$

$$E = 1 \times 40\% + 0.5 \times 0\% + 0 \times 60\% = 40\%$$

$$This\ set\ of\ probabilities\ also\ works: P(W) = 30\%, P(D) = 20\%, P(L) = 50\%$$

$$E = 1 \times 30\% + 0.5 \times 20\% + 0 \times 50\% = 40\%$$

In other words, the expected score suggests more than one set of probabilities, which I summarized with his own parameters as shown above (see equation 6). The parametric equation can be simplified as following.

$$E = 1 \times P(W) + 0.5 \times P(D) + 0 \times P(L)$$

$$\therefore E = P(W) + 0.5 \times P(D)$$

Clearly, the sum of a set of probabilities is always 100 %, as any chess game is guaranteed to have an ending of either win, draw or loss.

$$P(W) + P(D) + P(L) = 100\% \quad \text{(7)}$$

This means that a set of probabilities is bound to a certain range. To calculate a range of each probability based E, they were expressed in terms of $P(W)$.

$$E = P(W) + 0.5 \times P(D)$$

$$E - P(W) = 0.5 \times P(D)$$

$$2(E - P(W)) = P(D)$$

Using equation 7, the equation for probability of drawing can be used to find the probability of losing.

$$P(W) + P(D) + P(L) = 100\%$$

$$P(L) = 100 - \big(P(W) + P(D)\big)$$

Substitute $2\big(E - P(W)\big) = P(D)$,

$$P(L) = 100 - \big[P(W) + 2\big(E - P(W)\big)\big]$$

$$\therefore For\ any\ E\ and\ P(W),$$

$$P(D) = 2\big(E - P(W)\big)$$

$$P(L) = 100 - \big[P(W) + 2\big(E - P(W)\big)\big]$$

Of course, probability of winning can be as low as zero to as high as the expected score:

$$\{P(W) | 0 \leq P(W) \leq E\}$$

Each lower bound (0) and upper bound (E) of $P(W)$ was substituted into the equation for $P(D)$ to find its domain.

$$P(D) = 2\big(E - P(W)\big)$$

$$If\ P(W) = 0,$$

$$P(D) = 2(E - 0)$$

$$= 2E$$

$$If\ P(W) = E$$

$$P(D) = 2(E - E)$$

$$= 0$$

$$\therefore \{P(D) | 0 \leq P(D) \leq 2E\}$$

Domain of $P(L)$ was found by the same process.

$$P(L) = 100 - [P(W) + 2(E - P(W))]$$

$$If\ P(W) = 0$$

$$P(L) = 100 - [0 + 2(E - 0)]$$

$$= 100 - 2E$$

$$If\ P(W) = E$$

$$P(L) = 100 - [E + 2(E - E)]$$

$$= 100 - E$$

$$\therefore \{P(L)|100 - 2E \leq P(L) \leq 100 - E\}$$

$P(W), P(D)$ and $P(L)$ was then graphed at E of 40 % to help the visualization of how each probability alters as $P(W)$ changes (see Graph 3). The graph shows all sets of possibilities of each outcome (win, draw or loss) when the expected score is fixed to 40 %. Firstly, it corroborates the boundary for each outcome.

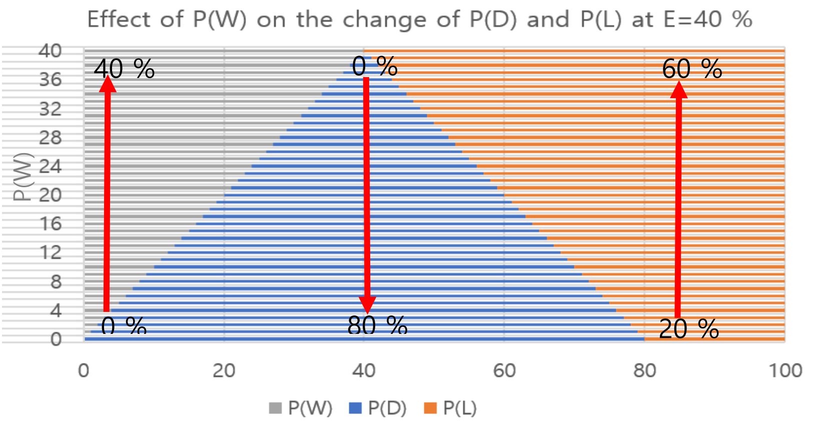
$$When\ E = 40\%,$$

$$\{P(W)|0 \leq P(W) \leq E\} = \{P(W)|0 \leq P(W) \leq 40\}$$

$$\{P(D)|0 \leq P(D) \leq 2E\} = \{P(D)|0 \leq P(D) \leq 80\}$$

$$\{P(L)|100 - 2E \leq P(L) \leq 100 - E\} = \{P(L)|20 \leq P(L) \leq 60\}$$

<u>Graph 3: Effect of P(W) on the change of P(D) and P(L) at E=40 %</u>



Also, it can be seen from the graph that as $P(W)$ increases to its upper boundary (40 %), $P(L)$ also increases to its upper boundary (60 %), while $P(D)$ decreases to its lower boundary (0 %).

To summarize, each player is given an expected score by equation 5 based on the difference between their rating and their opponent's rating. The expected score provides a good representation of how the game should end, which has multiple scenarios for each distinctive value of expected score (see Graph 3).

Equation of new rating

However, actual score may differ from the expected score, especially when the number of games increases. Hence, the difference between what was expected by the algorithm and what has actually happened determines the change of their rating after a game or series of games as following (Sas, 2020):

$$R_N = R_A + K(S_A - E_A) \quad \textbf{(8)}$$

$$, where\ R_N = new\ rating\ of\ player\ A$$

$$R_A = rating\ of\ player\ A$$

$$K = K\ factor\ (constant)$$

$$S_A = actual\ score\ of\ player\ A$$

$$E_A = expected\ score\ of\ player\ A$$

The updated rating will increase if an actual score of a player is higher than the expected score and vice versa, since the fact that players performed better than what the system expected indicates that their performance deserves higher rating. This mechanism is explained as following:

$$if\ S_A > E_A,$$

$$S_A - E_A > 0$$

$$K(S_A - E_A) > 0$$

Substitute this inequality into equation 8,

$$R_N = R_A + K(S_A - E_A)$$

$$R_N - R_A = K(S_A - E_A)$$

$$R_N - R_A > 0$$

$$R_N > R_A$$

$$\therefore New\ rating\ is\ increased\ if\ S_A > E_A$$

In this equation, the K-factor also plays a role in the magnitude of the rating fluctuation as a low value of K-factor will not consider the difference of actual result and expected result as significant as a higher value of K-factor does (The Internet Chess Club, 2002). In fact, the K-factor for grandmasters (players with rating higher than 2500) is 10, which is much lower than players with lower rating, where their K-factor varies from 20 to 40 in respect to their divisions (The Internet Chess Club, 2002). This is because high K-factor changes players' rating dramatically, which is not necessary for players in the highest division, where their rating has been built over thousands of games.

For example, if a grandmaster A with a rating of 2700 won against grandmaster B with a rating of 2600, the grandmaster A would have gained different rating compared to when player C with a rating of 1600 won against player D with a rating of 1500, since player A and C are in different divisions with different K-factors, although the outcome of both players and the rating difference (100) are same. The new ratings of player A and C can be calculated as following to show how different K-factor determines the magnitude of change in rating.

Expected score of player A (see equation 5):

$$E_A = \frac{1}{1 + 10^{\frac{R_B - R_A}{400}}}$$

$$= \frac{1}{1 + 10^{\frac{2600-2700}{400}}}$$

$$\approx 0.64$$

New rating of player A (see equation 8):

$$R_N = R_A + K(S_A - E_A)$$

Substitute $R_A = 2700,\ S_A = 1\ (since\ player\ A\ won), E_A = 0.64$ into equation 8,

$$R_N = 2700 + K(1 - 0.64)$$

$$R_N = 2700 + 0.36K$$

Substitute $K = 10,$ since player A is a grandmaster,

$$R_N = 2700 + 0.36 \times 10$$

$$R_N = 2703.6$$

$$R_N = 2704$$

Hence, player A gained 4 rating points from winning a game against player B with a rating of 2600.

Same calculation was used to find the expected score and the new rating of player C.

Expected score of player C (see equation 5):

$$E_C = \frac{1}{1+10^{\frac{R_B-R_A}{400}}}$$

$$= \frac{1}{1+10^{\frac{1500-1600}{400}}}$$

$$\approx 0.64$$

New rating of player C (see equation 8):

$$R_N = R_C + K(S_C - E_C)$$

Substitute $R_C = 1600,\ S_C = 1\ (since\ player\ A\ won), E_C = 0.64$ into equation 8,

$$R_N = 1600 + K(1 - 0.64)$$

$$R_N = 1600 + 0.36K$$

Substitute $K = 20,$ since this value corresponds to a division which player C is in,

$$R_N = 1600 + 0.36 \times 20$$

$$R_N = 1607.2$$

$$R_N = 1607$$

Hence, player C gained 7 rating points from winning a game against player D with a rating of 1500, while player A gained 4 rating points. As the example calculation shows, the K-factor determines how much rating a player gains or loses, along with the rating difference between players.

It should also be noted that equation 8 is limited to update the rating of a single game, meaning the equation must be used after every game. However, simply inserting sigma notation can solve the problem and allow this to be used for many games or competitions.

$$R_N = R_A + K\sum_{B=1}^{n}(S_A - E(s|R_A, R_B) \quad \textbf{(9)}$$

$$, where\ R_N = new\ rating\ of\ player\ A$$

$$R_A = rating\ of\ player\ A$$

$$K = K\ factor\ (constant)$$

$$n = number\ of\ games$$

$$S_A = actual\ score\ of\ player\ A$$

$$E(s|R_A, R_B) = expcted\ score\ given\ the\ rating\ of\ both\ players$$

Overall, Elo rating system appropriately predicts the outcome of a game and alters ratings of players based on K-factor and the difference between actual score and expected score.

Glicko rating system

Introduction of rating deviation (RD)

In Elo rating system, a player may simply not play after reaching their highest rating and that rating will be treated equally to other ratings, leading to their performance being worse than predicted. Hence, the Glicko rating system was developed by Dr. Mark E. Glickman in 1995, as an extension of Elo rating system, which takes the reliability of a player’s rating into consideration. This is why the Glicko rating system consists of another parameter called rating deviation (RD), which indicates the uncertainty of a player’s rating (Glickman, The Glicko system, 2016). Unlike Elo rating system, where the standard deviation of all players’ rating was fixed to 200, Glicko RD fluctuates depends on the degree of player’s activity in chess, which increases as the period of inactivity increases (Glickman, The Glicko system, 2016). Note that new players are given the maximum RD (350), as insufficient game records of them make their rating very unreliable.

$$RD = \sqrt{RD_{old}^2 + c^2 t}, for\ 0 < RD \le 350 \qquad \textbf{(10)}$$

$$, where\ RD = updated\ rating\ deviation$$

$$RD_{old} = previous\ rating\ deviation$$

$$c = constant$$

$$t = period\ of\ inactivity\ (months)$$

Although I attempted to derive equation 10 to fully deconstruct the meaning of the equation like I did for all the equations in the Elo rating system, the equation will be used as an accepted fact without going through any derivation, since such process was beyond the scope of high school mathematics.

The role of constant c

The constant c plays an important role of adjusting the rate of change of RD, like the k-factor. The value of c can be calculated by letting each variable equal to certain value with appropriate reason. For example, suppose RD_{old} is 50, as it is a reasonably common RD. Let the updated rating deviation (RD) be 350, which means that the period of inactivity represents how much time would need to pass before a rating for a typical player to become as uncertain as that of a new player. Hence, t was chosen to be 100 periods (months), a reasonable assumption that if a player with RD of 50 does not play for 100 months, the player's RD becomes as unreliable as that of a new player. Since all the variables except the constant c have been chosen, c can be calculated with rearranging the equation 10.

$$let\ RD = 350, RD_{old} = 50, t = 100$$

$$RD = \sqrt{RD_{old}^2 + c^2 t}$$

$$350 = \sqrt{50^2 + 100c^2}$$

$$c = \sqrt{\frac{350^2 - 50^2}{100}}$$

$$c \approx 34.6$$

Now that we found the value of c for the sample calculation, we can create parametric equation followed by substitution of each variable to graph.

$$RD = \sqrt{RD_{old}^2 + c^2 t}$$

$$let\ RD_{old} = O, c = C\ and\ t = T$$

$$RD = \sqrt{O^2 + C^2 T} \quad \textbf{(11)}$$

$$Substitue\ O = 50, C = 34.6,$$

$$RD = \sqrt{50^2 + 34.6^2 T}$$

$$\therefore Equation\ for\ updated\ RD\ with\ repect\ to\ T: RD = \sqrt{50^2 + 34.6^2 T}$$

Same process was used to find equations for updated RD with respect to each remaining variable.

$$Substitue\ O = 50, T = 100\ into\ eqution\ 11,$$

$$RD = \sqrt{50^2 + 100O^2}$$

$$\therefore Equation\ for\ updated\ RD\ with\ repect\ to\ C: RD = \sqrt{50^2 + 100C^2}$$

$$Substitue\ C = 34.6, T = 100\ into\ eqution\ 11,$$

$$RD = \sqrt{O^2 + 100 \times 34.6^2}$$

$$\therefore Equation\ for\ updated\ RD\ with\ repect\ to\ O: RD = \sqrt{O^2 + 100 \times 34.6^2}$$

I was interested in seeing how this relationship between the RD and each variable changes at what rate. To further explore each variable's instantaneous rate of change, partial differentiation was used (see equation 11).

Partial derivative with respective to O (old RD):

$$\frac{\partial RD}{\partial O}\left(\sqrt{O^2+C^2T}\right)=\frac{\partial RD}{\partial O}(O^2+C^2T)^{\frac{1}{2}}$$

$$=\frac{1}{2}\times(O^2+C^2T)^{\frac{-1}{2}}\times 2O\ (using\ chain\ rule)$$

$$=\frac{O}{\sqrt{O^2+C^2T}}$$

Partial derivative with respective to C (constant):

$$\frac{\partial RD}{\partial C}\left(\sqrt{O^2+C^2T}\right)=\frac{\partial RD}{\partial C}(O^2+C^2T)^{\frac{1}{2}}$$

$$=\frac{1}{2}\times(O^2+C^2T)^{\frac{-1}{2}}\times 2CT\ (using\ chain\ rule)$$

$$=\frac{CT}{\sqrt{O^2+C^2T}}$$

Partial derivative with respective to T (period of inactivity):

$$\frac{\partial RD}{\partial T}\left(\sqrt{O^2+C^2T}\right)=\frac{\partial RD}{\partial T}(O^2+C^2T)^{\frac{1}{2}}$$

$$=\frac{1}{2}\times(O^2+C^2T)^{\frac{-1}{2}}\times C^2\ (using\ chain\ rule)$$

$$=\frac{C^2}{2\sqrt{O^2+C^2T}}$$

To visually represent these partial derivates on the same set of axes (see Graph 4), the variables, which were treated as constants for each partial derivative, was substitued by the values from the sample calculation.

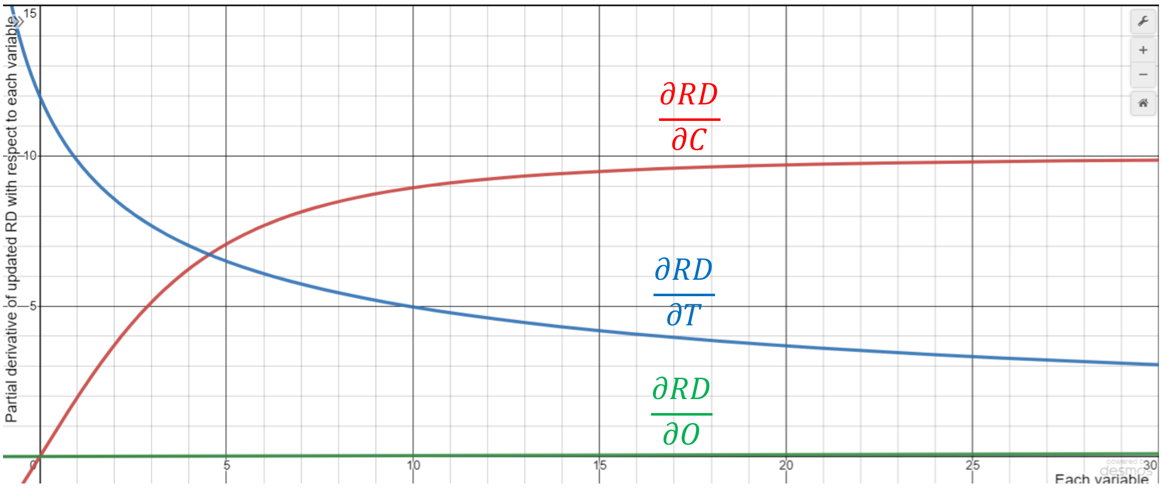
Substitue $O = 50, C = 34.6 \; and \; T = 100$ appropriately,

$$\frac{\partial RD}{\partial O} = \frac{O}{\sqrt{O^2 + C^2T}} = \frac{O}{\sqrt{O^2 + 34.6^2 \times 100}}$$

$$\frac{\partial RD}{\partial C} = \frac{CT}{\sqrt{O^2 + C^2T}} = \frac{100C}{\sqrt{50^2 + 100C^2}}$$

$$\frac{\partial RD}{\partial T} = \frac{C^2}{2\sqrt{O^2 + C^2T}} = \frac{34.6^2}{2\sqrt{50^2 + 34.6^2T}}$$

Graph 4: 3 Partial derivatives of updated RD with respect to each variable



The graph only shows 1st quadrant of the cartesian plane since all values of variables are positive in this case. As shown by the graph the partial derivative with respect to C is approaching to a maximum asymptote. This maximum asymptote will imply the maximum rate of change that C has over the change of updated RD, which was calculated as following:

$$\lim_{C\to\infty} \frac{\partial RD}{\partial C} = \lim_{C\to\infty} \frac{100C}{\sqrt{2500 + 100C^2}}$$

$$= 100\left(\lim_{C\to\infty} \frac{C}{\sqrt{2500 + 100C^2}}\right)$$

$$= 100\left(\lim_{C\to\infty} \frac{C}{C\sqrt{\frac{2500}{C^2} + \frac{100C^2}{C^2}}}\right)$$

$$= 100\left(\lim_{C\to\infty}\frac{1}{\sqrt{\frac{2500}{C^2}+100}}\right)$$

$$= 100\left(\frac{\lim_{C\to\infty} 1}{\lim_{C\to\infty}\sqrt{\frac{2500}{C^2}+100}}\right)$$

$$= 100 \times \frac{1}{\sqrt{0+100}}$$

$$= 10$$

Hence, the above calculation indicates that, no matter how large the value of constant c is set to be in the system, the maximum impact the constant can have on updated RD's instantaneous rate of change is 10 units.

Graph 4 further shows partial derivative of updated RD with respect to old RD has almost flat gradient, indicating there is practically no change in updated RD, as the value of old RD changes. This means that the change in updated RD heavily relies on the values of c (constant) and t (period of inactivity).

Approximating maximum RD

As explained in previous section, the maximum RD in the Glicko rating system is set to be 350 (see equation 10), which is given to a new player or can be achieved by a player who has not played chess for long enough that their RD equates to new players' RD.

As explained in previous section, higher RD indicates that a player has not been participating in chess for a long period, and thus the reliability of player's RD is also decreased. This reliability is another important variable in the Glicko rating system, as I thought that this can be used to explore the suitable value of maximum RD.

$$Equation\ for\ reliability\ of\ RD\ \text{(Glickman, The Glicko system, 2016):}$$

$$g(RD) = \frac{1}{\sqrt{1 + 3\left(\frac{q(RD)}{\pi}\right)^2}}$$

$$where\ g(RD) = reliability\ of\ RD$$

$$q = constant$$

$$RD = rating\ deviation\ of\ a\ player$$ **(12)**

Note that the constant $q = \frac{\ln(10)}{400}$.

I tried to understand how this equation was derived and the role of the constant q. To clarify these inquiries, I emailed Dr. Glickman, the founder of the Glicko rating system, and received a response that "q is introduced in order to translate $\frac{1}{1+10^{-(R_A - R_B)}}$ to $\frac{1}{1+e^{-(R_A - R_B)}}$ (Glickman, Inquires about Glicko rating system, 2021)".

With this equation 12, I tried to explain why the maximum RD is 350, by examining the rate of change of reliability of RD ($g(RD)$).

Firstly, the constant q was substituted to numeric value.

$$Substitue\ q = \frac{\ln(10)}{400}\ into\ equation\ 12$$

$$g(RD) = \frac{1}{\sqrt{1 + 3\left(\frac{q(RD)}{\pi}\right)^2}}$$

$$= \frac{1}{\sqrt{1 + 3\left(\frac{[\ln(10)](RD)}{400\pi}\right)^2}}$$

Now the function $g(RD)$ was differentiated.

$$g'(RD)\left(\frac{1}{\sqrt{1+3\left(\frac{[\ln(10)](RD)}{400\pi}\right)^2}}\right)=g'(RD)\left(1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right)^{\frac{-1}{2}}$$

The chain rule($\frac{d}{dx}\left[\left(f(x)\right)^n\right]=n\left(f(x)\right)^{-1}\times f'(x)$) was used (Haese, Humphries, Sangwin, & Vo, 2019).

$$g'(RD)=\frac{-1}{2}\left(1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right)^{\frac{-3}{2}}\times g'(RD)\left(1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right)$$

$$=\frac{-1}{2\sqrt{\left[1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right]^3}}\times\frac{6\ln^2(10)\,(RD)}{160000\pi^2}$$

$$=\frac{-3\ln^2(10)\,(RD)}{160000\pi^2\sqrt{\left[1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right]^3}}$$

From this, I thought converting denominator of the fraction above to binomial expression was an appropriate method to calculate the maximum RD, as such form in binomial expression allowed the test for convergence to happen. The reason why the test for convergence is thought to be a useful method for this case is explained as following:

Any number that is not included in the interval will result the expression to diverge, which for this case, makes the rate of change of reliability of player's RD ($g'(RD)$) uninterpretable. Hence, I thought that the upper limit of the interval would be the maximum RD as any higher RD will make the denominator of the $g'(RD)$ diverging, which makes the $g'(RD)$ diverging as well.

Converting the denominator of the $g'(RD)$ into binomial expression was done as following:

$$g'(RD) = \frac{-3\ln^2(10)\,(RD)}{160000\pi^2\sqrt{\left[1+3\left(\frac{\ln(10)\,(RD)}{400\pi}\right)^2\right]^3}}$$

$$= \frac{-3\ln^2(10)\,(RD)}{160000\pi^2\left[\frac{160000\pi^2+3\ln^2(10)\,(RD)^2}{160000\pi^2}\right]^{\frac{3}{2}}}$$

$$= \frac{-3\ln^2(10)\,(RD)}{160000\pi^2\,\frac{\left[160000\pi^2+3\ln^2(10)\,(RD)^2\right]^{\frac{3}{2}}}{160000^{\frac{3}{2}}\pi^3}}$$

$$= \frac{-1200\pi\ln^2(10)(RD)}{\left[160000\pi^2+3\ln^2(10)(RD)^2\right]^{\frac{3}{2}}}$$

Now that the denominator of the equation above is in binomial expression, following calculations were carried out to determine whether the expression is convergence or divergence series:

$$For\ n\in\mathbb{Q},\ (a+bx)^n = a^n\sum_{r=0}^{\infty}\binom{n}{r}\left(\frac{bx}{a}\right)^r\quad \text{(Haese, Humphries, Sangwin, \& Vo, 2019)}$$

$$\left[160000\pi^2+3\ln^2(10)\,(RD)^2\right]^{\frac{3}{2}} = (160000\pi^2)^{\frac{3}{2}}\left[1+\frac{3\ln^2(10)(RD)^2}{160000\pi^2}\right]^{\frac{3}{2}}$$

$$= 64000000\pi^3\sum_{r=0}^{\infty}\binom{\frac{3}{2}}{r}\left(\frac{3\ln^2(10)(RD)^2}{160000\pi^2}\right)^{\frac{3}{2}}$$

This expression needs to have an absolute value of r as less than 1 for the binomial expansion to converge (Haese, Humphries, Sangwin, & Vo, 2019),

$$\textit{The interval of convergence}: \left|\frac{bx}{a}\right| < 1$$

$$\left|\frac{3\ln^2(10)\,(RD)^2}{160000\pi^2}\right| < 1$$

This means,

$$\frac{-3\ln^2(10)\,(RD)^2}{160000\pi^2} < 1 \;and\; \frac{3\ln^2(10)\,(RD)^2}{160000\pi^2} < 1$$

$$\rightarrow \frac{3\ln^2(10)\,(RD)^2}{160000\pi^2} > -1 \;and\; \frac{3\ln^2(10)\,(RD)^2}{160000\pi^2} < 1$$

$$\rightarrow \frac{3\ln^2(10)\,(RD)^2}{160000\pi^2} + 1 > 0 \;and\; \frac{3\ln^2(10)\,(RD)^2}{160000\pi^2} - 1 < 0$$

$$\rightarrow 3\ln^2(10)\,(RD)^2 + 160000\pi^2 > 0 \;and\; 3\ln^2(10)\,(RD)^2 - 160000\pi^2 < 0$$

$$\rightarrow (RD) > -\frac{400\pi}{\sqrt{3}\ln(10)} \;and\; (RD) < \frac{400\pi}{\sqrt{3}\ln(10)}$$

$$\therefore \left|\frac{3\ln^2(10)\,(RD)^2}{160000\pi^2}\right| < 1 \rightarrow \frac{-400\pi}{\sqrt{3}\ln(10)} < (RD) < \frac{400\pi}{\sqrt{3}\ln(10)}$$

$$\frac{400\pi}{\sqrt{3}\ln(10)} \cong 315$$

$$since\ (RD) > 0, 0 < (RD) < 315$$

$$\therefore Any\ RD\ less\ than\ 315\ will\ lead\ the\ g'(RD)\ to\ be\ convergent.$$

Such finding means that having RD of equal to or less than 315 will lead the denominator of the derivative of g(RD) to settle towards a certain value, which provides a distinct value for the derivative. In other words, the reliability of RD's rate of change will only be interpretable when the RD is less than 315, as higher RD than 315 will lead the denominator of the derivative to be divergent and thus resulting diverging rate of change. Hence, I thought that setting maximum RD as 315 was appropriate. However, the original equation by Glickman manipulated 350 as the maximum RD, which is not too different from the maximum RD calculated by me.

Expected score and equation of new rating

The way the expected score is determined in the Glicko rating system is based on the expected score of the Elo rating system, except the expected score of the Glicko rating system takes g(RD) into consideration.

$$Equation\ 5: E(S|R_A, R_B) = \frac{1}{1 + 10^{\frac{R_B - R_A}{400}}}$$

$\rightarrow Expected\ score\ of\ player\ A\ in\ the\ Glicko\ rating\ system$ (Glickman, The Glicko system, 2016):

$$E(S_A|R_A, R_B, RD_B) = \frac{1}{1 + 10^{g(RD_B)(R_B - R_A)/400}} \quad \textbf{(13)}$$

$$, where\ E(S_A|R_A, R_B, RD_B) = expected\ score\ of\ player\ A, given\ rating\ of\ player\ A,$$

$$rating\ of\ player\ B\ and\ rating\ deviation\ of\ player\ B$$

$$g(RD_B) = reliability\ of\ rating\ deviatin\ of\ player\ B$$

Along with the modified expected score, equation for updated rating is also modified with consideration of variance (δ^2) of the posterior distribution. Note that the same variables are used as equation 13.

$Variance\ in\ the\ Glicko\ rating\ system$ (Glickman, The Glicko system, 2016):

$$\delta^2 = \left[q^2 \sum_{B=1}^{n} \left(g(RD_B)\right)^2 E(S|R_A, R_B, RD_B)\{1 - E(S_A|R_A, R_B, RD_B)\}\right]^{-1} \quad \textbf{(14)}$$

$$, where\ q\ (constant) = \frac{\ln(10)}{400}$$

With the modified expected score and introduction of variance, more sophisticated version of equation for updated rating in the Glicko rating system was developed.

$$Equation\ 9: R_N = R_A + K \sum_{B=1}^{n} \left(S_A - E(s|R_A, R_B) \right)$$

$\rightarrow Equation\ for\ updated\ rating\ in\ the\ Glicko\ rating\ system$ (Glickman, The Glicko system, 2016)

$$R_N = R_A + \frac{q}{\frac{1}{(RD_A)^2} + \frac{1}{\delta^2}} \sum_{B=1}^{n} g(RD_B)\{S_A - E(S_A|R_A, R_B, RD_B)\} \quad \textbf{(15)}$$

Consequently, rating and RD of each player is required to compute equations in the Glicko rating system, as supposed to the Elo rating system which only requires rating of each player.

For example, the updated rating of first two games in 10 minutes format (see appendix 1) will be as follows, according to the Glicko rating system:

$$The\ author's\ rating\ and\ RD: (R_A, RD_A) = (1343, 36)$$

$$The\ first\ opponent's\ rating\ and\ RD: (R_{B1}, RD_{B1}) = (1322, 51)$$

$$The\ second\ opponent's\ rating\ and\ RD: (R_{B2}, RD_{B2}) = (1251, 28)$$

Using equation 12, the reliability of the first opponent's RD was calculated.

$$g(RD_{B1}) = \frac{1}{\sqrt{1 + 3\left(\frac{q(RD_{B1})}{\pi}\right)^2}}$$

$$= \frac{1}{\sqrt{1 + 3\left(\frac{\ln(10) \times 51}{400\pi}\right)^2}}$$

$$\approx 98.715\ \%$$

$$Substitue\ g(RD_o) = 98.715\%\ into\ equation\ 13$$

$$E(S_A|R_A, R_{B1}, RD_{B1}) = \frac{1}{1 + 10^{g(RD_{B1})(R_{B1}-R_A)/400}}$$

$$= \frac{1}{1 + 10^{\frac{(0.98715)(1322-1343)}{400}}}$$

$$\approx 52.980\ \%$$

Following the same procedure, expected score against second opponent and his/her reliability of RD were calculated.

$$\therefore g(RD_{B2}) = 99.607\ \%\ and\ E(S_A|R_A, R_{B2}, RD_{B2}) = 62.890\ \%$$

Using equation 14, the variance of the game was calculated.

$$\delta^2 = \left[q^2 \sum_{B=1}^{n} \big(g(RD_B)\big)^2 E(S|R_A, R_B, RD_B)\{1 - E(S_A|R_A, R_B, RD_B)\}\right]^{-1}$$

$$= \left(\left(\frac{\ln(10)}{400}\right)^2 \times [0.99607^2 \times 0.62890 \times (1 - 0.62890) + 0.98715^2 \times 0.52980\ \times (1 - 0.52980\,)]\right)^{-1}$$

$$\approx 63625$$

The result of the two games were that I lost (score = 0) the first game but won the second game (score = 1). Hence the updated my rating after the two games were calculated with the above information and equation 15.

$$R_N = R_A + \frac{q}{\frac{1}{(RD_A)^2} + \frac{1}{\delta^2}} \sum_{B=1}^{n} g(RD_B)\{S_A - E(S_A|R_A, R_B, RD_B)\}$$

$$= 1343 + \frac{\left(\frac{\ln(10)}{400}\right)}{\left(\frac{1}{36^2} + \frac{1}{63625}\right)} \times [0.99607 \times (1 - 0.62890) + 0.98715 \times (0 - 0.52980)]$$

$$= 1343 + 7.3114 \times (0.36964 - 0.52992)$$

$$= 1343 - 1.1719 \approx 1342$$

Conclusion

To conclude, both rating systems provide a reasonably accurate way to predict the outcome of a chess game and change a player's rating according to the result of the game. In the Elo rating system, the predicted outcome is dependent on difference between ratings of two players (see equation 5) and K-factor (see equation 8) which is fixed based the players' rating division. This fixed K-factor, which determines the sensitivity of the change in rating, led Glicko rating system to start considering a reasonable way for K-factor to change. The Glicko rating system is more accurate than the Elo system because it ensures that all players' RD are changeable, which depends on their period of inactivity (see equation 10). With this flexible RD, the reliability of RD is also considered in the latter system (see equation 12), which then allows for better approximation of expected scores and updated ratings (see equation 13 and 15). As can be seen in the appendices, the difference between expected score of the games in two systems are negligible (see appendix 1 and 2). However, the two rating systems give disparate change in ratings for both formats of the game. This is because the K-factor for the Elo rating system was fixed to 40, while the Glicko rating system considered the period of inactivity, which resulted two different K-factors for both formats. Because I play chess games in 10 minutes format a lot more frequent than in 3 minutes format, the reliability of RD in the former format is higher than the reliability of RD in the latter format, resulting in less dynamic change of rating (lower K-factor) in the former format compared to the latter format.

Bibliography

Glickman, M. E. (2016, September 10). *The Glicko system.* Retrieved from Mark Glickman's World: http://www.glicko.net/glicko/glicko.pdf

Glickman, M. E. (2021, April 18). Inquires about Glicko rating system. (Y. Hyeonggeun, Interviewer)

Haese, M., Humphries, M., Sangwin, C., & Vo, N. (2019). *Mathematics: Analysis and Approaches HL.* Marleston: Haese Mathematics.

Khan Academy. (2021, n.d. n.d.). *u-substitution with definite integrals*. Retrieved from Khan Academy: https://www.khanacademy.org/math/ap-calculus-ab/ab-integration-new/ab-6-9/a/u-substitution-definite-integrals

Khan, R. (2017, August 20). *Logistic Distribution Basics*. Retrieved from RPubs: https://rpubs.com/riazakhan94/logstcdistbasics

Numberphile. (2019, February 16). *The Elo Rating System for Chess and Beyond* [Video]. YouTube.

https://www.youtube.com/watch?v=AsYfbmp0To0

Pelánek, R. (2016, July n.d.). *Applications of the Elo rating system in adaptive educational systems*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S036013151630080X

Sas, W. (2020, October 29). *Elo Calculator*. Retrieved from Omni Calculator: https://www.omnicalculator.com/sports/elo

Tenkanen, S. (2019, May 23). *Rating National Hockey League teams: the predictive power of Elo rating models in ice hockey.* Retrieved from Aalto University, School of Business: https://aaltodoc.aalto.fi/bitstream/handle/123456789/39337/bachelor_Tenkanen_Santeri_2019.pdf?sequence=1&isAllowed=y

The Internet Chess Club. (2002, October 18). *K-FACTOR*. Retrieved from The Internet Chess Club: https://www.chessclub.com/help/k-factor

Veisdal, J. (2019, September 2). *The Mathematics of Elo Ratings*. Retrieved from Cantor's Paradise: https://www.cantorsparadise.com/the-mathematics-of-elo-ratings-b6bfc9ca1dba

Weisstein, E. W. (2021, August 3). *Erf*. Retrieved from Wolfram MathWorld: https://mathworld.wolfram.com/Erf.html

Appendices

This sample calculation in page 33 to 35 is a part of data in appendix 1. The appendices were created to provide a clear representation of how the two rating systems differ. They include all the relevant data of 50 games that has been played by me and random opponents on the online chess website (chess.com) under two different time limits (10 mins and 3 mins). All the data in the appendices was calculated by using equation 5 and 9 for data related to the Elo rating system and equation 13 to 15 for data related to the Glicko rating system.

Appendix 1: Relevant data of 25 games in 10 mins format

Rapid (10 mins)									
$g(RD_A)$	R_A	RD_A	R_B	RD_B	$g(RD_B)$	S_A	E_A (Elo)	E_A (Glicko)	δ^2
99.354%	1343	36	1322	51	98.715%	0	0.53018	0.52980	0.242752
99.354%	1335	36	1251	28	99.607%	1	0.61858	0.61814	0.234194
99.354%	1341	36	1330	21	99.779%	1	0.51582	0.51579	0.248646
99.389%	1349	35	1348	26	99.661%	1	0.50144	0.50143	0.248307
99.389%	1357	35	1387	32	99.488%	0	0.45693	0.45715	0.245631
99.389%	1350	35	1338	28	99.607%	1	0.51726	0.51719	0.247748
99.389%	1358	35	1352	20	99.799%	1	0.50863	0.50862	0.248923
99.389%	1366	35	1416	26	99.661%	1	0.42854	0.42878	0.243271
99.389%	1375	35	1400	41	99.164%	0	0.46408	0.46438	0.24459
99.423%	1368	34	1341	30	99.550%	0	0.53878	0.53860	0.246277
99.423%	1359	34	1308	27	99.635%	1	0.57287	0.57261	0.242944
99.423%	1366	34	1367	37	99.318%	0	0.49856	0.49857	0.246598
99.389%	1358	35	1326	27	99.635%	0	0.54592	0.54576	0.246099
99.389%	1349	35	1286	25	99.687%	1	0.58968	0.58941	0.240492
99.389%	1356	35	1374	25	99.687%	1	0.47412	0.47420	0.247775
99.423%	1364	34	1325	67	97.813%	1	0.55589	0.55468	0.236325
99.423%	1371	34	1378	38	99.281%	1	0.48993	0.49000	0.246317
99.423%	1379	34	1400	36	99.354%	0	0.46982	0.47001	0.245891
99.423%	1371	34	1393	25	99.687%	0	0.46838	0.46848	0.247449
99.423%	1364	34	1184	23	99.735%	1	0.73811	0.73758	0.192531
99.423%	1361	34	1391	26	99.661%	0	0.45693	0.45708	0.246479
99.354%	1354	36	1403	27	99.635%	1	0.42995	0.43020	0.243341
99.389%	1363	35	1327	28	99.607%	1	0.55162	0.55142	0.245418
99.389%	1370	35	1377	41	99.164%	1	0.48993	0.49001	0.245739
99.389%	1378	35	1378	33	99.456%	0	0.50000	0.50000	0.247288
Overall data	1360	35	1348	32	99.455%	15	12.91198	12.90989	5181.873
						$S_A - E_A$	2.08802	2.09011	5.650699
						Rating change	+83.521	+11.81059	

Appendix 2: Relevant data of 25 games in 3 mins format

Blitz (3 mins)									
$g(RD_A)$	R_A	RD_A	R_B	RD_B	$g(RD_B)$	S_A	E_A (Elo)	E_A (Glicko)	δ^2
94.597%	1001	108	1023	25	99.687%	1	0.46838	0.46848	0.24745
94.597%	1034	108	986	24	99.711%	1	0.56864	0.56845	0.24390
95.050%	1060	103	1121	21	99.779%	0	0.41310	0.41329	0.24141
95.402%	1038	99	1048	30	99.550%	1	0.48561	0.48568	0.24755
95.743%	1068	95	1058	25	99.687%	0	0.51439	0.51434	0.24823
95.992%	1043	92	1003	25	99.687%	0	0.55731	0.55713	0.24519
96.235%	1018	89	1068	32	99.488%	0	0.42854	0.42890	0.24244
96.471%	1000	86	1095	36	99.354%	1	0.36659	0.36741	0.22943
96.625%	1025	84	973	23	99.735%	1	0.57428	0.57409	0.24322
96.777%	1041	82	972	21	99.779%	1	0.59801	0.59780	0.23937
96.925%	1056	80	1083	40	99.204%	0	0.46122	0.46153	0.24458
97.070%	1040	78	1002	42	99.123%	0	0.55447	0.55400	0.24277
97.212%	1022	76	1006	23	99.735%	1	0.52301	0.52295	0.24815
97.351%	1037	74	1123	23	99.735%	0	0.37870	0.37901	0.23411
97.420%	1026	73	996	39	99.243%	0	0.54307	0.54274	0.24443
97.554%	1010	71	935	22	99.757%	0	0.60629	0.60604	0.23760
97.620%	993	70	1041	31	99.520%	1	0.43136	0.43168	0.24298
97.750%	1008	68	973	25	99.687%	0	0.55020	0.55004	0.24595
97.813%	994	67	1067	45	98.996%	0	0.39646	0.39747	0.23470
97.876%	984	66	1000	24	99.711%	0	0.47699	0.47706	0.24803
97.938%	972	65	935	25	99.687%	1	0.55305	0.55288	0.24566
97.999%	983	64	1009	30	99.550%	1	0.46265	0.46282	0.24638
98.059%	995	63	918	49	98.812%	0	0.60903	0.60778	0.23275
98.119%	981	62	979	22	99.757%	0	0.50288	0.50287	0.24878
98.177%	970	61	953	26	99.661%	1	0.52445	0.52436	0.24772
Overall data	1016	79	1015	29	99.545%	11	12.54869	12.54881	4969.35851
						$S - E$	-1.54869	-1.54881	15.98957
						Rating change	-61.948	-24.76482	